

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FÍSICA

**Representações Reduzidas
por Redes Neurais
com Aprendizado Local Não-Supervisionado
sobre Campos Receptivos Restritos.***

Filipe Ronald Noal Souza

Dissertação realizada sob orientação do Prof. Dr. Marco Aurélio Pires Idiart (IF-UFRGS) e apresentada ao Instituto de Física da UFRGS em preenchimento parcial dos requisitos para a obtenção do título de Mestre em Física.

Porto Alegre
Dezembro de 2005.

* Trabalho financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq.

“Não temas o progresso lento, mas receia ficar parado.”
– Da sabedoria popular chinesa

Agradecimentos

★ Ao meu orientador, Prof. Dr. Marco Aurélio Pires Idiart, pelo ensino, pela paciência e pela compreensão, mormente nos momentos mais adversos pelos quais passei, jamais deixando de acreditar e investir em mim mesmo quando eu próprio não acreditava.

★ À minha família e mais particularmente ao meu estimado irmão Raphael Ronald, que sempre fez o que esteve ao seu alcance para me estimular e assistir.

★ Aos meus sogros e principalmente à minha querida namorada Daniela Rosa, que sempre soube estar presente nos momentos em que precisei oferecendo sua mão, seu ombro e seu coração.

★ Aos meus amigos da graduação no Departamento de Física da UFPR, que mesmo à distância me apoiaram e torciam por mim.

★ À Universidade da Região da Campanha sede Alegrete, nas pessoas do Prof. Fábio Rossi e do acadêmico Rafael Araújo Rodrigues, ambos do Departamento de Informática, cujos conhecimentos em programação foram de grande auxílio em minhas simulações.

★ Aos meus amigos da sala M201, do Instituto de Física da UFRGS, com quem compartilhei a maior parte dos momentos durante o período de 2000 a 2002 e que agora encontram-se espalhados por esta e outras instituições.

★ Aos colegas de aula das cadeiras cursadas nos anos de 2000 e 2001 no Instituto de Física da UFRGS com os quais compartilhei dúvidas e dos quais obtive estímulo sempre que precisei.

★ Aos professores e funcionários do Instituto de Física da UFRGS que de alguma forma contribuíram para a minha formação e/ou para a realização deste trabalho.

★ A Deus, especialmente.

Muito Obrigado.

Resumo

Algoritmos ótimos na extração de componentes principais com aprendizado não-supervisionado em redes neurais de múltiplos neurônios de saída são *não-locais*, ou seja, as modificações em uma dada sinapse entre dois neurônios dependem também da atividade de outros neurônios. Esta rede ótima extrairá as principais componentes dos dados \vec{e} submetidos à sua primeira camada. As principais componentes são as projeções destes vetores nos autovalores máximos da matriz de correlação $C_{ij} = \langle e_i e_j \rangle$, onde a média $\langle \cdot \rangle$ é sobre a distribuição de \vec{e} .

Existem fortes evidências indicando que sinapses biológicas só se modificam via regras locais, como por exemplo a regra de Hebb. Mas se aplicarmos regras locais numa rede com múltiplas saídas, todos os neurônios da saída serão equivalentes e darão respostas redundantes. A rede será bastante ineficiente.

Um modo de contornar este problema é através da restrição dos campos receptivos dos neurônios de saída. Se cada neurônio acessar diferentes partes dos estímulos de entrada, a redundância diminui significativamente. Em contrapartida, ao mesmo tempo que a redundância diminui, também diminui a informação contida em cada neurônio; assim, devemos balancear os dois efeitos otimizando o campo receptivo. O valor ótimo, em geral, depende da *natureza dos estímulos*, sua *estatística*, e também do *ruído intrínseco à rede*.

Objetivamos com este trabalho determinar a estrutura ótima de campos receptivos com aprendizado não-supervisionado para uma rede neural de uma camada em diversas condições medindo seu desempenho a partir de técnicas de reconstrução.

Abstract

Optimal algorithms to extract principal components with unsupervised learning in neural networks with multiple output neurons are *non-local*, that is, the changes in a given synapsis between two neurons depend on activities of other neurons as well. That optimal network will extract the principal components from the data \vec{e} submitted to the first layer. The principal components are the projections of these vectors over the correlation matrix $C_{ij} = \langle e_i e_j \rangle$ maximum eigenvalues, where the mean $\langle \cdot \rangle$ is over all distribution of \vec{e} .

There are strong evidences indicating that biological synapsis only change themselves through local rules, like the Hebb's rule. But if we apply local rules in a network with multiples units, all the output neurons will be equivalent and they will give redundant responses. The network will be very inefficient.

A good way to cope with that problem is restricting the output neurons receptive fields. If each neuron accesses different parts of the input stimulus the redundance falls in an expressive way. At the same time, the information content in each neuron also decreases; thus we must calibrate the two effects to optimize the receptive field. In general, the optimum value depends on the *stimulus nature*, its *statistics*, and on the *network intrinsic noise*.

With this work, we intend to determine the receptive fields' optimum structure submitted to unsupervised learning rule in a single layer network on several conditions measuring its performance by reconstruction techniques.

Sumário

Introdução	1
1. O Sistema Biológico e os Sistemas Artificiais	8
1.1 O Sistema Visual	10
1.2 O Sistema Visual Primário	11
1.3 Sistemas Artificiais	12
1.3.1 Simulações Computacionais	13
1.3.2 Modelo de McCulloch e Pitts	14
2. Redes Neurais	17
2.1 Modelos de Redes Neurais	17
2.1.1 Percéptrons Simples	19
2.1.2 Aprendizados de Redes Neurais	20
2.1.3 Aprendizagem com Retropropagação de Erro (<i>Back-propagation</i>)	24
2.2 Análise de Componentes Principais (PCA)	26
2.3 Campos Receptivos Restritos	30
3. Estatística de Imagens Naturais	36
3.1 Invariância de Escala e Auto-afinidade	37
3.1.1 Espectro de Potência	38
3.1.2 Previsibilidade em Cenas Naturais	39
3.2 Codificação de Imagens Naturais	40
3.2.1 Código Compacto e PCA	41
3.2.2 Código Esperso e Estatística Estacionária	43

3.3	Aproximações de Imagens Naturais.	45
3.3.1	Fractais	45
3.3.2	Caminhante Aleatório	49
4.	Procedimentos e Resultados	51
4.1	Espectro de Potência	53
4.2	Campos Receptivos em uma Rede Linear	55
4.2.1	Cenas Naturais	57
4.2.2	Cenas Artificiais	62
4.2.3	Cenas Geradas por um Caminhante Aleatório	65
4.2.4	Cenas Fractais	67
	Conclusão	73
	A. Componentes Principais: um Exemplo Simples	77
	B. Erro Quadrado Médio	80
	C. Modelos Matemáticos de Fractais	83
	D. <i>Ensembles</i> de Imagens	87
	Referências	92

Lista de Figuras

0.1	Visão esquemática simplificada da configuração biológica das células responsáveis pela visão. A luz, representada pela seta à esquerda, carrega a informação visual do meio externo através do olho até atingir a matriz de <i>células fotorreceptoras</i> (centro). No olho humano, existem em média 10^6 <i>células ganglionares</i> (matriz à direita), cada uma delas conectada ao seu próprio conjunto de células fotorreceptoras.	3
0.2	Diferença de sinais descartada em um processo de conversão de uma imagem de formato <i>bit-map</i> para o formato JPEG. A imagem mais acima à esquerda foi escaneada em tons de cinza de 8 bits. Em seguida, foi convertida para JPEG (acima à direita) com a qual a subtração das matrizes de pontos foi efetuada. A imagem mais abaixo apresenta os sinais que foram descartados neste processo.	4
1.1	Esquema simplificado da secção de um globo ocular representando o <i>processo visual primário</i> . À esquerda, a imagem de um objeto orientado é projetada pelo cristalino (a elipse hachurada no centro da qual se dá a intersecção das duas retas) no fundo do olho, onde encontra-se a <i>retina</i> ; à direita, uma ampliação da camada mais interna da retina, onde se localizam os <i>fotorreceptores</i> , as células visuais responsáveis pela percepção da luz e das informações contidas nesta. Estas células podem ocupar-se com características distintas da luz, dependendo de sua natureza.	12

- 1.2 Esquema simplificado de um *neurônio biológico típico* onde ficam indicadas suas principais partes: os impulsos são recebidos como sinais de entrada de outros neurônios através das *sinapses* (a) ligadas aos *dendritos* (b). O *corpo da célula* (c), onde o *núcleo* fica abrigado, recebe e processa os sinais captados por todos os dendritos gerando respostas, que são enviadas através do *axônio* (d) – ao final do qual se localizam outros dendritos (e) – para então serem transmitidas às demais células neurais (f). 14
- 1.3 Esquema da *unidade neural de disparo unitário* proposto em 1943 por McCulloch e Pitts. As conexões w_1, w_2, w_3 , se ativas, são primeiramente somadas no neurônio ao qual se ligam. Na etapa seguinte, o valor resultante é comparado com o *limiar* μ_i e, se superior, leva ao disparo do neurônio. 15
- 2.1 *Percéptrons*: os círculos cheios correspondem aos terminais de alimentação da rede e não contribuem no processamento, de modo que não são incluídos na contagem das camadas. Os vazados são as unidades onde as informações são processadas. Do lado esquerdo, uma rede de *percéptrons simples* (somente uma camada); do lado direito, uma rede de *percéptrons multicamadas*. A primeira camada está oculta por não ter ligação com o meio externo. 19
- 2.2 Esquema da *unidade neural de disparo unitário* com função de ativação do tipo *linear*. As componentes do vetor entrada são ponderadas pelo vetor peso e somadas em um escalar antes de servirem de parâmetro para a função de ativação. 20
- 2.3 Arquitetura de uma rede com duas camadas para treinamento baseado em *retropropagação de erro*. Os círculos brancos representam a *camada oculta*, que não mantém ligação direta com o ambiente externo; os círculos cinzentos representam a *camada de saída*, que tem a mesma dimensão dos terminais de entrada tendo em vista a aplicação da rede. 24

2.4 Projeção das *duas componentes principais em um espaço amostral bidimensional*. A distribuição gaussiana de padrões (pontos) apresenta maior variância ao longo da dimensão horizontal, como pode ser visto pela estreita elipse hachurada sobre o eixo. Sendo assim, a primeira componente principal segue a orientação de w_1 . A elipse hachurada mais arredondada sobre o eixo vertical indica a menor variância e, por conseguinte, indica também que a segunda componente principal corre ao longo de w_2 27

2.5 Resultado da aplicação em rede linear de entradas bidimensionais $\vec{e} = (e_1, e_2)$ distribuídas de acordo com os pontos da figura. A evolução da orientação do vetor peso \vec{w} é vista seguindo a linha mais fina: (A) percebe-se que \vec{w} atinge a orientação desejada, mas segue crescendo devido à inexistência na regra de Hebb de uma condição de parada; (B) a regra de Oja (a regra de Hebb modificada) encarrega-se de manter \vec{w} com tamanho finito uma vez que a orientação seja atingida. (*Adaptado de Hertz et al [9].*) 28

2.6 Modelo de arquitetura de *rede neural com campos receptivos restritos* composta por um vetor entrada \vec{e} N -dimensional e um vetor com M unidades de saída \vec{S} . Cada uma das unidades de saída abarca um campo receptivo contendo r elementos do vetor entrada. 31

2.7 Representação do processo de extração de uma matriz de $M = M' \times M'$ campos receptivos sobrepostos entre si, com dimensão $r_x \times r_y$ (onde $r_x = r_y$), a partir de um padrão de entrada $N \times N$ 34

3.1 *Espectro de potência bidimensional médio* extraído de 10 cenas naturais. A origem localiza-se no centro da superfície, onde se encontra o ponto máximo. Regiões equidistantes da origem têm a mesma frequência. 39

3.2 À esquerda, a *curva de uma distribuição esparsa* (linha contínua) sobrepondo uma curva gaussiana (linha interrompida). A distribuição esparsa tem altos valores em regiões próximas à origem e decai suavemente à medida que se afasta do zero. À direita, um esquema representa o comportamento de um *processo regido segundo a distribuição esparsa*. O número de descritores (círculos negros) é bastante grande, mas um número pequeno deles é suficiente para representar a imagem. 44

3.3 *Conjunto de Mandelbrot*: fractal descoberto por Benoit B. Mandelbrot, cujo nome foi emprestado para batizá-lo. Trata-se de uma imagem infinitamente detalhada “comprimida” em um espaço finito e que foi gerada no plano complexo a partir de processos iterativos sobre a expressão $z_{n+1} = z_n^2 + c$, onde $z_0 = 0$ e c é uma constante complexa. 46

3.4 Três imagens 128×128 pixels compostas por exemplos de *censo fractais* com dimensão $d_F = 2,5$ criadas sem somas aleatórias e com desvio padrão igual a 1 através do processo de geração de movimento browniano pelo método de *substituição do ponto médio* para duas dimensões. 48

3.5 Três imagens exemplificando possíveis *caminhantes aleatórios bidimensionais* com dimensão 128×128 pontos. Usamos condição de contorno periódica levando o caminhante a ser “transportado” de um lado para o outro quando alcança uma borda, sendo que o caminhante viajou pelo espaço até que a média aritmética dos pontos alcançasse um valor de escape γ . Da esquerda para a direita, os valores de escape utilizados foram respectivamente $\gamma = 1/5$, $2/5$ e $1/2$ 50

4.1 *Logaritmo da média do espectro de potência* de 6 *ensembles* distintos (com 30 imagens 128×128 cada) plotado em função do logaritmo da frequência espacial. As retas sobrepondo os pontos são ajustes lineares realizados através do método dos mínimos quadrados para $\log k \leq 3,5$ 54

- 4.2 Erro médio de reconstrução E pela variação do campo receptivo r_o . A rede foi treinada com *cenais naturais*, e os gráficos (A), (B), (C) e (D) representam seu desempenho quando submetida a ruídos de natureza gaussiana com variância 0, 1, 10 e 100, respectivamente. A curva tracejada separa os pontos obtidos de campos sem sobreposição dos pontos vindos de campos com sobreposição. A escala dos eixos está normalizada em $E \times r_o/N$ 58
- 4.3 Erro médio de reconstrução sobre *cenais naturais* para uma rede com *campos receptivos restritos* (as duas curvas com barras de erro indicadas pela legenda) e para outra treinada com *retropropagação de erro* (as duas retas paralelas indicadas por B_M). 59
- 4.4 Distribuição dos 30 autovalores mais significativos das correlações das *cenais naturais* com dimensão reduzida para 64×64 , 32×32 e 16×16 pixels. A magnitude dos mesmos foi normalizada à unidade segundo os λ_{\max} do quadro central. 61
- 4.5 Reconstrução da 15ª cena natural do *ensemble* D.1. Da esquerda para a direita, as imagens representam os pontos $E_{4,65}$, $E_{25,30}$ e $E_{25,65}$. As últimas colunas (em destaque), são a reconstrução da mesma imagem por outras duas redes: a primeira corresponde a uma *rede híbrida* treinada com *cenais artificiais* usando os parâmetros do ponto $E_{25,30}$, e a segunda é um *codificador* 128×128 -25- 128×128 62
- 4.6 Erro médio de reconstrução E pela variação do campo receptivo r_o . A rede foi treinada com *cenais artificiais*, e os gráficos (A), (B), (C) e (D) representam seu desempenho quando submetida a ruídos de natureza gaussiana com variância 0, 1, 10 e 100, respectivamente. A curva tracejada separa os pontos obtidos de campos sem sobreposição dos pontos vindos de campos com sobreposição. A escala dos eixos está normalizada em $E \times r_o/N$ 63
- 4.7 Erro médio de reconstrução sobre *cenais artificiais* para uma rede com *campos receptivos restritos* (as duas curvas com barras de erro indicadas pela legenda) e para outra treinada com *retropropagação de erro* (as duas retas paralelas indicadas por B_M). 64

4.8 As linhas superior e inferior contêm respectivamente as reconstruções das 6ª e 22ª cenas artificiais do *ensemble* D.2. Da esquerda para a direita, as imagens representam os pontos $E_{4,65}$, $E_{25,30}$ e $E_{25,65}$. A última coluna (em destaque), é a reconstrução das mesmas imagens por uma *rede híbrida* treinada com *cenais naturais* usando os parâmetros do ponto $E_{25,30}$ 65

4.9 Erro médio de reconstrução E pela variação do campo receptivo r_o . A rede foi treinada com cenas de um *caminhante aleatório*, e os gráficos (A), (B), (C) e (D) representam seu desempenho quando submetida a ruídos de natureza gaussiana com variância 0, 1, 10 e 100, respectivamente. A curva tracejada separa os pontos obtidos de campos sem sobreposição dos pontos vindos de campos com sobreposição. A escala dos eixos está normalizada em $E \times r_o/N$. 66

4.10 Erro médio de reconstrução sobre *caminhantes aleatórios* para uma rede com *campos receptivos restritos* (as duas curvas com barras de erro indicadas pela legenda) e para outra treinada com *retropropagação de erro* (as duas retas paralelas indicadas por B_M). 67

4.11 Erro médio de reconstrução E pela variação do campo receptivo r_o . A rede foi treinada com *cenais fractais*, os gráficos (A), (B), (C) e (D) representam seu desempenho quando submetida a ruídos de natureza gaussiana com variância 0, 1, 10 e 100, respectivamente. A curva tracejada separa os pontos obtidos de campos sem sobreposição dos pontos vindos de campos com sobreposição. A escala dos eixos está normalizada em $E \times r_o/N$ 68

4.12 Erro médio de reconstrução sobre *cenais fractais* para uma rede com *campos receptivos restritos* (as duas curvas com barras de erro indicadas pela legenda) e para outra treinada com *retropropagação de erro* (as duas retas paralelas indicadas por B_M). 69

- 4.13 Erro médio de reconstrução E pela variação do campo receptivo r_o . Em todos os casos, a rede foi treinada com *cenais naturais*, e os diversos quadrantes representam o desempenho da rede a partir da reconstrução de imagens de naturezas distintas e sem ruído: o gráfico (A) contém *cenais naturais*, o gráfico (B) contém *cenais artificiais*, o (C) contém *cenais do caminhante aleatório*, e o (D) contém *cenais fractais*. 70
- 4.14 Cada curva em cada gráfico apresenta o erro médio de reconstrução para uma arquitetura de rede neural distinta. O eixo das abscissas marca a natureza das imagens do *ensemble* trabalhado pela rede. Na escala do eixo é exibido também o valor médio da variância dos pixels. 71
- 4.15 Cada curva em cada gráfico apresenta o erro médio de reconstrução para uma arquitetura de rede neural distinta. O eixo das abscissas marca a variância σ^2 do ruído gaussiano ao qual a rede foi submetida no treinamento. 72
- A.1 Espaço de ϵ^2 padrões de treinamento para um *perceptron* simples linear com regra de aprendizagem de Hebb. No lado (A), os padrões (quadrados vazados) e o vetor sináptico w^0 , que está em sua posição inicial; no lado (B), a evolução do crescimento do vetor sináptico ao longo das iterações do processo de aprendizagem da rede. Somente w^0 e w^1 aparecem com a magnitude original. 77
- D.1 População de 30 imagens composta por *cenais naturais* com dimensão 128×128 pixels cada. 88
- D.2 População de 30 imagens composta por *cenais artificiais* (estruturas criadas pelo homem) com dimensão 128×128 pixels cada. 89
- D.3 Exemplos de 4 *ensembles*: estas populações possuem 30 padrões distintos cada com a mesma dimensão 128×128 pixels. A primeira é um exemplo da trajetória de um *caminhante aleatório bidimensional*, a segunda, um *fractal*, a terceira é uma *cena natural filtrada pelo logaritmo de Ruderman*, e a última é um *ruído branco*. 91

D.4	Imagens de dimensão 128×128 cujas intensidades representam as variâncias dos pixels correspondentes aos <i>ensembles</i> em estudo. Da esquerda para a direita, temos <i>cenais naturais</i> , <i>cenais artificiais</i> , <i>caminhante aleatório</i> e <i>cenais fractais</i>	91
-----	--	----

Lista de Tabelas

- A.1 Valores correspondentes às 4 primeiras iterações no processo de treinamento utilizando-se a regra de aprendizagem de Hebb. A última coluna apresenta a evolução do ângulo do vetor peso à medida em que o aprendizado é realizado. 78

Introdução

Introdução

Na era da informação, quase que completamente calcada em nosso sistema visual, verifica-se uma necessidade cada vez maior de se trabalhar com imagens (com ou sem movimento), melhorando a transmissão de dados e instruções e, deste modo, a comunicação entre pessoas. Compreender o processo biológico empregado na compactação de uma imagem, desde sua apreensão pelos receptores celulares até a decodificação feita no cérebro, é um avanço importante no desenvolvimento de ferramentas tecnológicas que agilizem a troca de informações.

A natureza dotou o olho humano com *fotorreceptores*, células com a tarefa específica de captar a luz, e com *células ganglionares*, para onde convergem os sinais emitidos pelos fotorreceptores. Existem aproximadamente 10^6 células ganglionares no olho humano e, em média, 125 células fotorreceptoras para cada célula ganglionar (conferir com a figura 0.1; mais detalhes serão discutidos na seção 1.2). Segundo alguns autores [7, 9], toda imagem pode ser decomposta em componentes mais representativas, dada a quantidade de informações repetidas. Se, numa hipótese extremamente rudimentar, supusermos que de cada matriz de 125 células fotorreceptoras uma célula ganglionar puder extrair uma dessas componentes, reduziremos em 125 vezes a imagem original e ainda conservaremos cerca de um milhão de “frações informativas” (uma para cada célula ganglionar) com as quais reconstituí-la no cérebro.

Essas componentes, que aqui podem ser chamadas de *componentes principais*, dependem da estatística do sistema. Existem muitos métodos para determinar componentes principais. A *análise de componentes principais por redes neurais* (conferir a seção 2.2) é um formalismo iterativo que prima por aproximar-se de seu equivalente biológico. Consiste em estudar os padrões que caracterizam um conjunto de dados de entrada (em nosso caso, um conjunto de imagens) para extrair os critérios de seleção entre o que é tido como representativo e o

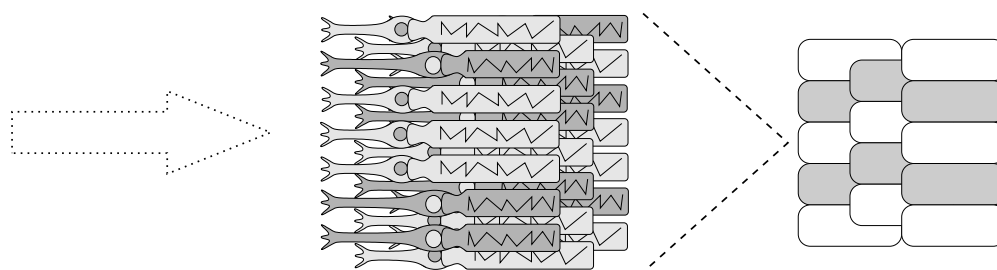


Fig. 0.1: *Visão esquemática simplificada da configuração biológica das células responsáveis pela visão. A luz, representada pela seta à esquerda, carrega a informação visual do meio externo através do olho até atingir a matriz de células fotorreceptoras (centro). No olho humano, existem em média 10^6 células ganglionares (matriz à direita), cada uma delas conectada ao seu próprio conjunto de células fotorreceptoras.*

que pode ser desprezado.

É evidente que os atuais modelos computacionais de redes neurais pouco reproduzem as propriedades e características do processamento biológico de imagens, dada a sofisticação deste. Todavia, inúmeros trabalhos vêm sendo realizados no intuito de melhorar nossa compreensão da natureza visando a conceber melhores modelagens. Ainda que pouco pretensiosos, os resultados são favoráveis e estimulantes.

O formato universalmente conhecido de representação de imagens digitalmente é o *bit-map* ou *mapeamento por bits*. Com ele, é possível definir eventos ou dados usando um arranjo de pontos de dimensão mínima distribuídos linear e uniformemente por uma matriz. Cada um desses pontos ou *pixels* pode ser individualmente controlado através da troca do valor de seu bit armazenado. Se for 1, estará ligado; se 0, desligado. Quando se armazena mais de um bit por pixel, consegue-se reproduzir as cores vermelha, verde e azul em cada ponto, compondo, deste modo, uma imagem cromática.

Microcomputadores normalmente usam vídeos de tubo de raios catódicos que se utilizam de mapas de bits para exibir imagens retratando textos, gráficos ou fotos. Também nos computadores, os bit-maps são empregados no armazenamento de imagens para inserção em documentos de texto ou trabalhos de artes gráficas. Neste caso, contudo, este mapeamento costuma ocupar muito espaço físico além de mostrar-se pesado para o manuseio computacional. O princípio para armazenar a imagem em uma mídia magnética é idêntico ao da sua exibição em um monitor: uma cor por ponto; uma seqüência de bits por cor.

Para uma imagem com dimensão 1024×1024 pixels e 16 milhões de cores, necessita-se de aproximadamente 3 megabytes de espaço em disco¹.

Com a dependência cada vez maior por imagens de alta qualidade sem perder de vista a economia de espaço, surgiram os formatos compactados, como o *Graphical Interchange Format* (GIF) e o *Joint Photographic Experts Group* (JPEG). Estes formatos foram criados para filtrar propriedades e características menos relevantes na definição da imagem de modo a permitir a redução do tamanho sem prejudicar seu reconhecimento. Um exemplo pode ser visto na figura 0.2.

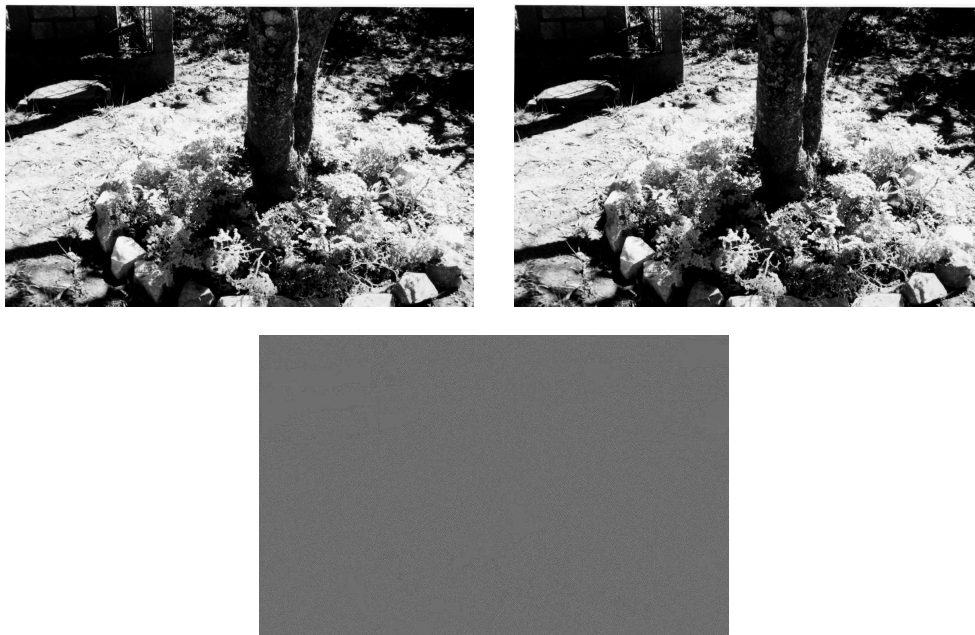


Fig. 0.2: *Diferença de sinais descartada em um processo de conversão de uma imagem de formato bit-map para o formato JPEG. A imagem mais acima à esquerda foi escaneada em tons de cinza de 8 bits. Em seguida, foi convertida para JPEG (acima à direita) com a qual a subtração das matrizes de pontos foi efetuada. A imagem mais abaixo apresenta os sinais que foram descartados neste processo.*

Ainda que os formatos GIF, JPEG, etc... utilizem critérios distintos entre si para a redução do tamanho da imagem, estes critérios se mantêm constantes para todas as naturezas de figuras, sejam fotos de família, paisagens naturais ou constelações interplanetárias. Por outro lado, um algoritmo neural adequadamente preparado é capaz de particularizar critérios, de

¹ Uma imagem de 16 milhões de cores é composta por pixels com palavra de 24 bits, isto é, $2^4 = 16.777.216$. Este valor de 3 bytes por palavra multiplicado pela dimensão da imagem resulta no espaço em bytes ocupado pela imagem.

maneira que tratará uma foto de um grupo de pessoas distintamente de uma paisagem com árvores e cachoeiras. Empregando o método de análise de componentes principais, uma rede de neurônios sujeita a estímulos adequados pode ser ensinada a determinar as principais componentes de um grupo de imagens de mesma natureza lhe possibilitando trabalhar com elas de modo muito eficiente. Retendo consigo o montante mais representativo de informações da imagem, essas componentes poderão ser armazenadas em um espaço otimizado. Em outras palavras, uma rede neural desta qualidade pode ser vista como um compactador personalizado para um dado tipo de necessidade.

Voltando ao sistema biológico, vamos considerar um exemplo ilustrativo para efeito de comparação.

Suponhamos um homem nascido e criado em ambiente urbano sendo subitamente jogado no interior de uma floresta selvagem e lá deixado a vagar abandonado. Descartando emoções básicas como medo e assumindo que não lhe faltasse as exigências fisiológicas mínimas como água e comida, podemos tentar teorizar rudimentarmente sobre a confusão pela qual seria tomado. Numa primeira etapa, este homem passaria horas caminhando ao longo de quilômetros sem notar a diferença entre a primeira árvore que viu e a última em que parou. Para qualquer lado em que olhasse, sempre veria as mesmas coisas: árvores, plantas, folhas e outros seres vivos. Mesmo que não estivesse andando em círculos, pensaria o contrário. Numa segunda etapa, se tivesse sobrevivido imerso à vegetação, teria delimitado uma pequena extensão da região para se estabelecer. Mantendo-se restrito a este território, alguns dias seriam suficientes para que se adaptasse melhor ao ambiente adquirindo senso de direção.

Na etapa inicial, um homem que passou a vida rodeado de prédios jamais tendo visto mais árvores do que as encontradas num parque urbano, não possuiria uma visão preparada para um ambiente selvagem. O tempo que passou na civilização condicionou seu sistema visual a distinguir formas não naturais. Deparando-se com a vegetação densa, será incapaz de extrair informações importantes que lhe permitam diferir um cenário natural de outro também natural. Não podendo estabelecer referências, fatalmente se perderá.

Na etapa seguinte, contudo, alguns dias passaram, e o aventureiro teve oportunidade de

conviver exaustivamente entre árvores e cipós. Esta exposição exaustiva permitiu a formação em seu sistema visual de uma estatística para as imagens naturais. Paulatinamente, elementos importantes são mais facilmente percebidos em sua vizinhança, suas referências melhoram e, conseqüentemente, seu senso de direção².

O cérebro dos mamíferos apresenta capacidades espantosas pelo tamanho que ocupa se comparado ao desempenho de um sistema artificial. Além destes últimos serem muito maiores, são menos eficientes quantitativa e qualitativamente em um grande número de tarefas, principalmente as que envolvem estímulos imprecisos, difusos ou distorcidos. Eis o porquê do seu estudo vir despertando tanto fascínio e movimentando tantos pesquisadores.

Nosso trabalho, pois, consiste em dar mais um passo em direção à sua compreensão focalizando-nos particularmente no sistema visual primário. Compreender este não se limita unicamente ao mecanismo biológico responsável pelo aprendizado sobre o meio, mas também ao estudo das estatísticas das imagens que compõem seu *habitat*. As redes neurais aqui empregadas mostram-se importantes ferramentas e fornecem excelentes resultados.

Estamos interessados em: reproduzir os resultados obtidos em Idiart *et al* [12] usando a técnica de teoria da informação com ênfase na obtenção de resultados analíticos; estudar a dependência dos campos receptivos ótimos com a dimensionalidade da rede – em particular, estudar em detalhe o caso mais realístico de uma rede bidimensional; estudar a dependência dos campos receptivos ótimos com a natureza dos estímulos – em particular, estudar o caso em que os estímulos possuem invariância translacional e invariância de escala, propriedades comuns em *imagens naturais*; estudar o comportamento dos campos receptivos ótimos quando o aprendizado é incompleto ou existe ruído na regra de aprendizagem.

O capítulo 1 aprofundará as idéias aqui introduzidas sobre o sistema biológico enfatizando as diferenças em relação aos sistemas artificiais. No capítulo 2, nos concentraremos mais diretamente nas redes neurais artificiais. Trataremos de seu conceito e veremos os modelos mais importantes para o desenvolvimento de nosso trabalho. Introduziremos também o modelo de *campos receptivos restritos* (seção 2.3) que visa a buscar soluções locais para problemas normalmente abordados com técnicas não-locais. O capítulo 3 apresen-

² A hipótese descrita acima não leva em consideração variáveis como iluminação do meio ou a idade do agente. Sabe-se, contudo, terem estes parâmetros grande importância na estatística do sistema. Animais jovens têm de aprender rapidamente sobre seu meio para que possam sobreviver.

tará propriedades e características estatísticas determinadas por diversos autores a respeito de imagens naturais. Veremos também imagens criadas artificialmente que, por suas propriedades estatísticas, servem como possíveis aproximações de cenas naturais. No capítulo 4, abordaremos os procedimentos e as técnicas adotados na realização deste trabalho e discutiremos os resultados obtidos. Apêndices também foram adicionados a fim de abordar temas ou detalhar cálculos que não pareceram pertinentes ao longo do texto.

Capítulo 1

O Sistema Biológico e os Sistemas Artificiais

A máquina biológica é um sistema dotado de capacidades fantásticas encapsuladas de forma eficientemente enxuta. Como todo sistema biológico, é o resultado de anos de evolução e adaptação às intempéries do meio, como relevo, clima e demais interações com a vizinhança.

Ao longo dos milhões de anos que separam os dias atuais do momento do surgimento do primeiro microorganismo, estas adversidades propiciaram que as espécies se ramificassem de forma muito rica nos organismos que conhecemos hoje. Seres indiscutivelmente distintos entre si portando habilidades próprias para enfrentar as dificuldades de seu *habitat*. Cada espécie com características únicas, mas todas elas com o mesmo objetivo: sobreviver e propagar seus gens para as gerações seguintes.

A capacidade do indivíduo de uma espécie depende essencialmente – entre outras coisas – da programação de seu código genético. Algumas espécies reagem puramente a partir daquilo que talvez possamos rudimentarmente caracterizar como uma motivação básica e inata conhecida popularmente como *instinto*. A outras espécies mais evoluídas é concedido o dom de *aprender*, dom este que dá ao ser a possibilidade de sobrepujar o instinto. O aprendizado é uma faculdade restrita aos organismos dotados de *sistema nervoso*.

O *cérebro* é o órgão mais importante de qualquer organismo vivo que o possua. Nele, encontramos o centro de operações responsável pela administração de todas as demais funções necessárias à sobrevivência do ser, sejam estas operações voluntárias e deliberadas, sejam involuntárias e transparentes à percepção consciente daquele. Por ele, passam questões como o controle do funcionamento dos órgãos, a regulação das temperaturas interna e externa, o aviso sobre a necessidade de alimento ou água, a administração dos sentidos, a tomada de decisões, etc. Raras são as ações que são executadas sem passar antes pelo crivo cerebral, e

certamente a existência delas está vinculada a uma necessidade maior de agilidade – como é o caso do *ato-reflexo*.

Nem todos os organismos possuem cérebro, mas sabemos que, dentre todos os seres vivos dotadas de um, os mamíferos apresentam a versão mais evoluída. Os cães, por exemplo, são treinados para atividades que variam desde a proteger a casa, passando por espetáculos circenses e culminando com salvamento de vidas em regiões inóspitas.

O próprio ser humano tem necessidade de passar os primeiros vinte anos de sua vida preparando-se em instituições de ensino para aprender a lidar com o mundo onde nasceu. Aliás, o cérebro humano destaca-se dentre todos os demais, seja pelo *tamanho relativo*¹, pela *plasticidade* ou pela *quantidade de funções complexas* que é capaz de executar. Apesar de não sermos os únicos seres a ensinar nossos descendentes, somos os únicos capazes de gerar conhecimento e os únicos dotados da curiosidade necessária para motivar esta geração. Assim, pesquisamos, descobrimos, aprendemos, aplicamos e passamos adiante como nenhuma outra espécie conhecida já fez.

Mas como funciona este cérebro tão especial? Como suas células, tão pequenas e simples, se comunicam e conseguem um efeito global tão rico e complexo? Inúmeros são os meios empregados pelos pesquisadores no intuito de compreendê-lo melhor. Um dos processos de vanguarda está na *ressonância magnética funcional*, capaz de mapear no espaço tridimensional o funcionamento coletivo das células neurais a partir de estímulos diretos ou indiretos.

O entendimento de seu funcionamento nos permitiria criar versões artificiais com habilidades equivalentes ou quiçá superiores? Poderíamos desenvolver ferramentas inteligentes que nos substituiriam em atividades perigosas, complexas ou mesmo entediadas? A eterna insatisfação do homem parece levá-lo a buscar estas respostas. Cada passo que avançamos no entendimento do sistema neural é um passo em direção a estes objetivos.

A despeito da assombrosa explosão tecnológica que presenciamos nas duas últimas décadas, ainda estamos milhares de anos atrás na esteira da evolução. O cérebro humano é dotado de habilidades bastante superiores às dos computadores digitais em muitas tarefas. Os sistemas de processamento sensoriais de um bebê – em incipiente desenvolvimento – são

¹ A razão entre a *massa encefálica* e a *massa corporal*.

bons exemplos. Sem qualquer dificuldade, mostra reconhecer a voz, o rosto ou mesmo o cheiro de seus pais em um tempo praticamente desprezível [9]. Para uma máquina funcionando a partir dos métodos convencionais introduzidos por von Neumann, esta mesma operação exigiria algoritmos complexos e horas de processamento.

Nas seções que seguem, restringiremos nossa discussão ao processo realizado pelo sistema visual dos mamíferos e aos modelos propostos por pesquisadores na tentativa de simulá-lo.

1.1 O Sistema Visual

O mundo que conhecemos através de nossos sentidos não é o mesmo que outras espécies conhecem através de seus sentidos. Cada um de nossos órgãos sensoriais está calibrado para receber uma determinada faixa de estímulos relevantes à nossa sobrevivência, sendo insensível fora desta faixa. Diferentes espécies têm diferentes faixas de sensibilidade porque têm diferentes necessidades de sobrevivência. Cada sentido responde a uma determinada forma de energia e, para a visão, o estímulo físico é a luz^2 [32].

Sabe-se que a habilidade visual varia de espécie para espécie de animal de acordo com as necessidades impostas pela natureza. Uma águia ou um lince, carnívoros predadores, têm de vislumbrar sua caça a metros, talvez quilômetros de distância. Para tanto, uma acuidade visual precisa é questão de sobrevivência. Algumas aves, como a coruja, são conhecidas pela caça noturna, para elas somente possível através de fotorreceptores muito sensíveis à intensidade de luz. Apesar de não possuir a maior acuidade visual do reino animal, o ser humano é privilegiado com a percepção de certas frequências luminosas – o que nos permite enxergar o mundo em milhares de matizes de cores.

Dentre os sentidos que o ser humano utiliza para retirar do meio as informações que precisa, a *visão* exerce papel fundamental nas interações do indivíduo com sua vizinhança e com outros indivíduos. O olho humano é capaz de extrair informações relevantes de ambientes com pouca ou muita iluminação. Sua faculdade de adaptação lhe permite, dentro de um certo espectro de intensidades, equilibrar a quantidade de luz e operar em diversos ambientes a que se sujeita. Este processo é automático e normalmente passa despercebidamente

² Aqui, adotamos o conceito de que a luz é a estreita faixa de frequência das ondas eletromagnéticas para a qual o olho humano é sensível, ou seja, o intervalo de 400 a 700 nanômetros.

ao acendermos a luz em um cômodo, por exemplo.

O sistema visual humano é constituído pelos *olhos*, por várias *partes do cérebro* e pelos *trajetos* que os conectam. Como predador, o homem requer uma visão preparada para se ajustar rapidamente ao seu alvo. Quando mudamos o olhar desviando a atenção do teclado do computador para uma árvore ao final da estrada vislumbrada através da janela, não mais do que frações de segundos são exigidas para que o olho se acomode ao novo foco e assimile uma imagem clara e nítida. Para tanto, músculos internos ao globo ocular se contraem ou relaxam de acordo com o ponto de focalização da visão.

Todas estas características fazem deste pequeno e discreto órgão um poderoso instrumento de absorção de informações. Por conseqüência, talvez não seja exagero dizer que a visão pode ser considerada como o mais importante dos sentidos de que somos dotados.

1.2 O Sistema Visual Primário

O meio ambiente dentro do qual nos inserimos é composto por milhares de matizes de cores, contrastes, diferenças de luminosidade, padrões, formas estruturais, tudo compondo uma imagem única que é projetada pelo cristalino à nossa retina, como pode ser observado na figura 1.1. Desta combinação complexa, pouca coisa parece ser efetivamente aproveitada pelo sistema visual primário, a estrutura constituída pelos olhos e que é responsável pela primeira etapa do processamento visual.

Uma vez que um feixe de luz oriundo do ambiente alcança o olho, atravessa o cristalino e é projetado sobre a retina, um sofisticado processo de transformação e interpretação de sinais elétricos e químicos tem início. No interior do globo ocular, um mosaico de fotorreceptores converte a energia luminosa para um potencial de membrana, que por sua vez é convertido para um sinal químico – o *neurotransmissor glutamato*. Este sinal químico chega até as células pós-sinápticas horizontal e bipolar após uma nova conversão para potencial de membrana. Este processo segue ciclicamente transformando sinal elétrico em químico e novamente em elétrico até que a luz (ou a sua ausência) seja finalmente convertida para uma alteração na freqüência de disparo do potencial de ação das células ganglionares [3]. A ilustração da figura 0.1 apresenta uma versão bastante simplificada deste processo.

É surpreendente, contudo, que para cada campo receptivo de, em média, 125 fotorrecep-

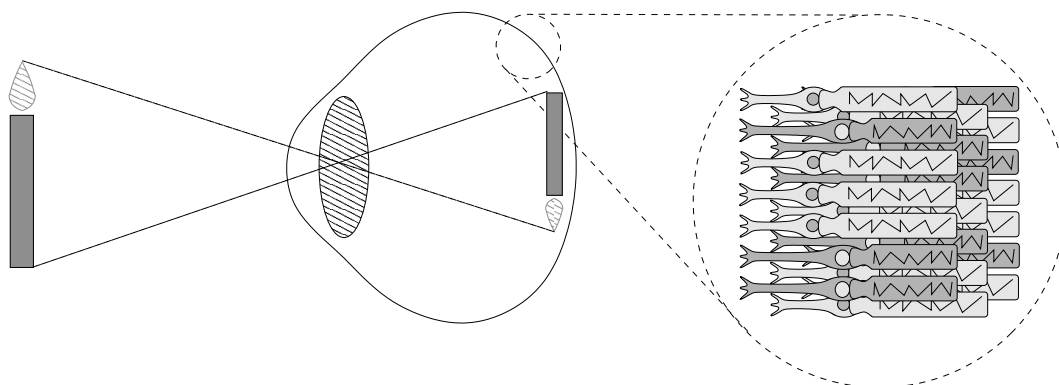


Fig. 1.1: *Esquema simplificado da secção de um globo ocular representando o processo visual primário. À esquerda, a imagem de um objeto orientado é projetada pelo cristalino (a elipse hachurada no centro da qual se dá a intersecção das duas retas) no fundo do olho, onde encontra-se a retina; à direita, uma ampliação da camada mais interna da retina, onde se localizam os fotorreceptores, as células visuais responsáveis pela percepção da luz e das informações contidas nesta. Estas células podem ocupar-se com características distintas da luz, dependendo de sua natureza.*

tores disparando, existe apenas uma célula ganglionar coletando³. Estes números, mais do que nos dizerem que existem muito mais células fotorreceptoras do que ganglionares, revelam uma “preocupação” da natureza em selecionar a informação com a qual trabalha. Se os estímulos captados por 125 células podem ser acomodados em uma única, certamente alguma coisa, senão a maior parte, foi descartada.

O tratamento dado à informação não a torna deficiente. Basta considerarmos que nossas interações com a vizinhança em nada são prejudicadas com a perda das informações descartadas. Certamente algum critério é empregado na seleção do que deve ser mantido e do que é desnecessário.

1.3 Sistemas Artificiais

O cérebro dos mamíferos é muito sofisticado, e sua compreensão passa por diversas etapas. Dentre as muitas atividades nas quais se mostra superior aos “cérebros artificiais”, podemos destacar a habilidade em trabalhar com informações imprecisas ou confusas, assim como a

³ Rigorosamente falando, a matriz de fibras do nervo ótico não é uniforme no mapeamento do espaço visual. A quantidade de fotorreceptores alimentando cada célula ganglionar é relativamente pequena no centro da retina (particularmente na *fóvea*) quando comparada aos milhares de receptores na periferia. Em conseqüência disto há, no “espaço neural”, uma “super-representação” do foco central do espaço visual. Esta especialização assegura grande acuidade na visão central, mas também exige que o olho mova-se para trazer as imagens de interesse diretamente para a fóvea [3].

rápida adaptação a ambientes novos. Seu estudo, entre tantas outras razões, é motivado pelo fato de que o entendimento dos processos neurais responsáveis por certas tarefas permite a criação de modelos mais precisos e melhor preparados para simular e reproduzir resultados com desempenho similar⁴.

O cérebro pode ser visto como um conjunto de milhares de processadores paralelos (os neurônios) que executam simultaneamente um programa muito simples e com baixa dissipação de potência. Isto o torna um sistema dotado de alta robustez, pois mesmo a morte diária de suas células nervosas – os processadores neurais – não necessariamente é acompanhada por perdas de desempenho.

Em vista disso tudo, um computador que seja capaz de trabalhar neurologicamente também deverá, em princípio, estar apto a manifestar muito mais facilmente essas características. A apreensão, o pré-processamento, a análise, o reconhecimento e o armazenamento de imagens ainda carecem de muita pesquisa, embora um esforço bastante grande nos últimos anos vem avançando mais e mais na busca por respostas que nos levem a explicar os mecanismos de funcionamento do sistema visual.

Nas seções que seguem, introduziremos as idéias nas quais estão baseadas as simulações computacionais.

1.3.1 Simulações Computacionais

Podemos obter testes não triviais para diferenciar entre duas formas paralelas de medida de qualidade de processamento: a *informação mútua* e a *técnica de reconstrução*. Ambos os métodos são hoje em dia usados extensivamente para a análise de como informações sobre o meio ambiente são codificadas no sistema nervoso biológico. Para a aplicação destes testes, interessam-nos particularmente as *redes neurais*, aqui tratadas como algoritmos computacionais que buscam “retratar” o comportamento coletivo de um sistema neural biológico.

Este retrato, por assim dizer, implica em simular características que se aproximem tanto quanto possível daquelas encontradas nos neurônios. Os modelos de redes neurais, por exemplo, são dotados da mesma robustez das células nervosas biológicas mencionada na

⁴ As aplicações para este campo são muito vastas e provavelmente até hoje não enumeradas totalmente. É inevitável pensarmos que as ciências médicas veriam uma oportunidade de contar com próteses capazes de substituir com segurança e eficiência partes do organismo humano gravemente lesionadas, e os programas de tecnologia robótica evoluiriam para uma inteligência artificial a partir do momento em que adquirissem a habilidade de processar informações como o cérebro humano.

seção anterior. Além disto, as redes são muito flexíveis podendo ser facilmente reajustadas para novos ambientes a partir de um *reaprendizado*. Elas não necessariamente precisam tratar com informações determinísticas; por realizarem um processo estatístico, lidam eficientemente com *ruído*, com informações *nebulosas*, *probabilísticas* e mesmo *inconsistentes*. E, além de serem pequenas e compactas, também trabalham com processamento *paralelo*, mais dinâmico que o paradigma computacional usual.

1.3.2 Modelo de McCulloch e Pitts

Em um neurônio biológico típico, como pode ser observado na figura 1.2, temos o *corpo da célula* abrigando seu *núcleo* e se ramificando em *dendritos*. Ambos, o corpo da célula e o conjunto de dendritos, contêm *sinapses* de entrada oriundas de outros neurônios. Também do corpo da célula se estende um filamento chamado *axônio* por onde sai a resposta do neurônio alcançando outros através das sinapses [9].

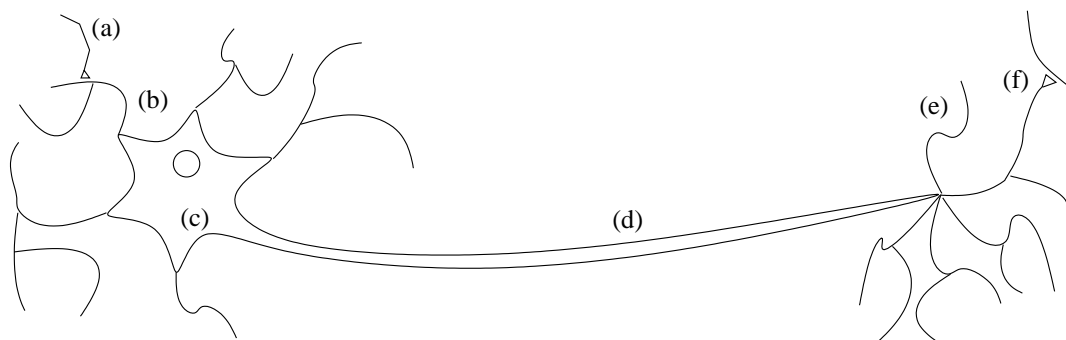


Fig. 1.2: Esquema simplificado de um neurônio biológico típico onde ficam indicadas suas principais partes: os impulsos são recebidos como sinais de entrada de outros neurônios através das sinapses (a) ligadas aos dendritos (b). O corpo da célula (c), onde o núcleo fica abrigado, recebe e processa os sinais captados por todos os dendritos gerando respostas, que são enviadas através do axônio (d) – ao final do qual se localizam outros dendritos (e) – para então serem transmitidas às demais células neurais (f).

A interação entre neurônios para a transmissão de algum sinal é um processo complexo aqui simplificado por um modelo: uma célula recebe estímulo de muitas outras durante o tempo. Os estímulos podem ser inibitórios ou excitatórios. Se o potencial elétrico dos estímulos gerado no corpo da célula alcançar um valor limite, um disparo em forma de pulso ao longo do axônio é efetuado. A este disparo, segue-se o *período refratário* da célula,

quando nenhum outro pulso pode ser produzido.

McCulloch e Pitts [9] utilizaram em 1943 a função de Heaviside $\Theta(x - x_0)$ – que assume valor 1 se $x \geq x_0$ e valor 0 se $x < x_0$ – para criar uma *unidade neural de disparo unitário*. Consiste na soma ponderada de um conjunto binário de entradas n_j que representa os sinais transmitidos por outros neurônios. A ponderação é feita por pesos w_{ij} representando sinapses excitatórias (quando positivos), inibitórias (quando negativos) ou inexistentes (quando nulos). Se o resultado for igual ou superior ao valor limite μ_i (chamado de *limiar*), o neurônio dispara.

$$n_i(t+1) = \Theta\left(\sum_j w_{ij}n_j(t) - \mu_i\right) \quad (1.1)$$

Assim, o estado $n_i(t+1)$ do neurônio i ao tempo $t+1$ pode ser 0 (quiescente⁵) ou 1 (ativo). O tempo t , aqui assumido discreto, é usado para indicar a precedência temporal do valor computado para o neurônio. Trata-se de um modelo não trivial, ainda que omita muitas características típicas em sistemas biológicos⁶.

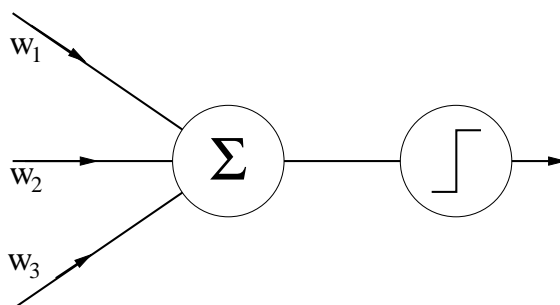


Fig. 1.3: Esquema da unidade neural de disparo unitário proposto em 1943 por McCulloch e Pitts. As conexões w_1, w_2, w_3 , se ativas, são primeiramente somadas no neurônio ao qual se ligam. Na etapa seguinte, o valor resultante é comparado com o limiar μ_i e, se superior, leva ao disparo do neurônio.

É possível romper os limites do espaço discreto dentro do qual se encontra a equação (1.1) escrevendo

$$n_i = f\left(\sum_j w_{ij}n_j - \mu_i\right), \quad (1.2)$$

⁵ Um neurônio é chamado de *quiescente* quando, ainda que não esteja ativo, também não está inativo. A inatividade pressupõe ausência absoluta de disparos mesmo que sob estímulos; um neurônio quiescente, por outro lado, está sujeito a pequenos disparos ocasionais mesmo sem receber estímulos.

⁶ Entre essas características, sabemos que eles têm respostas contínuas, alguns promovem somas não lineares de suas entradas, normalmente trabalham fora de sincronia uns dos outros, entre outras coisas.

onde n tem valores contínuos, e $f(\cdot)$ é uma *função de ativação* qualquer com características mais gerais. A função de ativação será discutida com mais detalhes na seção 2.1.

Um algoritmo montado com centenas de unidades que funcionam segundo a lei de McCulloch e Pitts oferece um enorme contraste quando comparado com o sistema tradicional idealizado por von Neumann [9]. São muitos processadores executando um programa muito simples cada, ao invés de alguns poucos processadores rodando programas pesados.

Capítulo 2

Redes Neurais

Neste capítulo, abordaremos conceitos sobre redes neurais necessários para a compreensão do trabalho que desenvolvemos. Discutiremos modelos de redes, as arquiteturas utilizadas e os métodos de aprendizado.

2.1 Modelos de Redes Neurais

Existem diversos modelos de rede neurais, distinguindo-se uns dos outros a partir de três fatores principais: sua *arquitetura*, sua *função de ativação*¹ e sua *regra de aprendizagem*.

A arquitetura determina o *número de camadas* da rede, a *dimensão dos vetores* e o tipo de *conexões sinápticas* empregadas. O modelo de arquitetura mais simples é composto tão somente pelo *vetor entrada*, responsável pela alimentação da rede com informações do ambiente, e um vetor composto por uma ou mais *unidades de saída*, caminho através do qual a rede responde ao meio externo. Por conveniência, adotamos a convenção que desconsidera o vetor entrada no cômputo do número de camadas da rede, o que torna o exemplo acima uma rede de camada única. Na figura 2.1, a arquitetura à esquerda é um exemplo de rede neural de uma camada.

Se forem introduzidos um ou mais vetores entre o vetor entrada e as unidades de saída (conferir o exemplo à direita na figura 2.1), a arquitetura da rede passa a ser multicamadas, potencialmente mais apta a realizar tarefas de complexidade superior. Estas camadas são normalmente referidas como *camadas ocultas*, pois não se comunicam com o ambiente. A dimensão destes vetores (vetor entrada, camadas ocultas e unidades de saída) depende da natureza dos dados com o qual se pretende trabalhar e do tratamento que eles receberão

¹ Outros nomes também estão disponíveis na literatura, como *função de transferência* ou *função ganho*, por exemplo.

quando submetidos à rede.

Uma vez enviado pelo vetor entrada, o sinal navega entre as camadas da rede percorrendo as vias sinápticas, a rede de comunicação que liga as componentes de um vetor com as componentes de outro vetor ou do mesmo. As formas desta ligação são flexíveis e estão intrinsecamente relacionadas com o objetivo para o qual a rede é projetada.

A função de ativação, por sua vez, é dada pela expressão que determina a lei matemática empregada na ativação ou disparo de um neurônio. Da natureza, sabemos que o disparo de um neurônio obedece regras bastante complexas que dependem não somente dos sinais de entrada que recebe de outras células a partir das conexões sinápticas, mas também (e entre muitas outras coisas) de sua própria estrutura neuro-química interna. A função de ativação é, rudimentarmente falando, o formalismo que se propõe a simular algebricamente a condição fisiológica da célula neural que a impele a disparar ou não.

Determinadas a arquitetura e a função de ativação da rede neural, faz-se necessário treiná-la a fim de que aprenda como executar sua tarefa. Seleciona-se, pois, uma lei matemática que determinará o processo com o qual as conexões sinápticas – daqui para adiante referidas como *vetores pesos* – terão seus valores alterados. A esta lei chamamos de *regra de aprendizagem*. Estas características fornecem uma *identidade* à rede neural.

O processo de treinamento consiste em reunir uma população de dados (os *padrões de entrada*) e submetê-los um após outro à rede segundo o algoritmo ilustrativo que segue:

1. Padrões submetidos individualmente através do vetor entrada;
2. Ponderação pelo(s) vetor(es) peso(s) e aplicação à função de ativação;
3. A saída é verificada (*aprendizado supervisionado*) ou não (*aprendizado não-supervisionado*);
4. O(s) vetor(es) peso(s) é(são) atualizado(s) segundo a regra de aprendizagem;
5. Retorna-se ao item 1, caso o critério de interrupção não tenha sido atingido.

Uma rede neural, quando treinada desta forma, torna-se capaz de desempenhar tarefas, tais como *determinação de familiaridade, codificação, mapeamento de características*, etc...

2.1.1 Percéptrons Simples

Em 1962, Rosenblatt batizou de *percéptrons* os modelos de redes *multicamadas feed-forward*, ou seja, redes cujo fluxo é unidirecional correndo somente no sentido dos terminais de entrada para a camada de saída. Estas redes, que representam muito bem as estruturas neurais reais, consistem fundamentalmente em uma camada inicial formada por terminais que alimentam uma camada seguinte que pode ser ou não seguida de camadas adicionais. A última camada oferece a saída da rede, isto é, o resultado da operação realizada pelos neurônios. As camadas intermediárias (se houver) entre a camada de entrada e a de saída também são consideradas camadas ocultas por não possuírem comunicação externa. As unidades que compõem cada camada não retornam para as camadas anteriores e não se comunicam com unidades da mesma camada ou com aquelas que vêm após a seguinte.

Dois exemplos podem ser vistos na figura 2.1.

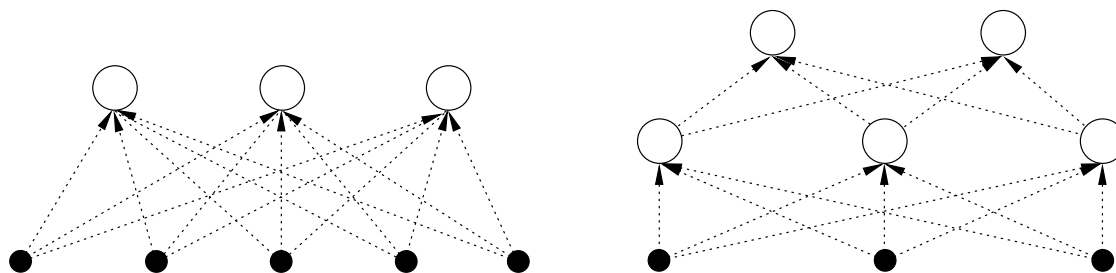


Fig. 2.1: Percéptrons: os círculos cheios correspondem aos terminais de alimentação da rede e não contribuem no processamento, de modo que não são incluídos na contagem das camadas. Os vazados são as unidades onde as informações são processadas. Do lado esquerdo, uma rede de percéptrons simples (somente uma camada); do lado direito, uma rede de percéptrons multicamadas. A primeira camada está oculta por não ter ligação com o meio externo.

Restringiremos nosso estudo ao percéptron simples, que contém somente uma camada e é alimentado por um conjunto de N entradas.

Percéptrons Simples Lineares

Seja uma rede neural descrita pelas características que seguem:

- camada simples (primeira arquitetura da figura 2.1);
- *feed-forward*;

- linear: função de ativação do tipo $f(x) = ax + b$ (a, b constantes);

Uma rede que obedece estas determinações é chamada de *perceptron simples linear*. Para uma arquitetura com N entradas e uma saída, tem-se a forma

$$S = \sum_{i=1}^N w_i e_i = w_1 e_1 + w_2 e_2 + \cdots + w_N e_N = \vec{w} \cdot \vec{e}, \quad (2.1)$$

onde w_i é a i -ésima componente do vetor peso \vec{w} correspondendo à respectiva componente do vetor entrada \vec{e} . Observe que a resposta S da rede linear, o produto interno entre o vetor de entrada \vec{e} e o vetor sináptico \vec{w} , nada mais é do que uma grandeza escalar que representa a *familiaridade* do padrão de entrada.

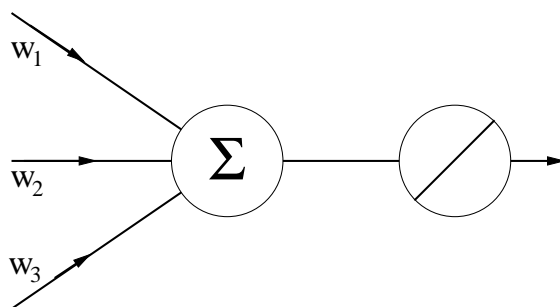


Fig. 2.2: Esquema da unidade neural de disparo unitário com função de ativação do tipo linear. As componentes do vetor entrada são ponderadas pelo vetor peso e somadas em um escalar antes de servirem de parâmetro para a função de ativação.

2.1.2 Aprendizados de Redes Neurais

A fim de que possamos utilizar uma rede neural com o intuito de realizar alguma tarefa, aplicaremos algoritmos que nos permitam “ensinar” a rede dentro dos moldes da necessidade. Uma rede é capaz de desempenhar uma tarefa quando utiliza pesos adequados que a faz convergir para o resultado esperado. Partindo de valores arbitrários para os pesos, o aprendizado consiste em realizar uma série de iterações que os levarão a tenderem para os valores apropriados. Frequentemente, esta iteração é o procedimento mais simples para atingir-se o valor adequado das conexões; de outro modo, a escolha dos valores pode envolver um grande problema de otimização ou uma inversão matricial consideravelmente complexa.

Aprendizado Supervisionado

O aprendizado supervisionado utiliza-se de *conjuntos entrada-alvo* $(\vec{e}^\tau, \vec{R}^\tau)$ para treinamento. A entrada é submetida à rede, e a saída resultante \vec{S}^τ é comparada com o respectivo alvo \vec{R}^τ do conjunto, que indicará se os pesos devem ou não ser reajustados. A unidade de saída S_i^τ é dada por

$$S_i^\tau = f\left(\sum_{k=0}^N w_{ik} e_k^\tau\right) = f\left(\sum_{k=1}^N w_{ik} e_k^\tau - \mu_i\right), \quad (2.2)$$

onde e_k^τ e w_{ik} são respectivamente o elemento k da entrada e o elemento k relacionado à unidade S_i^τ da matriz de pesos. A função de ativação $f(\cdot)$ depende explicitamente da entrada (propriedade de redes *feed-forward*) e normalmente não é linear. A forma mais à direita em (2.2) é obtida quando definimos $e_0^\tau = -1$ e $w_{i0} = \mu_i$.

Para um conjunto de P padrões de entrada-alvo para treinamento, utilizamos o índice τ para identificar qual padrão está sendo aplicado, de tal forma que $\tau = 1, 2, \dots, P$. Quando a entrada \vec{e}^τ passa pela rede, resulta na saída \vec{S}^τ , que por sua vez é comparada com o alvo \vec{R}^τ – a resposta desejada pelo padrão. Esta atividade – buscar $S_i^\tau = R_i^\tau$ para cada i e τ – recebe o nome de *tarefa de associação geral*. O aprendizado então consiste em aplicar um padrão \vec{e}^τ , comparar a saída \vec{S}^τ com o alvo \vec{R}^τ e alterar \vec{w} se o resultado não for o desejado. Matematicamente,

$$\vec{w}^\tau = \vec{w}^{\tau-1} + \Delta \vec{w}^\tau, \quad (2.3)$$

onde $\Delta \vec{w}^\tau$ é escolhido de acordo com o tipo de função de ativação $f(\cdot)$ da rede. A equação (2.3) nos mostra que, apesar de τ apenas indicar o padrão aplicado à iteração, também serve para estabelecer uma ordem cronológica no processo.

Após o treinamento, é conveniente testar o desempenho da rede com outros modelos de entrada visando a averiguar se ela será capaz de generalizar o que aprendeu. As entradas, saídas e alvos podem assumir valores booleanos ou contínuos. Dependendo da necessidade, pode-se usar um tipo diferente de valor para cada uma das três entidades.

Aprendizado Não-Supervisionado

Muitas vezes não é possível dispor de um conjunto de treinamento que prepare uma rede para dada tarefa. Uma vez, contudo, que exista uma correlação entre os dados de entrada,

um método de *aprendizado não-supervisionado* torna-se viável. Neste caso, espera-se que a rede crie categorias a partir destas correlações e produza sinais de saída correspondendo às categorias de entrada. Neste trabalho, nos centraremos no *aprendizado não-supervisionado*, tendo que o sistema visual se utiliza da estatística dos estímulos ambientais para a formação do conjunto sináptico.

No aprendizado não-supervisionado, somente as entradas e saídas da rede são levadas em consideração, isto é, não existe um retorno dado pelo ambiente orientando o aprendizado da rede. O aprendizado, então, depende de ela descobrir padrões, características e correlações que advêm da existência de *redundância* nas entradas. Segundo Barlow [2],

“*Redundância proporciona conhecimento*”.

A natureza deste aprendizado depende essencialmente da arquitetura da rede. Dentre as características que podem ser extraídas, destacamos: *familiaridade* (detecta a similaridade de novos padrões de entrada), *análise de componentes principais* (é uma extensão do caso anterior para várias unidades), *codificação* (a saída pode ser uma versão codificada da entrada, em pedaços menores, mantendo tanta informação relevante quanto possível), entre outras.

Um modelo de rede neural pode ser construído com muitas arquiteturas diferentes e obedecer diversas leis de aprendizagem de acordo com a tarefa que lhe será imposta. Em sua maior parte, a arquitetura de redes com aprendizado não-supervisionado é bastante simples; as complicações surgem predominantemente das regras de aprendizagem. Na lei de aprendizagem hebbiana para uma rede com uma única unidade de saída S , por exemplo, a correção no peso é feita a partir de

$$\Delta \vec{w} = \eta S \vec{e}, \quad (2.4)$$

onde η controla a taxa de aprendizagem. Esta regra é chamada de *Regra de Hebb* devido à hipótese de Donald Hebb que, em 1949, propôs que os pesos sinápticos no cérebro mudam em resposta a experiências. Hebb sugeriu mudanças proporcionais à correlação entre os disparos dos neurônios pré e pós-sinápticos [29].

O aprendizado hebbiano ensina o neurônio a executar sua tarefa quando este é submetido a um padrão de treinamento apropriado. A regra (2.4) altera o valor das conexões sinápticas incrementando-as com o vetor $\Delta \vec{w}$ paralelo ao vetor entrada \vec{e} . Esta expressão, contudo,

não possui um critério de parada, de modo que a rede não tem conhecimento de quando seu aprendizado foi suficiente a fim de interromper o processo. Assim, os pesos, sujeitos à iteração dada em (2.3), crescem indefinidamente podendo divergir para o infinito.

Na seção 2.2, veremos como isto pode ser evitado aplicando uma regra de Hebb modificada [9] com valores contínuos para as unidades de saída. O objetivo principal será, ou a medição da familiaridade, ou a projeção sobre as componentes principais dos dados de entrada. Para tanto, consideraremos apenas percéptrons simples com muito menos unidades de saída do que de entrada.

Regras de Aprendizado Local e Não-Local

Não se pode esperar que um neurônio biológico realize interações com outro neurônio sem que haja uma sinapse para promover a troca de sinais. Da mesma forma, a coerência nos faz crer que um neurônio pertencente a um cérebro real não poderá dispor de informações de neurônios não ligados a ele ou tampouco enviá-las para estas mesmas células.

Um modelo matemático de rede neural que associe a um padrão de entrada \vec{e}^τ múltiplas unidades de saída \vec{S}^τ , por outro lado, pode apresentar tanto *regras locais* de aprendizagem quanto *regras não-locais*.

A regra de aprendizagem que modifica uma componente de peso w_{ij} será *local* quando, analogamente ao sistema biológico, o aprendizado da rede depender exclusivamente do que está disponível no elemento de entrada e_j e no elemento de saída S_i . Este é o caso, por exemplo, da regra de Hebb (2.4) e da regra de Oja para uma unidade (2.12), ambas abordadas na referência [9].

Não obstante, alguns ganhos no desempenho de uma tarefa podem ser obtidos se implementada uma rede com aprendizado *não-local*, ainda que isto implique em um distanciamento de sistemas biológicos. A análise de componentes principais (seção 2.2) é mais facilmente computada quando aplicada a regra de Sanger (2.14) ou a de Oja (2.15) para múltiplas unidades (também tratadas em [9]). Em ambas, a atualização do peso w_{ij} exige mais informação do que está disponível em e_j e em S_i .

2.1.3 Aprendizagem com Retropropagação de Erro (*Back-propagation*)

A retropropagação de erro é um dos métodos de *treinamento supervisionado* que pode ser empregado para ensinar uma rede multicamada. O algoritmo orienta a alteração dos vetores sinápticos \vec{w} empregando um conjunto de treinamento composto por pares entrada-saída $(\vec{e}^\tau, \vec{S}^\tau)$ fazendo uso do método conhecido como *gradiente descendente*.

Seja um *perceptron* com duas camadas (uma camada oculta e uma de saída) tal que e_k^τ representa a componente k do padrão τ de um conjunto de P padrões. Sejam também w_{jk} a matriz que conecta as entradas com a camada oculta O_j^τ , e W_{ij} a matriz que conecta a camada oculta com a camada de saída S_i^τ .

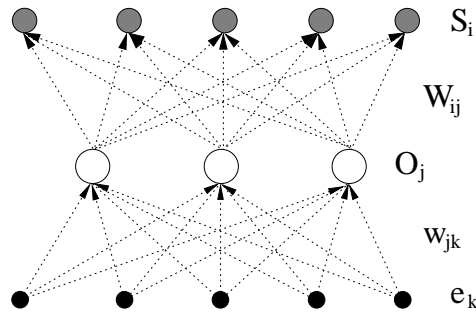


Fig. 2.3: Arquitetura de uma rede com duas camadas para treinamento baseado em retropropagação de erro. Os círculos brancos representam a camada oculta, que não mantém ligação direta com o ambiente externo; os círculos cinzentos representam a camada de saída, que tem a mesma dimensão dos terminais de entrada tendo em vista a aplicação da rede.

Em uma rede linear, a função de ativação $f(\cdot)$ para as unidades da camada oculta de um dado padrão τ será representada como

$$O_j^\tau = f\left(\sum_k w_{jk}e_k^\tau\right) = \sum_k w_{jk}e_k^\tau, \quad (2.5)$$

enquanto a camada de saída, que herdará como entrada a saída da camada oculta, ficará

$$S_i^\tau = f\left(\sum_j W_{ij}O_j^\tau\right) = \sum_j W_{ij} \sum_k w_{jk}e_k^\tau. \quad (2.6)$$

O problema pode ser otimizado de forma bastante simples definindo uma *função custo* $H(\vec{w})$ cuja minimização fornece o melhor desempenho da rede, ou seja,

$$H(\vec{w}) = \frac{1}{2} \sum_{\tau i} (R_i^\tau - S_i^\tau)^2 = \frac{1}{2} \sum_{\tau i} \left(R_i^\tau - \sum_j W_{ij} \sum_k w_{jk}e_k^\tau \right)^2. \quad (2.7)$$

A diferença entre o resultado desejado \vec{R}^τ e o obtido \vec{S}^τ é condição do treinamento com supervisão, ao passo que a lei quadrática garante que haverá somente um mínimo global. Tendo em vista que as funções de ativação são contínuas e diferenciáveis, esta função também será contínua e diferenciável para cada peso². Isto, por sua vez, torna possível usar o método do gradiente descendente como forma de minimizar o custo e, portanto, aproximar o obtido do desejado. Por este método, a atualização dos pesos entre a camada oculta e a camada de saída se dá pela expressão

$$\Delta W_{ij} = -\eta \frac{\partial H}{\partial W_{ij}} = \eta \sum_{\tau} (R_i^\tau - S_i^\tau) O_j^\tau, \quad (2.8)$$

e a atualização dos pesos entre a camada de entrada e a camada oculta é feita por

$$\Delta w_{jk} = -\eta \frac{\partial H}{\partial w_{jk}} = \eta \sum_{\tau i} (R_i^\tau - S_i^\tau) W_{ij} e_k^\tau. \quad (2.9)$$

Normalmente, a soma sobre todos os padrões τ não é realizada completamente antes da atualização dos pesos, como é mostrada pelas regras de aprendizagem acima. Ao invés disto, emprega-se um procedimento incremental: o padrão τ é apresentado à entrada, e então todos os pesos são calculados antes de ser considerado o próximo padrão – o que decresce a função custo para η suficientemente pequeno. Seguindo este método, obter-se-á melhor desempenho na maior parte dos casos, especialmente para conjuntos de treinamento muito regulares ou redundantes.

A codificação e a compressão de imagens são possíveis aplicações do método de retropropagação de erros. A idéia consiste em encontrar um número M de unidades ocultas tal que $M < N$ force a redução dimensional de um grande número de pares entrada-saída. Para tanto, empregam-se redes de duas camadas e adotam-se entradas e saídas unárias e idênticas.

Um aspecto interessante é que a não-linearidade das unidades ocultas de uma rede teoricamente não oferece vantagem ao processo de codificação e compressão de imagens. Aplicando unidades lineares, obtemos uma análise teórica detalhada que mostra que a rede projeta a

² É de hábito utilizar-se *funções sigmóides* como função de ativação por serem diferenciáveis e saturáveis em ambos os extremos. Por exemplo:

$$f(x) = \frac{1}{1 + e^{-x}}$$

entrada sobre o subespaço abarcado pelas M primeiras componentes principais da entrada, tal que M ainda é o número de unidades ocultas [9].

Por convenção, redes com N entradas, M unidades ocultas e N saídas são citadas como *codificadores N - M - N* .

2.2 Análise de Componentes Principais (PCA)

Neste trabalho, nos focaremos particularmente na *codificação* e na *análise de componentes principais* (PCA)³. A primeira se dá quando os dados de saída se tornam uma versão codificada dos padrões de entrada, ocupando uma porção de memória muito menos expressiva e conservando tanta informação relevante quanto possível – como foi discutido na seção 2.1.3. A segunda, por sua vez, é realizada por uma rede neural quando seus vetores pesos constroem uma base com múltiplos eixos em um subespaço do espaço dos padrões de entrada, base esta que mede a similaridade dos exemplos.

A PCA tem por objetivo reduzir a dimensão de um espaço N -dimensional para um M -dimensional de forma a determinar M vetores ortogonais que estejam relacionados com a variância das entradas. Dependendo da estatística dos dados, tem-se $M \ll N$, o que significa que obtemos uma redução dimensional expressiva. É uma ferramenta muito utilizada para análises estatísticas, principalmente nas situações em que a população sob análise possui distribuição gaussiana.

Matematicamente, o efeito de uma PCA equivale a tomar-se a primeira componente principal ao longo da direção de máxima variância da distribuição; a segunda, contida no subespaço perpendicular à primeira, é tomada ao longo da direção de maior variância dentro daquele subespaço; a terceira, contida no subespaço perpendicular às duas primeiras, é tomada ao longo da direção de maior variância dentro deste novo subespaço. E assim sucessivamente. Em consequência disto, a projeção do espaço N sobre o M freqüentemente retém a maior parte das informações significativas dos dados. Um exemplo matemático simplificado é ilustrado no apêndice A.

Pode ser mostrado [9] que a k -ésima componente principal está ao longo do autovetor

³ Do inglês, *principal components analysis*, também conhecida por *transformação de Karhunen-Loève*. Está muito relacionado com outros métodos, tais como o de mínimos quadrados.

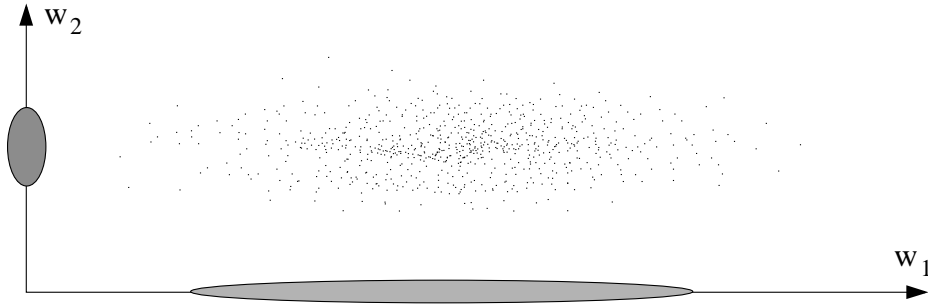


Fig. 2.4: *Projeção das duas componentes principais em um espaço amostral bidimensional. A distribuição gaussiana de padrões (pontos) apresenta maior variância ao longo da dimensão horizontal, como pode ser visto pela estreita elipse hachurada sobre o eixo. Sendo assim, a primeira componente principal segue a orientação de w_1 . A elipse hachurada mais arredondada sobre o eixo vertical indica a menor variância e, por conseguinte, indica também que a segunda componente principal corre ao longo de w_2 .*

relacionado ao autovalor de k -ésima ordem de grandeza da *matriz covariância* completa

$$C_{ij} = \langle (e_i - \langle e_i \rangle)(e_j - \langle e_j \rangle) \rangle. \quad (2.10)$$

No caso particular em que os dados estão centrados (têm média zero), a equação (2.10) se reduz à (2.11) a seguir, indicando que as componentes principais se reduzem aos autovetores da matriz correlação \mathbf{C} .

$$C_{ij} = \langle e_i e_j \rangle, \quad (2.11)$$

onde definimos $\langle \cdot \rangle$ como a média sobre a distribuição de entrada $\mathcal{P}(\vec{e})$. \mathbf{C} é simétrica ($C_{ij} = C_{ji}$), tem autovalores reais, positivos e não nulos e autovetores ortogonais.

A regra de aprendizado não-supervisionado de Hebb vista na equação (2.4) não oferece à rede um critério para que ela saiba quando deve parar de alterar o valor do vetor sináptico. Como consequência, mesmo após este ter sido orientado para a direção desejada, o incremento de sua norma não é interrompido, e esta cresce indefinidamente, como está ilustrado na figura 2.5 (A).

Oja propôs uma regra de aprendizagem que não somente elimina a divergência do peso, mas também o normaliza [9]. As sucessivas iterações corrigem sua orientação e o aproximam do *autovetor máximo* da *matriz correlação* \mathbf{C} apresentada em (2.11). Aqui, chamamos de autovetor máximo o autovetor correspondente ao maior autovalor.

A regra de Oja – cuja representação gráfica pode ser vista na figura 2.5 (B) – consiste na

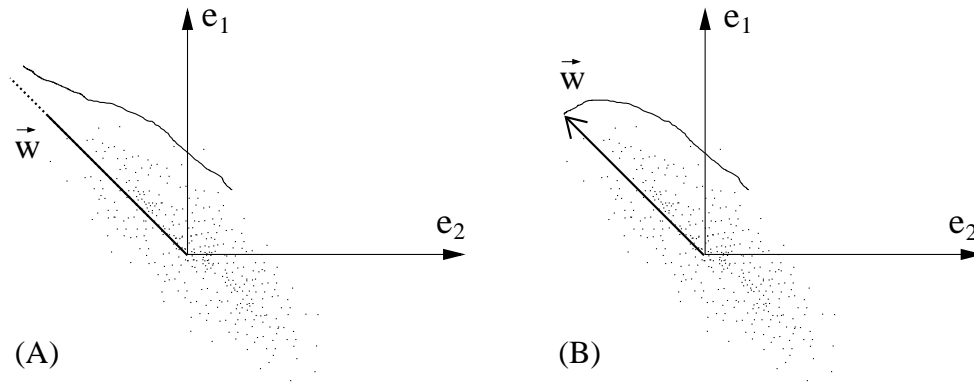


Fig. 2.5: Resultado da aplicação em rede linear de entradas bidimensionais $\vec{e} = (e_1, e_2)$ distribuídas de acordo com os pontos da figura. A evolução da orientação do vetor peso \vec{w} é vista seguindo a linha mais fina: (A) percebe-se que \vec{w} atinge a orientação desejada, mas segue crescendo devido à inexistência na regra de Hebb de uma condição de parada; (B) a regra de Oja (a regra de Hebb modificada) encarrega-se de manter \vec{w} com tamanho finito uma vez que a orientação seja atingida. (Adaptado de Hertz et al [9].)

alteração da regra de Hebb pela subtração de um termo proporcional ao quadrado da saída S , ou seja,

$$\Delta w_i = \eta S(e_i - S w_i). \tag{2.12}$$

Uma vez que a regra de Oja seja implementada, e a rede trabalhe com um conjunto de entradas, a saída S será somente a componente da entrada \vec{e} ao longo da direção de \vec{w} . Em outras palavras, verifica-se que a rede desenvolve um índice de familiaridade para a distribuição como um todo e não necessariamente para qualquer exemplo \vec{e} em particular, o que nos leva a inferir que esta regra permite um estudo estatístico do comportamento da rede ao invés de somente um exemplo isolado.

De fato, a regra de Oja escolhe a direção de \vec{w} que maximize $\langle S^2 \rangle$. Para dados centrados, isto equivale a dizer que a regra maximiza a variância da saída – o que, por sua vez, corresponde à extração da primeira componente principal.

A extração das M primeiras componentes principais foi obtida por Sanger e Oja [9] a partir da ampliação da rede linear apresentada em (2.1) desenvolvendo as redes de uma camada, lineares e *feed-forward* dadas por

$$S_a = \sum_i^N w_{ai} e_i = \vec{w}_a \cdot \vec{e}, \tag{2.13}$$

onde \vec{w}_a é o vetor peso para cada uma das unidades S_a , com $a = 1, 2, \dots, M$.

A regra de aprendizado de Sanger para múltiplas unidades, modificada da regra de Oja para uma unidade dada pela equação (2.12), é

$$\Delta w_{ai} = \eta S_a \left(e_i - \sum_{b=1}^a S_b w_{bi} \right). \quad (2.14)$$

Em outras palavras, serão subtraídos da expressão do aprendizado hebbiano $\Delta w_{ai} = \eta S_a e_i$ tantos termos quantas forem as unidades de saída em consideração. Então, se observarmos a unidade S_3 (com $M > 3$), veremos que o aprendizado de seu vetor peso \vec{w}_3 depende das saídas anteriores (S_1 e S_2), do quadrado dela mesma (S_3^2) e dos respectivos vetores pesos (\vec{w}_1 , \vec{w}_2 e \vec{w}_3) – o que a torna uma *regra de aprendizagem não-local*. Os vetores pesos convergem para as M primeiras componentes principais sempre tendendo aos autovetores normalizados da matriz correlação \mathbf{C} . Suas saídas são estatisticamente descorrelacionadas, e suas variâncias decrescem com o incremento de i . Estas variâncias, que são os autovalores correspondentes, podem ser vistas como uma medida da utilidade da saída.

Segundo Hertz *et al* [9],

“A regra de Sanger é normalmente mais útil nas aplicações porque extrai as componentes principais individualmente e em ordem, além de fornecer resultados que podem ser reproduzidos (...) para um determinado conjunto de dados se os autovalores forem não degenerados. Realiza exatamente a transformação de Karhunen-Loève”.

A regra de aprendizado de Oja para múltiplas unidades, por sua vez, também é derivada de (2.12) e tem a forma

$$\Delta w_{ai} = \eta S_a \left(e_i - \sum_{b=1}^M S_b w_{bi} \right). \quad (2.15)$$

Neste caso, serão subtraídos da expressão do aprendizado hebbiano *todas* as M unidades de saída – o que também a torna uma *regra não-local*. Os M vetores pesos convergem para o mesmo subespaço dos M autovetores de \mathbf{C} , mas não necessariamente se orientam segundo as direções destes. As orientações podem ser várias, diferindo a cada tentativa dependendo das condições iniciais e das particularidades dos dados empregados na aprendizagem.

Em ambos os casos, os vetores \vec{w}_a convergem para vetores ortonormais, assim como as saídas projetam o padrão de entrada \vec{e} sobre o subespaço das M primeiras componentes

principais. Apesar disto, um cérebro real teria mais probabilidade de comportar-se segundo a regra de Oja do que a de Sanger. Sobre isto, Hertz *et al* comentam:

“(...) se qualquer algoritmo deste tipo fosse implementado em cérebros reais, provavelmente se pareceria mais com o de Oja do que com o de Sanger: não existe vantagem óbvia para um animal em ter informação ajustada segundo componentes principais individuais”.

Sanger sugeriu uma reformulação da sua regra que permite uma implementação local; em contrapartida, traz consigo o aumento de complexidade da rede [9]. Para tanto, evidenciou o termo do somatório em (2.14) correspondendo ao índice $b = a$

$$\Delta w_{ai} = \eta S_a \left(\left[e_i - \sum_{b=1}^{a-1} S_b w_{bi} \right] - S_a w_{ai} \right) = \eta S_a (e'_i - S_a w_{ai}). \quad (2.16)$$

É interessante notar que a expressão mais à direita é equivalente à regra de Oja para uma unidade de acordo com a equação (2.12) e surge quando operamos a transformação $e'_i = e_i - \sum_{b=1}^{a-1} S_b w_{bi}$. Nesta nova forma, a regra de Sanger (2.16) fica reduzida a uma *regra de aprendizagem local*.

2.3 Campos Receptivos Restritos

Como comentamos anteriormente, regras de aprendizagem locais como Oja (2.12) podem realizar reduções dimensionais de conjuntos de dados de entrada se aproveitando da redundância devido à correlação das entradas. Para uma unidade simples, estes modelos levam a uma redução ótima. Se a rede, por outro lado, for composta por mais de uma unidade, e insistirmos em usar tais regras, o resultado será uma representação altamente redundante constituída de componentes principais iguais entre si – um resultado evidentemente indesejado. Regras mais complexas como (2.14) ou (2.15), que visam a reduzir ou a remover esta redundância, podem desenvolver ótimas representações de componente principal, pagando o preço de serem não-locais e, portanto, pouco realistas do ponto de vista biológico.

Como foi comentado na seção 1.2, as células fotorreceptoras presentes na retina representam a extremidade de entrada do estímulo visual, enquanto as ganglionares representam

a saída da retina para a alimentação do cérebro. Se olharmos para as células ganglionares, veremos que cada uma delas coleta uma fração da imagem projetada sobre as células fotorreceptoras. Em outras palavras, cada célula toma individualmente um bloco que representa uma fração da imagem – fração esta que passaremos a chamar de *campo receptivo*. Dentro deste contexto, a primeira etapa do processamento visual pode ser simplificada em um modelo artificial descrito por redes de neurônios.

Para estudar os efeitos dos campos receptivos, Idiart *et al* [12] tomaram redes lineares, sujeitas ao aprendizado hebbiano modificado, que consistem em N entradas distribuídas em M campos receptivos com r elementos cada. Um modelo esquemático pode ser visto na figura 2.6.

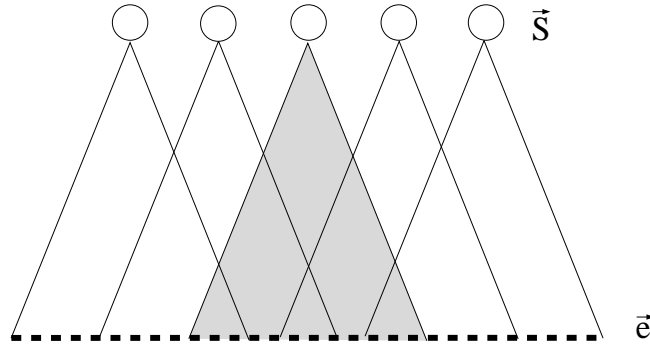


Fig. 2.6: Modelo de arquitetura de rede neural com campos receptivos restritos composta por um vetor entrada \vec{e} N -dimensional e um vetor com M unidades de saída \vec{S} . Cada uma das unidades de saída abarca um campo receptivo contendo r elementos do vetor entrada.

Tomando a rede linear⁴ para múltiplas unidades de saída apresentada na equação (2.13) e acrescentando um ruído ρ_a , tem-se

$$S^a = \sum_{i=1}^N w_{ai}e_i + \rho_a = \vec{w}_a \cdot \vec{e} + \rho_a, \quad (2.17)$$

tal que o ruído tem variância $\sigma^2 = \langle \rho_a^2 \rangle$, é centrado e é descorrelacionado, ou seja, se $a \neq b$, então $\langle \rho_a \rho_b \rangle = 0$.

Como a matriz correlação, definida em (2.11), é uma matriz real e simétrica – e por conseguinte hermitiana –, com autovetores e autovalores reais, sua equação de autovalores pode ser escrita na forma

$$\mathbf{C}\vec{v}_a = \lambda_a\vec{v}_a, \quad (2.18)$$

⁴ A partir da equação (2.17), a notação de S_a será mudada para S^a para facilitar a organização dos índices nas fórmulas.

onde $a = 1, 2, \dots, M$ e a definição da regra de ordenamento $\lambda_1 > \lambda_2 > \dots > \lambda_M$ ⁵. Convém lembrar que \vec{v}_a , o autovetor de \mathbf{C} correspondente ao autovalor λ_a , é também a a -ésima componente principal do espaço de entrada com distribuição $\mathcal{P}(\vec{e})$ para a qual tende o vetor peso \vec{w}_a no processo de aprendizagem da rede. As saídas, que formam a representação reduzida dos N elementos de entrada, consistem em M unidades S^a , tal que $N > M$.

Um possível método de avaliar quantitativamente o processo de redução de informações consiste em reconstruir o conjunto completo de dados do espaço de dimensão N a partir das M saídas e compará-lo com o original. Tendo que a redução é feita por uma rede linear, a reconstrução também será feita pela consideração de uma estimativa linear otimizada \tilde{e} do vetor \vec{e} , isto é, o vetor dado pela expressão

$$\tilde{e} \equiv \sum_{a=1}^M S^a \vec{D}^a = S^1 \vec{D}^1 + S^2 \vec{D}^2 + \dots + S^M \vec{D}^M, \quad (2.19)$$

com \vec{D}^a apropriadamente escolhido de forma a minimizar o *erro quadrado médio* definido como

$$E = \frac{1}{2} \left\langle \left(\frac{\vec{e}}{|\vec{e}|} - \frac{\tilde{e}}{|\tilde{e}|} \right)^2 \right\rangle. \quad (2.20)$$

A grandeza aqui definida representa a medida mais básica da qualidade de reconstrução, pois quantifica a similaridade da imagem reconstruída em relação à original. Este processo, cuja operação consiste fundamentalmente nos *mínimos quadrados*, será tanto melhor sucedido quanto menor for a diferença ponto a ponto entre \vec{e} e \tilde{e} . Os módulos dos respectivos vetores são introduzidos como forma de mantê-los na mesma escala.

No trabalho de Idiart *et al*, é assumido, por hipótese, que $|\vec{e}| = |\tilde{e}|$. Esta premissa, associada à propriedade de *auto-mediância*⁶ das distribuições gaussianas, permite que a equação (2.20) seja reescrita na forma simplificada de (2.21)

$$E = \frac{\langle |\vec{e} - \tilde{e}|^2 \rangle}{\langle |\vec{e}|^2 \rangle}. \quad (2.21)$$

⁵ Rígorosamente falando, uma matriz \mathcal{M} com dimensão $n \times n$ possui n autovetores correspondentes a n autovalores. Para o formalismo das redes neurais abordado aqui, contudo, nos importa somente o número de autovetores e autovalores igual ao número de unidades de saída.

⁶ Um sistema será *auto-mediante* quando satisfizer

$$|\vec{e}| = \sum_i^N e_i^2 = N \langle e_i \rangle.$$

Inserindo (2.19) em (2.21), minimizamos o erro fazendo $\partial E/\partial D^a = 0$, que nos levará a

$$\vec{D}^a = \sum_{b=1}^M [Q^{-1}]^{ab} \vec{L}^b \quad (2.22)$$

para a qual definimos

$$\vec{L}^a = \langle \vec{e} S^a \rangle \quad \text{e} \quad Q^{ab} = \langle S^a S^b \rangle. \quad (2.23)$$

Portanto, a forma analítica do erro mínimo de reconstrução de um sinal \vec{e} para o caso particular de um sistema com auto-mediância é dado por

$$E_{\min} = 1 - \frac{\sum_{a=1}^M \vec{D}^a \cdot \vec{L}^a}{\sum_{i=1}^N \lambda_i}. \quad (2.24)$$

O detalhamento deste cálculo está disponível no apêndice B.

Padrões de Imagens

Com o acréscimo de mais um índice, expandimos o modelo da rede de modo a torná-la capaz de trabalhar com a dimensão de imagens bidimensionais. Sendo assim, seja um *perceptron* simples, com função de ativação linear, matriz de entrada $N \times N$ e M campos receptivos (um para cada unidade de saída)

$$S^{a\tau} = \sum_{i,j} w_{ij}^a I_{ij}^{a\tau}. \quad (2.25)$$

Analogamente aos modelos anteriores, w_{ij}^a é a componente (i, j) do vetor peso \vec{w}^a – que, neste caso, refere-se ao campo receptivo a . A saída para este mesmo campo receptivo pertencente ao padrão τ é representada por $S^{a\tau}$, com $\tau = 1, 2, \dots, P$. A grandeza $I_{ij}^{a\tau} = \theta_{ij}^a I_{ij}^\tau$, por sua vez, é o elemento (i, j) do campo receptivo a relativo ao padrão de imagem I_{ij}^τ do *ensemble* de imagens, e θ_{ij}^a é a *matriz de campo receptivo* definida com valor 1 se $(i, j) \in a$ e valor 0 se $(i, j) \notin a$. A figura 2.7 ilustra este modelo.

Partimos da sugestão de Idiart *et al* [12] para uma expressão de reconstrução

$$\tilde{I}_{kl}^\tau = \sum_{a=1}^M S^{a\tau} D_{kl}^a \quad (2.26)$$

com valores $k, l = 1, 2, \dots, N$, a mesma dimensão dos padrões de entrada. A partir desta, generalizamos a solução analítica apresentada na expressão (2.22) reescrevendo-a sob a

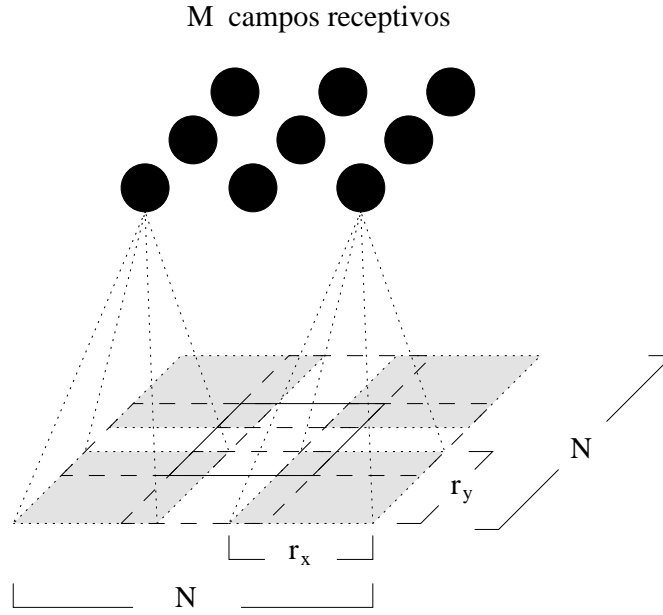


Fig. 2.7: Representação do processo de extração de uma matriz de $M = M' \times M'$ campos receptivos sobrepostos entre si, com dimensão $r_x \times r_y$ (onde $r_x = r_y$), a partir de um padrão de entrada $N \times N$.

notação matricial

$$\mathbf{D}^a = \sum_{b=1}^M [Q^{-1}]^{ab} \mathbf{L}^b \quad (2.27)$$

e nas duas expressões em (2.23)

$$\mathbf{L}^a = \langle \mathbf{I} S^a \rangle \quad \text{e} \quad Q^{ab} = \langle S^a S^b \rangle. \quad (2.28)$$

Vale reforçar que esta solução refere-se ao caso particular de um sistema assumido como *auto-mediante*, caso este em que o erro quadrado médio pode ser escrito como a razão entre duas médias como na equação (2.21).

Uma vez que os P padrões de imagens tenham sido empregados no treinamento da rede, ela terá aprendido a reconstruí-los – o que será feito a partir da equação (2.26) com o peso ótimo \mathbf{D}^a obtido através de (2.27) e (2.28). A qualidade deste aprendizado, refletida na fidelidade da reconstrução, é avaliada pelo erro médio E reescrito da equação (2.20) como

$$E = \frac{1}{2} \left\langle \left(\frac{\mathbf{I}}{|\mathbf{I}|} - \frac{\tilde{\mathbf{I}}}{|\tilde{\mathbf{I}}|} \right)^2 \right\rangle. \quad (2.29)$$

Campos receptivos pequenos, que obedecem a relação $M' r_o < N$ (onde $M' = M^{1/2}$ e r_o pode representar tanto a orientação r_x quanto a r_y), são responsáveis por um baixo desempenho

(ou erro alto), tendo em vista que apenas uma parte da informação total da imagem é lida pela rede. Campos receptivos maiores, por outro lado, para arquiteturas com a relação $M'r_o \geq N$ (situação de *sobreposição* ou “*overlap*” dos campos), apresentam desempenho melhor. A dimensão ótima do campo receptivo será discutida mais adiante no capítulo 4.

Capítulo 3

Estatística de Imagens Naturais

As cenas naturais possuem uma considerável complexidade em sua distribuição estatística que as impede de obedecer a qualquer tipo elementar de distribuição empregada mais habitualmente nas modelagens – como a distribuição gaussiana, por exemplo. E esta complexidade, na opinião de Ruderman [25], não deverá ser removida com qualquer tipo de transformação simples do espaço. Se a característica estatística do problema fosse especificada, estaríamos prontos não somente a quantificar seu comportamento típico, mas também a decidir qual é a arquitetura ótima de um sistema neural em relação a um determinado critério. Como é impossível coletar todos os dados necessários para uma perfeita descrição do *ensemble* natural, nossa estatística sempre estará muito longe de se completar. Este irrisório conhecimento estatístico também não nos permitirá saber o quanto nos aproximamos de imagens realísticas – o que, por si só, pode nos levar a questionar quanto conhecimento é realmente necessário para fazer algo útil dentro deste contexto.

Seja um *ensemble* \mathcal{I} de imagens naturais unidimensionais, tal que uma imagem qualquer $I(x)$ obtida aleatoriamente desta distribuição represente completamente ambientes naturais. Para este *ensemble*, de probabilidade $\mathcal{P}[I(x)]$ de ocorrência de uma dada imagem, definimos a medida $\psi(x) = (f * I)(x) + \rho(x)$, onde $f(x)$ é um filtro com o qual $I(x)$ é convoluído, e $\rho(x)$ é um ruído adicionado ao sinal. Se $I(x)$ tiver banda limitada, média zero e for estacionária (ver seção 3.2.2), podemos demonstrar [25] que a melhor estimativa $I_{\text{est}}(x)$ será

$$I_{\text{est}}(x) = \frac{\int I(x) \mathcal{P}[\psi|I] \mathcal{P}[I] dI}{\int \mathcal{P}[\psi|I] \mathcal{P}[I] dI}, \quad (3.1)$$

sendo $\mathcal{P}[c|d]$ a probabilidade de obter-se c dada a ocorrência de d . Esta equação diz como ir da medida $\psi(x)$ do sinal $I(x)$ para a melhor suposição deste sinal, isto é, $I_{\text{est}}(x)$. Fica claro que o conhecimento do *ensemble* dado por sua distribuição $\mathcal{P}[I]$ desempenha um

importante papel na compreensão de quão bem o sistema operará. Quando os estudos começam a alcançar dimensões superiores, torna-se muito complexo obter todas as informações necessárias para se ter um bom conhecimento sobre o que é $\mathcal{P}[I]$. Um sistema simples com imagens 4×4 e 16 escalas de cinza permite 16^{16} imagens distintas; nestes casos, novas técnicas têm de ser empregadas.

É bastante razoável esperarmos – como o próprio Ruderman sustenta – que não exista um *ensemble* correto, pois cada ser vivo tem seu próprio *habitat* e, cada *habitat*, uma natureza de cenário com suas próprias características e propriedades. Deste modo, os animais de determinado *habitat* estabelecerão uma estatística muito em particular. Isto, por sua vez, nos permite assumir como idéia básica que os sistemas sensoriais de um indivíduo estão adequadamente adaptados para processar os tipos de sinais presentes na natureza específica de seu ambiente original.

Muitos dos critérios propostos para uma arquitetura neural eficiente são baseados em medidas estatísticas – como a redução de redundância e a fidelidade de reconstrução –, o que torna o estudo das propriedades de cenas naturais tão importante. Este capítulo inicia revisando os principais estudos sobre imagens naturais encontrados na publicação científica. Diversos autores foram consultados na composição deste trabalho, cada qual seguindo uma linha distinta de pesquisa enfocando uma característica e sustentando a importância de sua escolha. Trabalhos de pesquisadores como Abbott, Ruderman, Shouval entre tantos outros nos fornecerão contribuição fundamental para a compreensão das características e propriedades físicas de um sistema tão complexo. Apresentaremos as propriedades estatísticas verificadas em cenas naturais e os modelos matemáticos propostos para satisfazê-las.

3.1 Invariância de Escala e Auto-afinidade

Intuitivamente, esperamos que o *ensemble* de imagens naturais contenha a propriedade de *invariância de escala* tendo em consideração que um sistema visual é treinado para captar uma mesma imagem com diversas escalas diferentes à medida que o indivíduo se aproxima ou se afasta do alvo. Trata-se de uma característica discutida com unanimidade pelos pesquisadores que estudam cenas naturais e pode ser provada com histogramas locais ou com a informação mútua entre dois pixels. Particularmente, os autores costumam enfatizar

o caso do *espectro de potência*.

3.1.1 Espectro de Potência

Seja um sinal $f(x)$ no espaço unidimensional com transformada de Fourier $F(k)$ [4]. Definimos o espectro de potência $S(k)$ para o sinal $f(x)$ como

$$S(k) = |F(k)|^2, \quad (3.2)$$

cuja versão para um *sinal discreto bidimensional* I_{ij}^T é

$$\langle S(k_{mn}) \rangle = \left(\frac{1}{2\pi N^2} \right)^2 \frac{1}{P} \sum_{\tau=1}^P \left| \sum_{i,j=0}^{N-1} I_{ij}^T \exp(-ik_{mn} \cdot r_{ij}) \right|^2. \quad (3.3)$$

Em (3.3), r_{ij} é a componente (i, j) do vetor posição no espaço bidimensional, e k_{mn} é a correspondente frequência espacial do vetor na coordenada (m, n) . O somatório externo faz uma média sobre todos os padrões de imagens P .

No escopo das cenas naturais, o *espectro de potência bidimensional* foi empregado exaustivamente como importante ferramenta de estudo. Uma imagem do *ensemble* apresenta uma superfície bastante característica onde se percebe a predominância das baixas frequências. A figura 3.1 exibe um exemplo. A frequência cresce do centro para a periferia; pontos equidistantes do centro possuem o mesmo valor para a frequência. A queda abrupta e o desvio assintótico sugerem uma relação de potência entre o espectro e a frequência espacial.

Operando uma média azimutal, um gráfico $\log \times \log$ revela um comportamento aproximadamente regido pela lei

$$S(k) \propto \frac{1}{|k|^{2-\mu}}, \quad (3.4)$$

sendo μ uma grandeza com valor pequeno. Independente da escolha da escala angular, o espectro de potência conserva-se invariante. Da mesma forma, $S(bk)$ não fere a invariância, pois apenas multiplica o espectro pelo fator constante $1/b^{2-\mu}$, o que não se verificaria em um espectro de potência dado por uma lei exponencial.

Ruderman [25] provou que esta invariância de escala é robusta quanto ao *reescalamento logarítmico* da intensidade do pixel I_{ij} ao empregar no cálculo do espectro de potência a expressão

$$\phi_{ij} = \ln \left(\frac{I_{ij}}{I_0} \right), \quad (3.5)$$

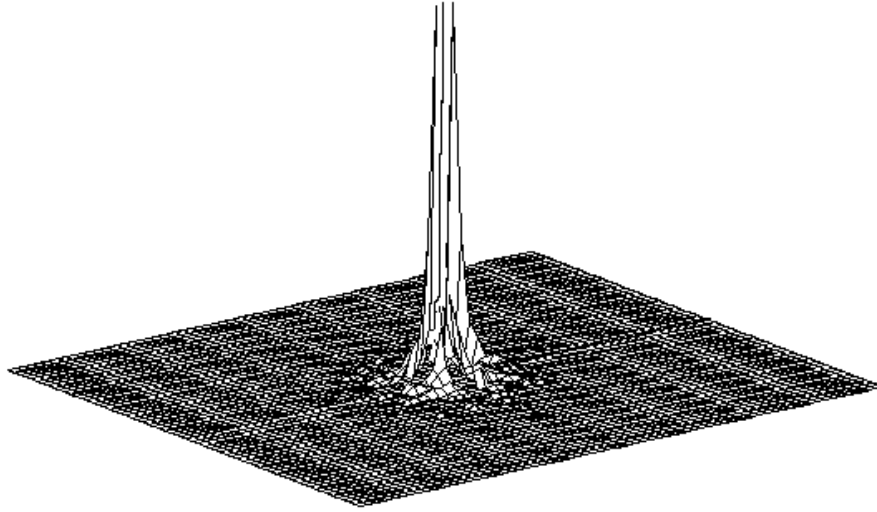


Fig. 3.1: Espectro de potência bidimensional médio *extraído de 10 cenas naturais*. A origem localiza-se no centro da superfície, onde se encontra o ponto máximo. Regiões equidistantes da origem têm a mesma frequência.

onde I_0 é definido para cada imagem tal que $\sum_{ij} \phi_{ij} = 0$. O escalonamento também é robusto a outras mudanças, como diferenças na escolha do *ensemble* ou métodos de captura de imagem. No entanto, os expoentes podem diferir entre autores.

Esta invariância leva à propriedade de *auto-afinidade*. Genericamente falando, o novo *ensemble* pode ser *auto-afim* ao original, significando que as novas imagens também devem ser multiplicadas por uma constante adequada após o reescalonamento, tornando a estatística idêntica à original [25]. Se $\mathcal{P}[\varphi(\alpha x)]$ é qualquer *ensemble* estatístico de $\varphi(x)$ numa escala α , então a invariância de escala implica que

$$\mathcal{P}[\varphi(x)] = \mathcal{P}[\alpha^\nu \varphi(\alpha x)], \quad (3.6)$$

para ν um expoente que independe de α e \mathcal{P} .

3.1.2 Previsibilidade em Cenas Naturais

Um ruído branco é um sinal que, entre outras coisas, possui variância infinita e uma distribuição completamente aleatória. Quando consideramos uma distribuição aleatória, esperamos que *não seja possível* fazer qualquer previsão sobre uma banda ausente do sinal baseado em outra disponível.

Tratando de imagens naturais, por sua vez, constata-se que uma lacuna pode ser intuitivamente preenchida ao levar-se em consideração os dados da vizinhança. Em outras palavras, as cenas naturais possuem um *grau não-nulo de previsibilidade*. Isto nos leva a concluir que cenas naturais não possuem estatística aleatória. Assim, adotamos a definição de Ruderman [25], que nos diz que

“*previsibilidade é a medida da não-aleatoriedade das imagens.*”

Ruderman [28] realizou um interessante estudo a respeito deste tema esperando encontrar grande previsibilidade entre pixels tendo em vista as correlações de longo alcance destas imagens. Empregando a *informação mútua de Shannon* entre os valores de dois pixels, foi capaz de fazer uma estimativa da previsibilidade de um ponto da imagem dispondo das informações de outros pontos, próximos ou não, de sua vizinhança.

Como resultado, obteve uma nova lei de potência relacionando a informação Z com a distância d , ou seja,

$$Z(d, \theta) \propto \frac{1}{d^{\alpha(\theta)}}, \quad (3.7)$$

para α um expoente parametrizado pelo ângulo θ da função. Apesar de não haver uma razão *a priori* que justifique este escalamento da informação mútua – ainda que o *ensemble* de imagens o possua –, identifica-se por meio deste resultado mais uma estatística das cenas naturais com *invariância de escala*.

3.2 Codificação de Imagens Naturais

Muitos algoritmos genéricos de compressão de imagens dividem-nas em blocos que são tratados como entidades independentes. Diferente desses, a compressão mais eficiente depende do conhecimento prévio da estrutura estatística.

O *espaço de estado* é um método adequado de se descrever a redundância em um conjunto de dados. Para descrevermos o espaço de estado de uma população de imagens, definimos um sistema de eixos ortogonais, cada qual indicando o nível de intensidade de um certo pixel, tal que imagens com dimensão $n \times n$ serão representadas por um único ponto em um espaço de estado n^2 -dimensional. Tomando todas as imagens 256×256 existentes em um espaço de dimensão 65.536 (desde as aleatórias, as com estatística gaussiana e as próprias cenas natu-

rais), o volume do espaço que é ocupado por cenas naturais é infinitesimalmente pequeno. Existe um conjunto de infinitos tipos de imagens, dentro do qual poucas delas podem ser realmente identificadas como imagens naturais. Nas palavras do próprio Ruderman [25],

“o fato de que o espaço é amplamente desocupado expressa a redundância na distribuição”.

Ou por outra, qualquer redundância no espaço de estado implica que a população não preenche o espaço uniformemente [7].

Autores como Field, Shouval, Ruderman entre tantos outros estudam as propriedades de cenas naturais em seu espaço de estado com o propósito de encontrar os meios mais eficientes de codificação.

3.2.1 Código Compacto e PCA

Na busca por uma arquitetura que maximize a informação na retina com o menor consumo possível de espaço, surge a proposta de compactação do sinal de uma imagem. Para tanto, a redução dimensional de cenas naturais (conferir o modelo biológico na figura 0.1) a partir da análise de componentes principais tratada na seção 2.2 é um dos principais métodos utilizados. Aproveita-se da redundância dos padrões de entrada para determinar a orientação do vetor sináptico que seja mais representativa para a maximização da variância dos dados.

Em um importante trabalho, Hancock *et al* [8] analisaram e estudaram conjuntos de imagens naturais. A partir delas, realizaram mais de 120.000 apresentações para treinar uma rede linear utilizando-se da regra de aprendizado não-supervisionado de Sanger para múltiplas unidades apresentada na equação (2.14). A amostragem deste experimento incluiu o mundo vegetal, visto tanto à distância quanto a grande proximidade, envolveu paisagens amplas, animais em seu comportamento natural e mesmo seres humanos reunidos em grupo. O propósito era determinar as primeiras 15 componentes principais.

Esta operação foi realizada com ampliações 1/2, 1 e 2 para o mesmo grupo a fim de testar a dependência dos resultados com a escala das imagens. Estes resultados foram confrontados com os de uma análise idêntica efetuada sobre um *ensemble* de imagens de texto composto por palavras da língua inglesa.

A primeira componente principal desta população de cenas naturais é aproximadamente

gaussiana, cujo tamanho reflete a escala de correlação das imagens, ou seja, o nível da relação existente entre pontos distantes da mesma imagem. As segunda e terceira componentes principais se assemelham à primeira derivada da função gaussiana nas direções horizontal e vertical, respectivamente. Segundo Hancock, que também testou as imagens giradas em 45° , estas orientações têm origem na própria natureza das imagens, descartando a possibilidade de ser consequência de um efeito de borda residual.

As correlações que a PCA detecta nas imagens naturais têm suas origens na probabilidade de que pixels adjacentes tenham causa similar, ou seja, pertençam ao mesmo objeto. A forma geral dos campos receptivos das primeiras componentes lembram as componentes verificadas em células simples. A dúvida, contudo, ainda reside no fato de desconhecermos se o sistema visual primário é capaz de executar este processo.

A respeito do trabalho de Hancock [8], Field [7] sustenta que é esperada uma certa semelhança dos primeiros coeficientes com campos receptivos corticais, tendo em vista que uma janela gaussiana senoidal de baixa frequência (os primeiros coeficientes de Fourier com maiores amplitudes) produzirão a *função de Gabor*¹. Mas, insiste ele, não há razão para apegar-se a este resultado. Passados os primeiros três ou quatro coeficientes de Fourier, o perfil dos campos receptivos deixa de parecer-se com perfis de campos receptivos corticais, tornando-se substancialmente diferente daqueles encontrados nos sistemas visuais de mamíferos.

Field e Olshausen [21] mostram que as regras lineares de aprendizado hebbiano são incapazes de retirar as propriedades destes campos receptivos quando realizam PCA. Eles analisaram usando a regra de Sanger partes de cenas naturais com dimensão 8×8 . As componentes principais apresentadas pelos autores, em sua grande maioria, não se parecem com qualquer campo receptivo cortical conhecido. As reconstruções com estas componentes também resultaram imagens pouco claras.

A ineficiência do algoritmo linear hebbiano, sustentam, está em seu aprendizado limitar-se a correlações de primeira e segunda ordem, unicamente. Estruturas *localizadas, orientadas*

¹ A função de Gabor popularizou-se por fornecer bons modelos de campos receptivos corticais. Uma função de Gabor bidimensional orientada verticalmente é definida pela equação

$$g(x, y) = \exp \left\{ - \left[x^2 / 2(\Delta W)^2 + y^2 / 2(\Delta L)^2 \right] \right\} \cos(2\pi f x + \theta), \quad (3.8)$$

onde ΔL é o tamanho espacial do sensor ao longo do seu comprimento (o eixo paralelo à orientação preferencial) e ΔW representa a largura do sensor no espaço (ortogonal à orientação preferencial). Mais detalhes em [6].

e *passa-banda* não são vistas por análises estatísticas que consideram funções correlacionando somente pares de pontos – como discutiremos na seção 3.2.2 adiante.

3.2.2 Código Esperso e Estatística Estacionária

Alguns autores – dentre os quais se destaca Field – caminham na contramão do grande fluxo de pesquisadores que se dedicam a desvendar os segredos do *ensemble* das cenas naturais. Como linha de pesquisa, normalmente questionam a validade da extração de componentes principais assim como a da estatística realizada considerando correlações obtidas com menos de três pontos.

A partir de seus estudos de cenas naturais, Field *et al* [21] afirmam que este *ensemble* é composto por estruturas *orientadas*, *localizadas* e *passa-banda*.

Galhos e troncos, horizontes e a borda de uma folha são exemplos de estruturas orientadas em cenas naturais. Este tipo de estrutura exige pelo menos três pontos para ser identificado, o que sugere que uma função correlação envolvendo apenas dois pixels não carrega suficiente informação sobre esta propriedade.

Imagens naturais são localizadas por conterem estruturas bem determinadas em coordenadas espaciais, o que é caracterizado em termos de Fourier pelo *espectro de fase*². Estruturas localizadas apresentam fases alinhadas através de diferentes frequências espaciais. Da mesma forma, estas imagens contêm estruturas *passa-banda*, que também exigem conhecimento a respeito do espectro de fase.

As componentes principais obtidas pela PCA descrevem unicamente o espectro de amplitude das cenas naturais, tendo em vista que o espectro de fase é irrelevante para a determinação da direção destas. Por conseqüência, não há possibilidade de caracterizar nem estruturas localizadas, nem estruturas *passa-banda*. Mais do que isto, estruturas orientadas, que dependem de pelos menos três pontos, também são desconsideradas, tendo em vista que a análise se utiliza de correlação de apenas dois.

A distribuição de energia, compacta e localizada, sugere que cenas naturais têm *distribuição esparsa* [21]. Das próprias palavras de Field *et al*,

“qualquer dada imagem (natural) pode ser representada com um número relativa-

² O espectro de fase de um sinal $f(x)$ é dado por $\arg[F(k)]$, onde $F(k)$ é a transformada de Fourier de $f(x)$.

mente pequeno de ‘descritores’ contidos em um conjunto muito maior”.

Assim sendo, Field propõe que uma forma eficiente de codificação para imagens naturais não é a feita pela redução dimensional operada com a PCA, mas sim através da *maximização dessa característica esparsa* (conferir a figura 3.2).

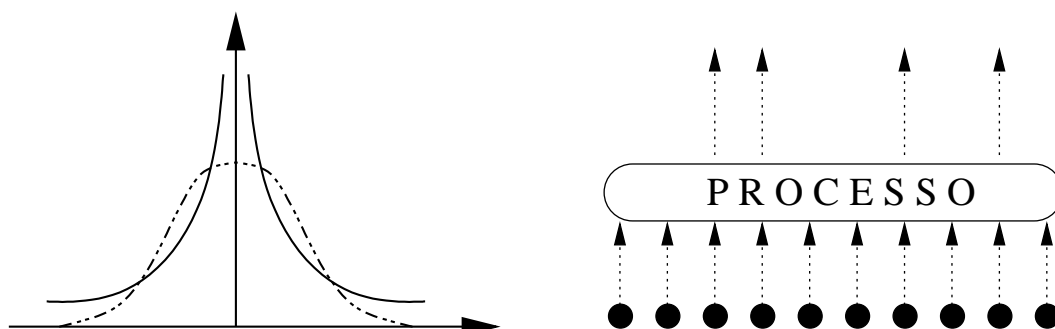


Fig. 3.2: À esquerda, a curva de uma distribuição esparsa (linha contínua) sobrepondo uma curva gaussiana (linha interrompida). A distribuição esparsa tem altos valores em regiões próximas à origem e decai suavemente à medida que se afasta do zero. À direita, um esquema representa o comportamento de um processo regido segundo a distribuição esparsa. O número de descritores (círculos negros) é bastante grande, mas um número pequeno deles é suficiente para representar a imagem.

Diferente da PCA, a codificação esparsa transforma a redundância da entrada reduzindo os padrões de disparo das células. Neste modelo, todas as células têm igual probabilidade de resposta sobre uma classe de imagens, mas baixa probabilidade para uma imagem qualquer. A informação é distribuída sobre todas as células e não existe redução de redundância, tendo em vista que esta é convertida na redundância dos padrões de disparo das células. Em outras palavras, o problema pode ser descrito como a *busca por uma codificação na qual somente umas poucas células respondam a qualquer entrada*.

Ao contrário de Field, para quem a utilização de uma estatística de pelo menos três pontos é fundamental, Ruderman [25] mostra-se incrédulo quanto à possibilidade de usar correlações de ordem superior a dois. Seriam difíceis de interpretar, além de entrar no domínio perturbativo de cálculos estatísticos. Para ele, encontrar propriedades de invariância na distribuição, como as discutidas na seção 3.1, reduz sua complexidade e nos poupa do sacrifício que seria medir a função correlação.

É razoável para Field [7], também, que cenas naturais não tenham localizações especiais na imagem onde a estatística tenda a ser diferente. Em outras palavras, sobre a população

de imagens no conjunto de dados (neste caso, todas as imagens naturais), a estatística em um local não deve ser diferente que em qualquer outro, ou seja, todas as características têm a mesma probabilidade de ocorrer tanto em uma localização quanto em outra. Matematicamente,

$$\mathcal{P}(x_i|x_{i+1}, x_{i+2}, \dots) = \mathcal{P}(x_j|x_{j+1}, x_{j+2}, \dots), \quad (3.9)$$

tal que $\mathcal{P}(c|d)$ está definida na equação (3.1). Esta relação deve valer para todas as imagens em todos os pixels x_i e x_j . Se as cenas naturais satisfizerem esta propriedade, terão *estatística estacionária*.

Quando a estatística de um conjunto de dados é estacionária, toda a redundância refletida na correlação entre pixels é capturada pelo espectro de amplitude dos dados, ou seja, o espectro de amplitude dá uma descrição completa das suas correlações. Além disto, com estatísticas estacionárias, o espectro de amplitude descreve os eixos principais dos dados no espaço de estado. Não fornece uma descrição completa da redundância, mas descreve as amplitudes relativas dos eixos principais.

3.3 Aproximações de Imagens Naturais.

Algoritmos computacionais podem ser utilizados para gerar valores que, uma vez plotados, revelam padrões cujas características (visuais e/ou estatísticas) assemelham-se muito com cenas naturais. Dentre as possíveis aproximações de imagens naturais, destacaremos nesta seção as *cenas fractais* e as *trajetórias de um caminhante aleatório bidimensional*.

3.3.1 Fractais

Como sabemos, montanhas, costas, nuvens e quaisquer outras entidades criadas pela natureza não são facilmente descritas pela geometria euclideana estudada tradicionalmente. Nuvens não são esferas, montanhas não são cones, linhas costeiras não são círculos, assim como árvores não são suaves. Mais especificamente, muitos padrões da natureza são tão irregulares e fragmentados que, uma vez comparados com o mundo euclidiano, exibem altos níveis de complexidade. Para Mandelbrot³ [17],

³ O Dr. Benoit B. Mandelbrot foi quem descobriu a teoria de fractais durante suas atividades como pesquisador do *Thomas J. Watson Research Center* da IBM. Ele, de fato, batizou as imagens com a palavra *fractal*. O conjunto que descobriu e que recebeu seu nome é apresentado na figura 3.3. Com o fim de introduzir a noção de fractal, propôs o desafio de medir

“o número de escalas distintas de comprimento de padrões naturais é, para todos os propósitos práticos, infinito”.

Na abordagem de um tema tão amplo e intrigante quanto os fractais, consideramos pertinente mostrar que a noção de *dimensão geométrica* se divide em várias componentes distintas. O percurso do movimento browniano, por exemplo, é topologicamente uma curva de dimensão 1. Contudo, no caso em que esta curva praticamente preenche um plano, é um fractal de dimensão 2. Mandelbrot considera que o principal fato é que a noção de dimensão possui muitas facetas matemáticas que não somente são conceitualmente distintas, mas podem levar a diferentes valores numéricos.

O espaço euclidiano é limitado a conjuntos para os quais todas as dimensões úteis coincidem, de modo que podemos chamá-los de conjuntos *dimensionalmente concordantes*. Por outro lado, as diferentes dimensões dos conjuntos aos quais nos reportamos aqui não coincidem. Estes são conjuntos *dimensionalmente discordantes*.

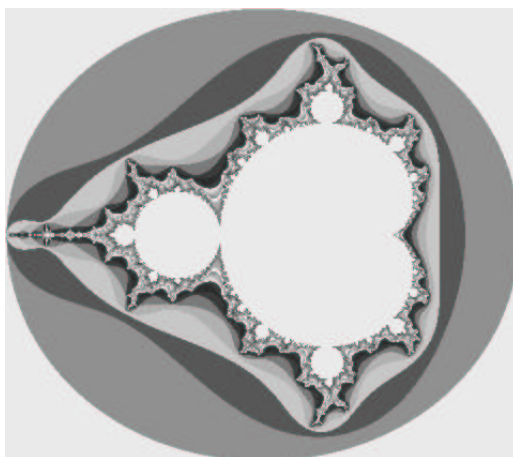


Fig. 3.3: Conjunto de Mandelbrot: *fractal descoberto por Benoit B. Mandelbrot, cujo nome foi emprestado para batizá-lo. Trata-se de uma imagem infinitamente detalhada “comprimida” em um espaço finito e que foi gerada no plano complexo a partir de processos iterativos sobre a expressão $z_{n+1} = z_n^2 + c$, onde $z_0 = 0$ e c é uma constante complexa.*

Parece-nos conveniente adotar – como o fez também Mandelbrot [17] – dois conceitos básicos para a dimensão matemática, cada um deles atribuindo para cada conjunto no espaço euclidiano um número real cujos alicerces formais e intuitivos estabelecem a correspondência

o comprimento da costa da Bretanha [17]. Por outro lado, a variação do comprimento aparente de linhas costeiras com o tamanho da régua que mede foi estudada por Richardson [18]. Pode ser interessante registrar que linhas costeiras reais têm dimensão fractal variando entre 1,15 e 1,25.

para ser chamado de sua dimensão. A mais intuitiva é a *dimensão topológica* d_T , de acordo com Brouwer, Lebesgue, Menger e Urysohn. A segunda dimensão, denotada por d_F , foi formulada por Hausdorff em 1919 e colocada em sua forma final por Besicovitch, o que a levou a ser chamada de *dimensão Hausdorff-Besicovitch*. Aqui, contudo, preferiremos o termo *dimensão fractal*⁴.

Trabalhando no espaço euclidiano \mathfrak{R}^E , tanto d_T quanto d_F têm em zero seus possíveis valores mínimos e em E seus possíveis valores máximos. Esta é a única característica comum entre estas dimensões. Por outro lado, enquanto d_T é necessariamente inteiro, d_F pode assumir valores reais. Além disto, devem satisfazer a desigualdade de Szpilrajn

$$d_F \geq d_T. \quad (3.10)$$

Para todo o espaço euclidiano, $d_F = d_T$; fractais, contudo, satisfazem $d_F > d_T$. Isto nos permite recorrer novamente a Mandelbrot, que define em suas palavras a geometria por ele descoberta:

“um fractal é, por definição, um conjunto para o qual a dimensão Hausdorff-Besicovitch estritamente excede a dimensão topológica”.

A dimensão d_F do espaço ocupado por um fractal depende do número de iterações empregadas para gerá-lo e é obtida aplicando-se a expressão

$$d_F = \frac{\log(P)}{\log(1/R)}, \quad (3.11)$$

onde P é o número de partes no qual dividiu-se o segmento original, e R é o comprimento de um dos semi-segmentos. Mandelbrot foi o primeiro a utilizar com sucesso d_F na descrição da natureza.

A falta de correspondência com o espaço cartesiano não pára no processo não determinístico de geração de uma curva nem em sua dimensão não inteira. Sabe-se, por exemplo, da propriedade de *invariância de escala* (ver seção 3.1) sob certas transformações e que se dá quando o grau de suas irregularidades e/ou fragmentações é idêntico em todas as escalas; a propriedade de *homogeneidade* pode ser detectada quando a invariância é modificada e/ou restringida em seu alcance à medida que nos deslocamos através de um fractal;

⁴ Está fora do escopo deste trabalho apresentar o formalismo desenvolvido por trás da conceituação destas duas dimensões. Detalhes sobre isto podem ser encontrados na referência [17].

a *auto-similaridade estatística* é uma invariância sob similaridade geométrica ordinária. De Mandelbrot *et al* [18],

“a propriedade de objetos por meio da qual subconjuntos ampliados parecem-se com (ou são idênticos a) o todo e as demais partes é conhecida como auto-similaridade”.

Quando um pequeno pedaço de um fractal é similar ao seu todo, dizemos que é auto-similar. Objetos auto-similares são, em geral, fractais, mas a recíproca nem sempre é verdade. Um fractal é definido pela *irregularidade*, que deve existir em todas as escalas, mas essa irregularidade não precisa parecer a mesma.

A propriedade de que objetos possam parecer estatisticamente similares enquanto, ao mesmo tempo, diferentes em detalhes a diferentes escalas é a característica central dos fractais na natureza. A maneira como os detalhes variam com a ampliação da escala é novamente caracterizada por uma dimensão fractal, ou seja, a dimensão fractal determina o detalhamento a diferentes escalas. Quanto maior for d_F , tanto mais rudimentar a imagem parecerá.

A figura 3.4 ilustra 3 imagens fractais cujos detalhes relacionados à sua criação podem ser encontrados no apêndice C.

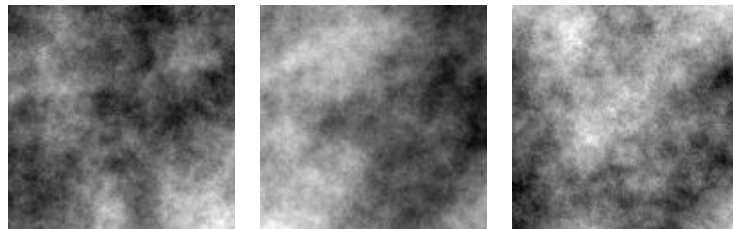


Fig. 3.4: Três imagens 128×128 pixels compostas por exemplos de cenas fractais com dimensão $d_F = 2,5$ criadas sem somas aleatórias e com desvio padrão igual a 1 através do processo de geração de movimento browniano pelo método de substituição do ponto médio para duas dimensões.

A listagem a seguir expõe algumas das diferentes qualidades dos fractais. Vale ressaltar que nem todas essas qualidades se aplicam a todo fractal:

- dimensão fracionária;
- estrutura complexa em todas as escalas;
- subdivisão infinita;

- auto-similaridade estatística;
- dinâmica caótica.

3.3.2 Caminhante Aleatório

Considere uma partícula percorrendo um caminho aleatório em uma linha real do espaço unidimensional. Suponha que inicie de $x = 0$ e se mova uma unidade para esquerda com probabilidade $1/2$ e uma unidade para a direita com probabilidade $1/2$ levando uma unidade de tempo. No passo seguinte, se move um passo para a esquerda ou para a direita com igual probabilidade, independentemente de sua localização depois do primeiro passo. Continuando este procedimento, obtemos um caminho aleatório que é o modelo matemático mais simples do movimento browniano linear.

O movimento unidimensional acima pode ser modelado mais genericamente para um caminhante que, partindo da origem, segue em passos de comprimento l com probabilidade p de tomar a direita e probabilidade q de tomar a esquerda. Para valores normalizados, assumimos $q = 1 - p$, tal que $p = q = 1/2$ se os sentidos forem equiprováveis⁵. O problema pode ser colocado como a busca da probabilidade $P_N(m)$ de o caminhante ser encontrado na posição $x = ml$ após N passos. Aqui, definimos m como a distância medida em número de passos entre a posição final x e a origem, de modo que m assume os valores inteiros contidos no intervalo $-N \leq m \leq N$.

Após N passos, a probabilidade de ter dado N_1 passos para a direita e N_2 passos para a esquerda vem da distribuição binomial

$$P_N(N_1) = \frac{N!}{N_1!N_2!} p^{N_1} q^{N_2}, \quad (3.12)$$

tal que o elemento fracionário é um fator combinatório que expressa o número de seqüências com N_1 passos para a direita e N_2 passos para a esquerda, com probabilidade dada por $p^{N_1} q^{N_2}$. A probabilidade $P_N(N_1)$, já devidamente normalizada, está sujeita aos vínculos $p + q = 1$ e $N_1 + N_2 = N$. Como $m = N_1 - N_2$, a probabilidade $P_N(m)$ será dada por

$$P_N(m) = \frac{N!}{\left(\frac{N+m}{2}\right)! \left(\frac{N-m}{2}\right)!} p^{\frac{N+m}{2}} q^{\frac{N-m}{2}} \quad (3.13)$$

⁵ Os sentidos podem não ser equiprováveis quando o caminhante estiver sujeito a forças externas privilegiando um sentido ou se o ambiente não for homogêneo, por exemplo.

que, para um número muito grande de passos, tende para uma distribuição gaussiana [26].

No plano bidimensional, o caminhante, além da esquerda e da direita, ainda pode optar por seguir um passo acima ou um passo abaixo a cada unidade de tempo. Neste caso, uma situação equiprovável impõe $p_1 = p_2 = p_3 = p_4 = 1/4$ as probabilidades em todos os 4 sentidos possíveis. Ademais, pode-se determinar um ponto de partida aleatório dentro do espaço de navegação e arbitrar a soma de um valor numérico fixo ν para as coordenadas que fizerem parte do itinerário do caminhante de tal modo que a coordenada visitada n vezes acumule o valor $n\nu$ em sua posição. Assim, criamos um mapa da trajetória deste caminhante (figura 3.5) que apresenta similaridade com cenas naturais – similaridade esta cuja estatística será investigada no capítulo 4.

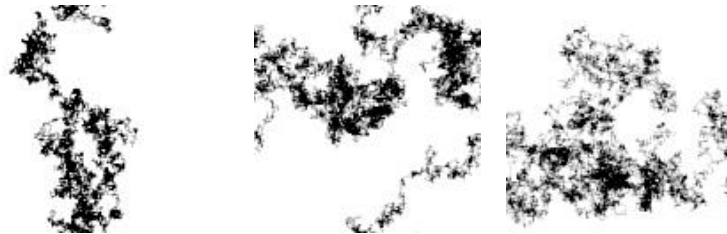


Fig. 3.5: Três imagens exemplificando possíveis caminhantes aleatórios bidimensionais com dimensão 128×128 pontos. Usamos condição de contorno periódica levando o caminhante a ser “transportado” de um lado para o outro quando alcança uma borda, sendo que o caminhante viajou pelo espaço até que a média aritmética dos pontos alcançasse um valor de escape γ . Da esquerda para a direita, os valores de escape utilizados foram respectivamente $\gamma = 1/5$, $2/5$ e $1/2$.

Capítulo 4

Procedimentos e Resultados

Como ponto de partida para nossa pesquisa, nos questionamos a respeito dos métodos empregados pelos diversos autores no processo de análise das propriedades de cenas naturais com ou sem o uso de redes neurais. É evidente que a natureza do trabalho exige o computador como ferramenta básica e fundamental nesta análise, mas muitas escolhas, na ausência de um critério quantitativo explícito, dependem essencialmente de noções empíricas ou subjetivas – para não dizer intuitivas – do pesquisador.

A primeira destas escolhas foi a origem das imagens. A realização do estudo de cenas naturais com redes neurais sugere a idéia de que uma grande quantidade de padrões de treinamento será necessária para a formação de conjuntos sinápticos representativos. Ainda não somos capazes de sintetizar cenas naturais, o que nos força a retirá-las da própria natureza. Dentre os processos possíveis, destacam-se particularmente as câmeras fotográficas e a internet.

A maior parte dos autores consultados (Shouval [29], Liu [31], Hancock *et al* [8], Ruderma [25], Field [7, 6], Parga *et al* [33]) optaram por selecionar um ambiente natural adequado e, dispondo de uma câmera, extraíram imagens fotográficas que satisfizessem certas condições. A variedade do equipamento citado é bastante grande e nem sempre um procedimento de correção das irregularidades óticas foi empregado antes ou depois da digitalização das fotos. Os ambientes de onde foram colhidas estendem-se em diversas regiões; Field, inclusive, foi mais longe e tomou fotos de locações na Inglaterra e na Grécia.

Um número bem menor de autores, como Parga, Turiel *et al* [19, 33] e Lazzaro *et al* [13], usou a internet como fonte de seus padrões. O banco virtual *Natural Stimuli Collection*¹

¹ A referência é feita do modo como solicitado pela página do banco virtual, que pode ser localizada através do endereço <http://hlab.phys.rug.nl/archive.html>.

[35], a única referência citada, é composto de aproximadamente 4.000 imagens calibradas obtidas com uma câmera *Kodak DCS420* e que podem ser baixadas livremente para uso científico e não comercial. As imagens têm dimensão 1.536×1.024 pixels e contêm cenas puramente naturais, puramente artificiais, ou uma composição destas.

Fosse através de fotografias, fosse através da navegação na internet, todos concordavam quanto aos critérios de escolha das imagens: evitaram enquadramentos que incluíssem objetos ou estruturas feitos pelo homem (construções, estradas, placas, etc.). No contexto da natureza, horizontes retilíneos ou grandes regiões vazias (céu ou chão unicamente) também eram indesejados.

O passo seguinte foi estabelecer o número e as características das imagens selecionadas. Aqui, o critério determinante foi a escolha do processo usado no estudo. Field [6], por exemplo, considerou que apenas 06 imagens com dimensão de 256×256 pixels e 256 níveis de cinza seriam suficientemente representativas. Outros autores aumentaram sua população de imagens variando de 15 (Hancock *et al* [8]) a 70 (Ruderman [25]). O próprio Field aplicou 55 imagens para a publicação em [7]. Parga *et al*, em [19], consideraram mais apropriado o uso de 200 imagens de 16 bits com dimensão de 1.536×1.024 pixels; em [33], por outro lado, dividiu os padrões em dois conjuntos: o primeiro composto por 45 imagens de 13 bits com dimensão de 256×256 pixels, e o segundo composto por 25 imagens de 16 bits com dimensão de 1.536×1.024 . Lazzaro e Wawrzynek [13] experimentaram 699 padrões, um número expressivamente maior, enquanto Shouval *et al* [30] elevou para 1.000. O processo de ambos distingüe-se, contudo, pelo fato de que os padrões de Lazzaro eram compostos por imagens de 451×438 pixels; Shouval, por outro lado, extraiu os seus (com dimensão de 13×13) de somente 12 cenas naturais.

O leque de possibilidades de escolha do número de imagens e suas características só foi superado pela variedade de processos adotados. O pré-processamento (normalmente visando a simular os campos receptivos da retina através de filtros de vizinhança circular) foi realizado em muitos casos, tais como Shouval *et al* [29, 30, 31]; Hancock *et al* [8] e Field [6, 7] acautelaram-se com procedimentos que minimizassem resultados espúrios advindos de efeitos de borda.

Antes das análises, Field também determinou o logaritmo das imagens, e os cálculos

sobre as mesmas foram baseados nesta transformação. Este procedimento foi adotado a fim de permitir com que as unidades respondessem em termos do contraste (razões das intensidades) ao invés das amplitudes (diferenças de intensidades), dada a propriedade logarítmica $\log(a) - \log(b) = \log(a/b)$. Isto, porque acredita que células produzam uma resposta linear maior quando trabalham com o contraste ao invés da amplitude.

Com o propósito de tornar o histograma de cada imagem individualmente centrado, Ruderman [25] empregou o logaritmo do contraste definido através da equação (3.5). Todas as análises que seguiram empregaram a grandeza ϕ_{ij} .

As transformações de Fourier e de Gabor também fizeram parte de muitos trabalhos.

Nossos estudos foram feitos utilizando 6 *ensembles* distintos. O primeiro deles, foco principal do trabalho, é o conjunto das *cenais naturais*, cujas imagens foram extraídas da rede mundial de computadores, um mecanismo rápido que dispõe de ampla variedade de exemplos possíveis (incluindo exemplos de [35], mas não somente eles). Para fins comparativos, um conjunto de *cenais artificiais*, de imagens retiradas também da internet, foi preparado. Os terceiro e quarto *ensembles* incluíram dois grupos de imagens criadas através de algoritmos computacionais: as cenais geradas por um *caminhante aleatório bidimensional* e as *cenais fractais*. Um conjunto adicional de imagens naturais criado a partir do *filtro logarítmico de Ruderman* (como citado no parágrafo acima) e um *ensemble* de cenais de *ruído branco bidimensional* completaram o quadro. Mais detalhes sobre os *ensembles* podem ser obtidos no apêndice D.

4.1 Espectro de Potência

Antes de iniciar o longo estudo sobre as propriedades de cenais naturais na abordagem das redes neurais, é inevitável que desejemos averiguar a validade da equação (3.4) mencionada por Ruderman como colocado no capítulo 3.

A média do espectro de potência foi calculada sobre 30 imagens de dimensão 128×128 pertencentes aos 6 *ensembles*, resultando em uma superfície com aproximada simetria azimutal. Operando uma média circular em torno do eixo de simetria e aplicando o logaritmo, obtemos um gráfico bidimensional $\log(S) \times \log(k)$ – exibido na figura 4.1 –, tal que S é a

média do espectro de potência, e k é a frequência espacial.

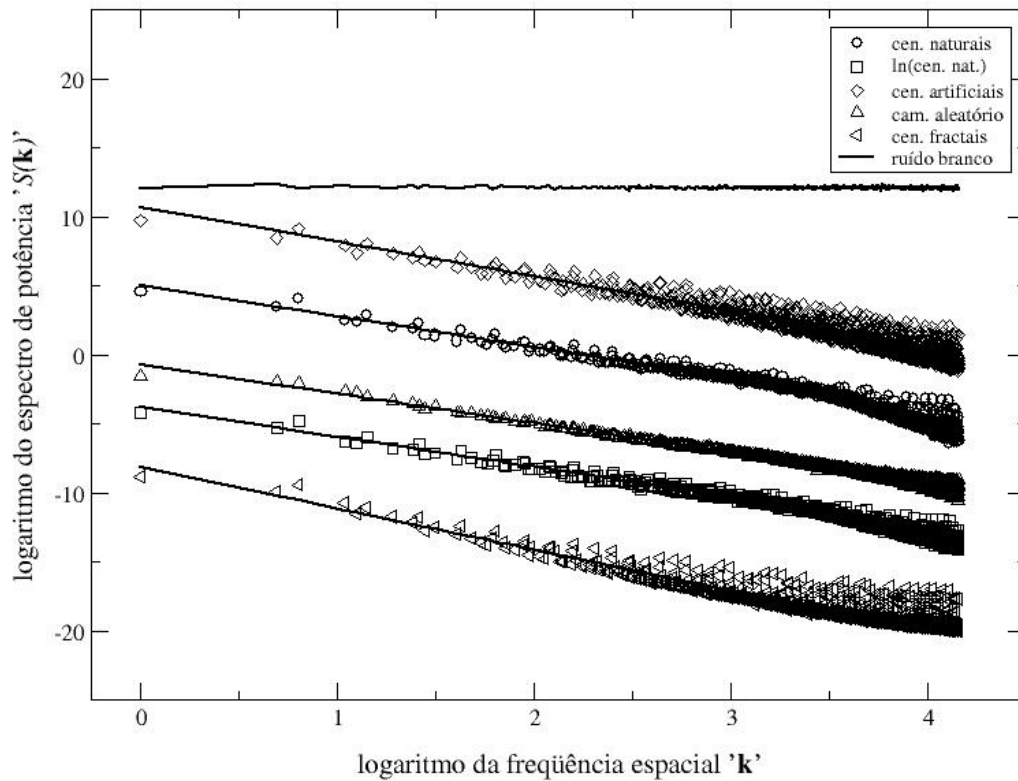


Fig. 4.1: Logaritmo da média do espectro de potência de 6 ensembles distintos (com 30 imagens 128×128 cada) plotado em função do logaritmo da frequência espacial. As retas sobrepondo os pontos são ajustes lineares realizados através do método dos mínimos quadrados para $\log k \leq 3,5$.

Ajustes lineares das curvas foram feitos dentro do intervalo $\log k \leq 3,5$, intervalo este que descarta um volume expressivo de pontos nas frequências altas do espectro. Este procedimento foi adotado tendo em vista que a região de altas frequências é calculada no limite da resolução da imagem, quando não há mais informação suficiente para uma determinação precisa do resultado através do método da transformada discreta de Fourier – discutida na seção 3.1.1. O valor $\log k = 3,5$ foi escolhido como limite por ser o ponto aproximado a partir do qual as curvas iniciam uma queda e se desviam da reta que descreviam.

Os pontos também foram transladados verticalmente visando a evitar que as curvas se sobrepusessem.

No topo do gráfico, uma reta horizontal descreve o espectro de potência de imagens representadas por ruído branco, isto é, matrizes com 128 linhas e 128 colunas compostas

por valores aleatórios entre 0 e 1. A reta é horizontal porquê, diferente das demais, não privilegia qualquer frequência. Um exemplo de ruído branco pode ser visto na imagem mais à direita da figura D.3.

Os outros *ensembles* apresentaram predominância de baixas frequências. Para as populações utilizadas, houve mais afinidade entre as cenas naturais (figura D.1), os caminhantes aleatórios e as cenas naturais com pré-processamento logarítmico (exemplos 1 e 3 da figura D.3, respectivamente). Seus ajustes resultaram em retas praticamente paralelas entre si, o que é parcialmente explicado pelo fato de terem menor dispersão em altas frequências. Também apresentaram os menores valores de μ em módulo, respectivamente -0,23, -0,12 e -0,17.

Com $\mu \approx -0,51$, as cenas artificiais (figura D.2) revelaram menor afinidade com a lei de potência das cenas naturais dada em (3.4), perdendo unicamente para as cenas fractais (exemplo 2 da figura D.3), que apresentaram $\mu \approx -0,98$.

4.2 Campos Receptivos em uma Rede Linear

O propósito desta seção, dentre outras coisas, é apresentar o erro médio determinado por nosso algoritmo a partir da reconstrução de conjuntos de padrões de imagens. Para tanto, empregamos a rede neural linear *feed-forward* descrita pela equação (2.25) treinada sobre entradas restritas a campos receptivos. Dentre os parâmetros passíveis de variação, destacamos o número M de unidades de saída e a dimensão \vec{r} do campo receptivo. Tratando-se de padrões de entrada bidimensionais $N \times N$, os campos receptivos como concebidos nos trabalhos de Idiart *et al* [12] têm de ser ampliados em mais uma dimensão (conferir a seção 2.3). Por conta disto, eles serão tratados através da notação $\vec{r} = (r_x, r_y)$, onde $0 < r_x, r_y \leq N$ podendo cada componente variar independentemente do valor da outra. Todavia, em uma primeira aproximação, conservamos sempre a relação $r_o \equiv r_x = r_y$, de tal forma a impor um número de unidades de saída obedecendo a regra $M = (M')^2$, onde M' é o número de unidades em uma dada orientação. Em outras palavras, como os campos receptivos tinham lados com comprimento sempre igual, o número de campos extraídos na orientação r_x foi o mesmo da orientação r_y . Isto implicou em um mesmo número M' de unidades de saída para cada uma das duas orientações, ou seja, em uma matriz $M' \times M'$, como pode ser visto

no esquema da figura 2.7.

A regra de aprendizagem não-supervisionada é a regra local de Oja para uma única unidade adaptada da equação (2.12) para aceitar dois índices. Para o processo de treinamento, utilizamos taxa de aprendizagem $\eta = 10^{-12}$ e 2.000 iterações. Isto significa que a rede foi submetida 2.000 vezes a cada um dos P padrões de imagens, sendo os vetores pesos atualizados em cada uma destas iterações. O conjunto de pesos gerado pelo treinamento é empregado pela mesma rede sobre um conjunto de imagens (de mesma natureza ou não) buscando reconstruí-las. A reconstrução de cada uma das imagens do conjunto apresenta uma diferença em relação à original correspondente, propiciando, deste modo, o cálculo do erro de reconstrução.

Erro Médio de Reconstrução

O erro médio de reconstrução E é determinado através da equação (2.29), enquanto a forma analítica da reconstrução $\tilde{\mathbf{I}}$ vem da forma (2.26) para E , ou seja,

$$E = \frac{\langle |\mathbf{I} - \tilde{\mathbf{I}}|^2 \rangle}{\langle |\mathbf{I}|^2 \rangle} \quad \text{onde} \quad \tilde{I}_{kl}^\tau = \sum_{a=1}^M S^{a\tau} D_{kl}^a, \quad (4.1)$$

e depende das grandezas $\mathbf{D}^a = \sum_{b=1}^M [Q^{-1}]^{ab} \mathbf{L}^b$ – função de $\mathbf{L}^a = \langle \mathbf{I} S^a \rangle$ e $Q^{ab} = \langle S^a S^b \rangle$ – e $S^{a\tau}$ ainda é a saída a do padrão τ como dado em (2.25).

Reconstrução de Imagens

Da mesma forma como determinamos quantitativamente o desempenho da rede no aprendizado sobre campos receptivos, podemos fazer uma avaliação qualitativa observando as imagens reconstruídas propriamente ditas. Tendo findado o aprendizado, e os pesos convergido para valores estáveis, podemos usar a rede para reconstruir imagens, sejam as utilizadas em seu treinamento, sejam outras quaisquer.

É importante reforçar, contudo, que a expressão (4.1), com a qual as reconstruções são realizadas, dependem das grandezas $\mathbf{L}^a = \langle \mathbf{I} S^a \rangle$ através da matriz \mathbf{D}^a – grandezas estas que têm seu caráter estatístico expresso através da média $\langle \cdot \rangle$ sobre o *ensemble* de imagens. Os gráficos analisados nas seções posteriores nos revelarão que as diversas arquiteturas de rede, quando submetidas a entradas restritas a campos receptivos, são capazes de obter ótimo

desempenho. Para efeitos práticos, contudo, é necessário que este bom desempenho não seja apenas no valor médio; em outras palavras, a rede precisa ter condições de reconstruir cada imagem isoladamente separando e direcionando suas características particulares de modo a que ela possa ser reconhecida dentre aquelas que foram utilizadas no treinamento. A rede deve também mostrar-se tanto mais eficiente nesta tarefa quanto menor for o erro médio observado.

Para fins de estudo, escolhemos parâmetros referentes a três pontos (vide figuras 4.3, 4.7, 4.10 e 4.12) designados pela notação E_{M,r_o} , onde M é o número de unidades de saída, e r_o é a dimensão de uma das orientações do campo receptivo. São os pares $(M, r_o) = (4, 65)$, $(25, 30)$ e $(25, 65)$. Estes pontos foram escolhidos por serem considerados estratégicos ao refletirem algumas particularidades: $E_{4,65}$ e $E_{25,30}$ marcam o início da sobreposição dos campos nas arquiteturas com $M = 4$ e $M = 25$, respectivamente. Da mesma forma, os pontos $E_{25,30}$ e $E_{25,65}$ comparam o desempenho da mesma rede (a “rede ótima”) sob condições de baixo e alto índice de sobreposição dos campos respectivos.

Retropropagação de Erro

Utilizamos as mesmas populações de imagens para treinar uma rede linear *feed-forward* com método de aprendizagem de retropropagação de erro já discutido na seção 2.1.3. Para uma arquitetura de duas camadas, a regra de aprendizagem das conexões entre a camada oculta e a camada de saída é apresentada pela equação (2.8), enquanto a expressão (2.9) corresponde ao aprendizado das conexões entre os terminais de entrada e a camada oculta. A taxa de aprendizagem mais eficiente foi $\eta = 10^{-8}$, e o parâmetro variado, na ausência de campos receptivos, foi apenas o número M de unidades ocultas.

4.2.1 Cenas Naturais

Uma rede treinada com a população de imagens de cenas naturais pode ser vista nos gráficos de desempenho da figura 4.2. Todos os gráficos referem-se ao desempenho nas reconstruções do mesmo *ensemble* de imagens empregado na geração dos vetores pesos da rede. Além disto, cada gráfico compara curvas obtidas para 4 arquiteturas distintas: $M = 4$ unidades, $M = 9$, $M = 16$ e $M = 25$.

O gráfico 4.2 (A) procura simular um caso ideal no qual não existe ruído interferindo no treinamento da rede ou no reconhecimento do sinal reconstruído. Enquanto os campos receptivos não se juntam, e a imagem permanece incompleta para a rede (região à esquerda da curva tracejada), verificamos uma queda bastante aguda do erro para todas as unidades à medida que aumenta a dimensão r_o do campo receptivo. Após este limite, o erro se estabiliza, e as curvas alcançam seu patamar mínimo. Maior número de unidades implica em mais campos receptivos, implica em mais pontos da imagem lidos e também em mais correlação entre estes pontos. Tudo isto favorece a melhora do desempenho, que é refletida pelo decréscimo do erro.

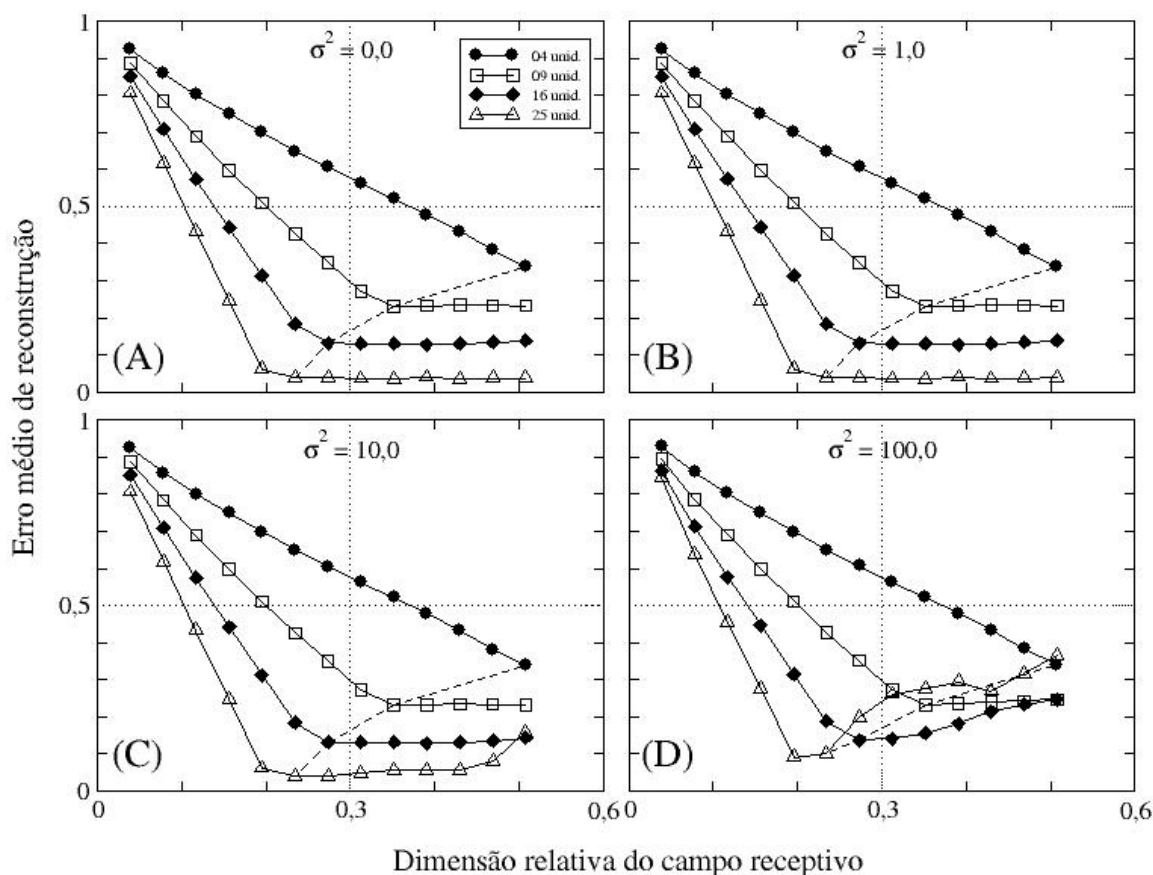


Fig. 4.2: Erro médio de reconstrução E pela variação do campo receptivo r_o . A rede foi treinada com cenas naturais, e os gráficos (A), (B), (C) e (D) representam seu desempenho quando submetida a ruídos de natureza gaussiana com variância 0, 1, 10 e 100, respectivamente. A curva tracejada separa os pontos obtidos de campos sem sobreposição dos pontos vindos de campos com sobreposição. A escala dos eixos está normalizada em $E \times r_o/N$.

Os gráficos 4.2 (B), (C) e (D) correspondem ao mesmo caso anterior com a diferença de estarem sujeitos à presença de ruído gaussiano, tanto no treinamento, quanto no reconhecimento do sinal. Em (B), o ruído (variância $\sigma^2 = 1,0$) não se mostra relevante, e o gráfico é idêntico ao do quadrante (A); em (C) e em (D) ($\sigma^2 = 10,0$ e $\sigma^2 = 100,0$, respectivamente) sua presença já se faz notar. Ainda que as arquiteturas para $M = 4$ e $M = 9$ sigam visivelmente insensíveis ao ruído, para $M = 16$ e $M = 25$ observamos alterações no comportamento das curvas. À medida que a dimensão do campo receptivo cresce, o desempenho decresce – o que é verificado somente para os pontos à direita do limite tracejado. Maior número de unidades implica em mais ruído e, por conseguinte, em maior erro. No caso extremo para $M = 25$ e $\sigma^2 = 100,0$, os pontos superaram em erro as curvas com menos unidades.

Estes resultados, assim como os de Idiart *et al*, sugerem que, sob a presença de ruído no sinal recebido por células neurais, os campos receptivos podem ser ajustados de modo a maximizar a acuracidade das reconstruções dos dados de entrada a partir de uma representação reduzida.

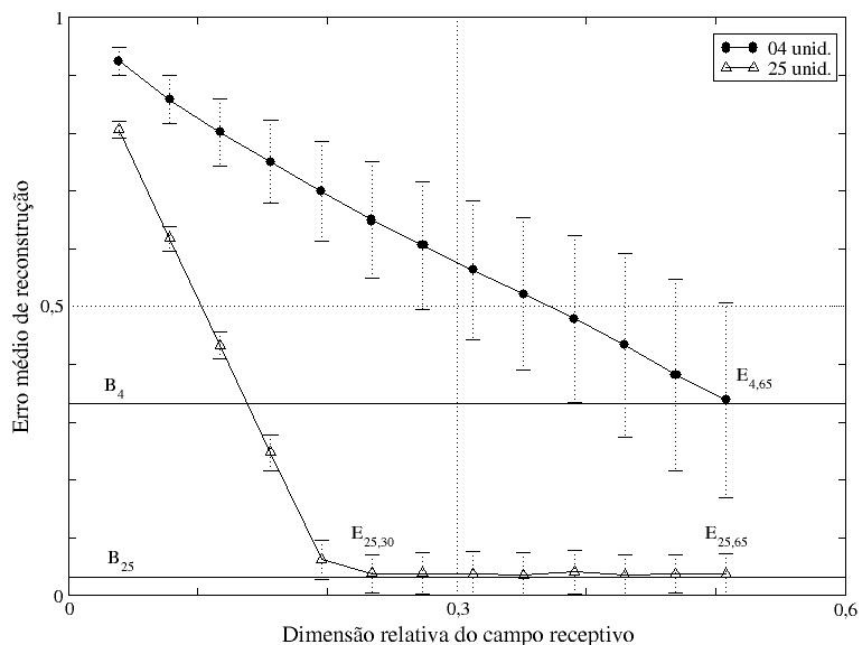


Fig. 4.3: Erro médio de reconstrução sobre cenas naturais para uma rede com campos receptivos restritos (as duas curvas com barras de erro indicadas pela legenda) e para outra treinada com retropropagação de erro (as duas retas paralelas indicadas por B_M).

O grupo de imagens escolhido tem características muito heterogêneas dada a diversidade de ambientes naturais envolvidos. Em alguns exemplos, como 4, 6 ou 11, é possível perceber a predominância da propriedade de *orientação* que Field cita (conferir a seção 3.2.2). Em outros, como 16, 17 e 18, a *auto-similaridade* mostra-se muito presente. É evidente que a predominância de uma propriedade em algumas imagens não exclui a presença das demais propriedades nestas mesmas imagens. Todas contêm estas características em graus diferenciados. Contudo, este fato certamente influi na eficiência da rede.

Para saber o quanto os valores do erro se distanciam da sua média, podemos analisar o gráfico da figura 4.3, que exhibe a primeira e a última curvas do gráfico 4.2 (A) com barras de erro. Para uma mesma arquitetura, ou seja, para o mesmo número M de unidades de saída, os desvios crescem com o crescimento da dimensão do campo receptivo, como se o aumento de informação sobre os padrões levasse a rede a ter seu desempenho mais diferenciado para cada imagem individualmente. Comparando as duas arquiteturas, verificamos que o erro médio de reconstrução se torna mais representativo em uma rede com mais unidades de saída. Além disto, uma vez que a curva inferior (25 unidades) alcança seu patamar de estabilidade, seus desvios também estabilizam-se – fato este que não pode ser analisado na outra curva, pois a estabilização deve ser alcançada somente no último ponto.

O gráfico na figura 4.3 também apresenta o desempenho de uma rede treinada com as 30 cenas naturais pelo método da retropropagação de erro ao longo de 1.000 iterações². O algoritmo de retropropagação simulou dois codificadores $N \times N - M - N \times N$ distintos: B_4 (a reta superior) é um codificador $128 \times 128 - 4 - 128 \times 128$, e B_{25} (a inferior) é um codificador $128 \times 128 - 25 - 128 \times 128$. Tendo-se em consideração que a retropropagação de erro realiza, em última instância, a extração das M primeiras componentes principais, a validade deste resultado está condicionada ao número de autovalores significativos da matriz correlação dos padrões de entrada. Espera-se que a correlação de imagens $N \times N$, que tem dimensão $N^2 \times N^2$ – e, por conseguinte, N^2 autovetores –, concentre a maior parte de seu traço nos primeiros M autovalores.

As imagens dos nossos *ensembles* geram correlações com 16.384 linhas e 16.384 colunas, matrizes estas que exigem um trabalho computacional descomunal para serem resolvidas.

² Os resultados, no entanto, demonstram que a saturação, ou seja, o aprendizado já é obtido a partir da 90ª iteração para B_4 e da 580ª iteração para B_{25} .

Todavia, análises podem ser feitas com matrizes de ordem inferior se reduzirmos a dimensão das imagens para 64×64 , 32×32 e 16×16 fazendo, por exemplo, uma média aritmética do valor da intensidade de blocos de 2×2 pixels.

O estudo destas correlações revelou uma propriedade comum: os autovalores mais significativos são sempre representados pelos 29 primeiros, que concentram praticamente 100% do traço. Quando normalizados à unidade, as curvas da figura 4.4 quase se sobrepõem, tanto em magnitude quanto em forma.

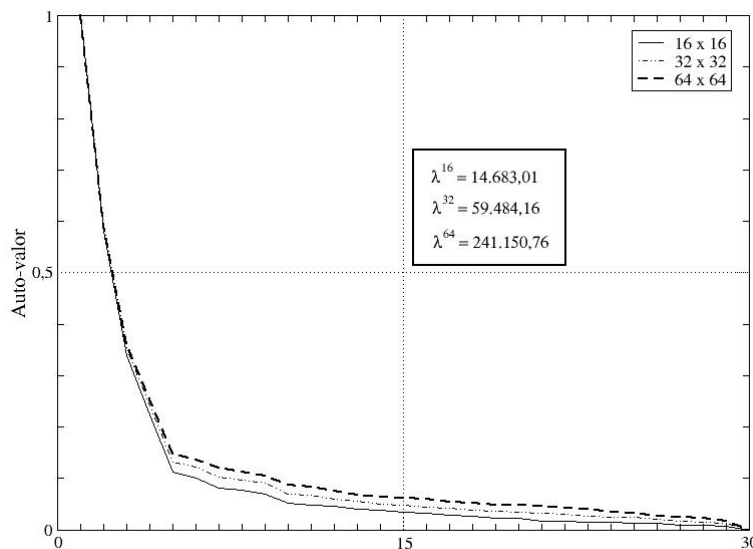


Fig. 4.4: Distribuição dos 30 autovalores mais significativos das correlações das cenas naturais com dimensão reduzida para 64×64 , 32×32 e 16×16 pixels. A magnitude dos mesmos foi normalizada à unidade segundo os λ_{\max} do quadro central.

Os resultados mostram que uma rede com regra de aprendizado local não-supervisionado (arquitetura de campos receptivos restritos com M unidades de saída) é capaz de um desempenho quantitativo e qualitativo equivalente ao aprendizado supervisionado não-local (retropropagação com M unidades ocultas).

Os testes de reconstrução das imagens foram bastante animadores. A figura 4.5, por exemplo, apresenta a 15ª imagem natural da população D.1 reconstruída para os parâmetros mencionados no início da seção 4.2. Como $E_{4,65}$ tem valor alto, temos uma imagem pouco clara; da mesma forma, como $E_{25,30} \approx E_{25,65}$ têm valor baixo, as reconstruções são bem definidas e muito similares. A quarta imagem, oriunda de uma rede híbrida treinada com

cenar artificiais sob os parâmetros $E_{25,30}$, e a quinta imagem, reconstruída sob os parâmetros B_{25} , são tão claras quanto as duas anteriores, revelando um desempenho bastante eficiente.



Fig. 4.5: *Reconstrução da 15ª cena natural do ensemble D.1. Da esquerda para a direita, as imagens representam os pontos $E_{4,65}$, $E_{25,30}$ e $E_{25,65}$. As últimas colunas (em destaque), são a reconstrução da mesma imagem por outras duas redes: a primeira corresponde a uma rede híbrida treinada com cenar artificiais usando os parâmetros do ponto $E_{25,30}$, e a segunda é um codificador 128×128 - 25 - 128×128 .*

4.2.2 Cenar Artificiais

Analogamente ao caso das cenar naturais, os gráficos em 4.6 exibem o desempenho da rede quando treinada a partir do *ensemble* de cenar artificiais apresentado na figura D.2.

Os gráficos 4.6 (A) e (B) apresentam curvas de magnitude praticamente idênticas às de seus análogos em 4.2. Mais do que isto, a evolução delas segue o mesmo comportamento, seja na variação do número de unidades de saída, seja no incremento da dimensão dos campos receptivos. Logo, pode-se concluir que um conjunto heterogêneo de cenar artificiais contém propriedades que permitem à rede treinada obter desempenho similar ao de uma rede treinada com um conjunto heterogêneo de cenar naturais, ainda que sob a presença de ruído gaussiano de baixa variância.

Esta conclusão também se mostra válida nos treinamentos sujeitos a ruído com variâncias mais altas para os casos de menor número de unidades de saída ou de menor dimensão dos campos receptivos. Apenas nos pontos extremos da última curva se faz perceber a diferença entre as curvas das imagens naturais e as das imagens artificiais. O modo como estes pontos ascendem e distorcem a forma da função é diferente em cada caso, mas ainda obedecendo às mesmas regras. Novamente aqui, os resultados sugerem que os campos receptivos podem ser ajustados para maximizar a acuracidade das reconstruções quando sob a presença de ruído.

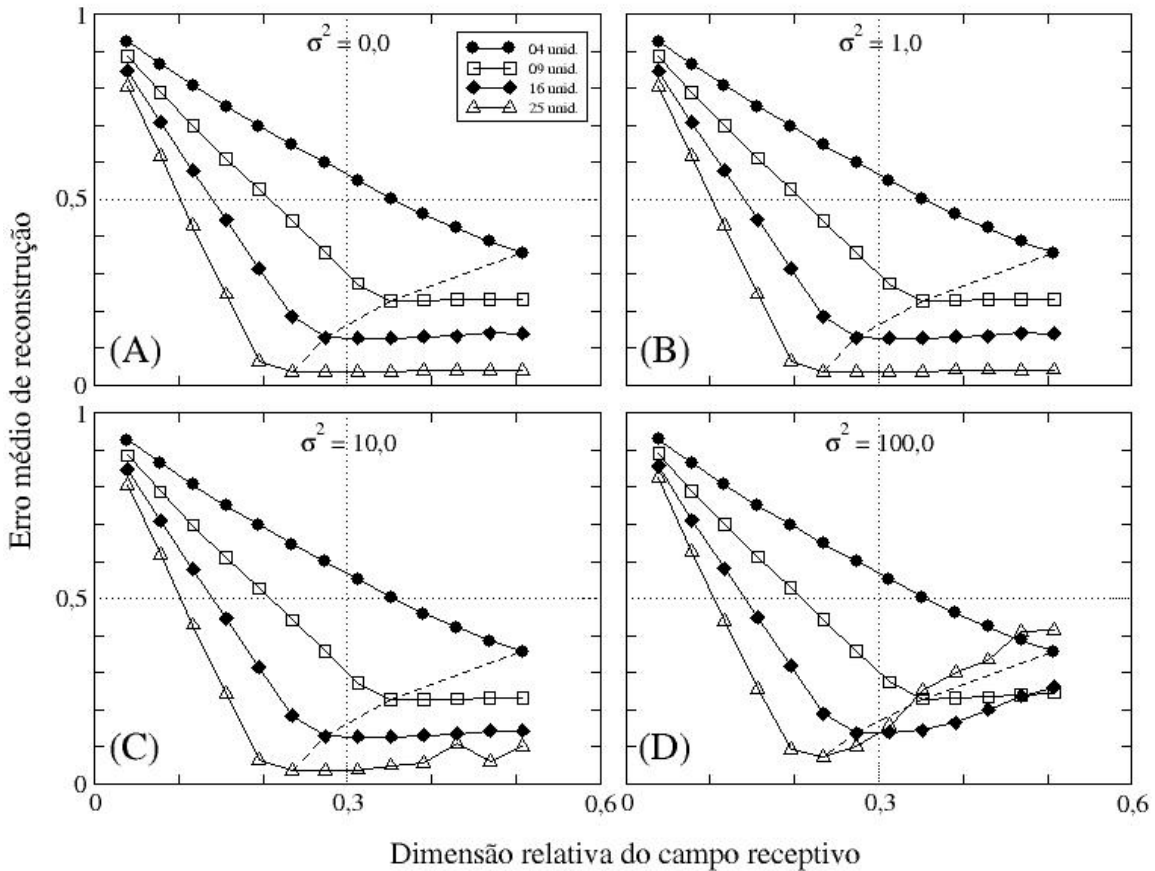


Fig. 4.6: Erro médio de reconstrução E pela variação do campo receptivo r_o . A rede foi treinada com cenas artificiais, e os gráficos (A), (B), (C) e (D) representam seu desempenho quando submetida a ruídos de natureza gaussiana com variância 0, 1, 10 e 100, respectivamente. A curva tracejada separa os pontos obtidos de campos sem sobreposição dos pontos vindos de campos com sobreposição. A escala dos eixos está normalizada em $E \times r_o/N$.

A semelhança entre imagens naturais e artificiais pode também ser localizada quando comparamos os desvios da média do gráfico 4.7 com os do gráfico 4.3, já discutidos na seção anterior. Além disto, uma rede treinada com retropropagação de erro é igualmente eficiente com as cenas artificiais e também se equivale quantitativamente à arquitetura de campos receptivos restritos.

O mesmo se verifica comparando as propriedades que são observadas visualmente. Diferente do primeiro *ensemble*, contudo, este ostenta *orientação* em virtualmente todos os casos. A *auto-similaridade*, por seu turno, é mais patente apenas nos exemplos 5, 6, 12 e 14. Estas propriedades em comum podem ajudar a compreender a semelhança no desempenho

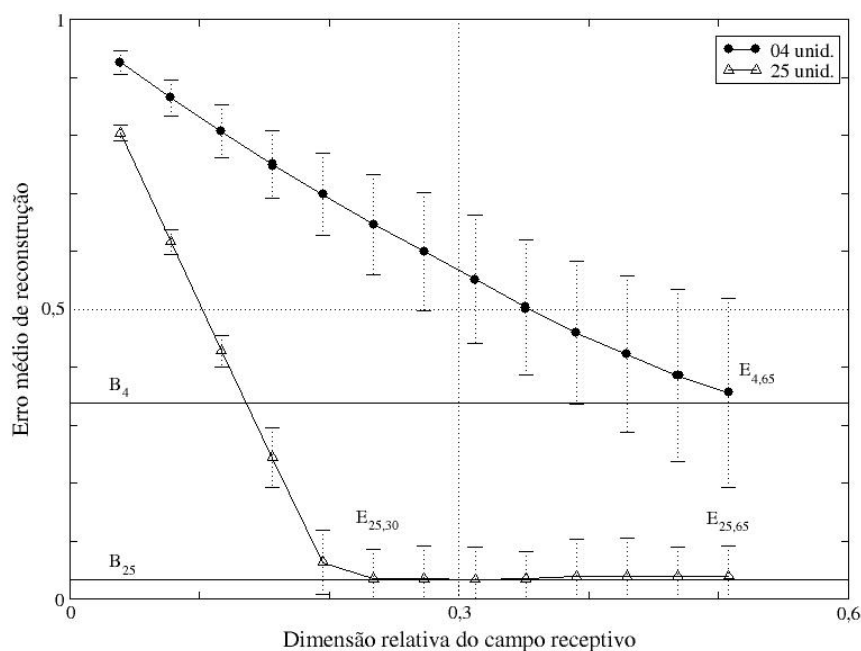


Fig. 4.7: Erro médio de reconstrução sobre cenas artificiais para uma rede com campos receptivos restritos (as duas curvas com barras de erro indicadas pela legenda) e para outra treinada com retropropagação de erro (as duas retas paralelas indicadas por B_M).

da reconstrução da rede, pois sugerem que uma rede treinada com cenas naturais não terá dificuldade maior na reconstrução de cenas artificiais do que a que teria para reconstruir as cenas naturais com as quais aprendeu. A confirmação disto é exibida nos gráficos (A) e (B) da figura 4.13, onde comparamos lado a lado o desempenho de uma rede normal e o de uma rede híbrida.

Reconstruindo nossas cenas artificiais, podemos observar aqui também o desempenho qualitativo. A figura 4.8 apresenta duas imagens reconstruídas: na primeira linha, expondo a 6ª cena, a reconstrução é bastante eficiente, seja nos pontos $E_{25,30}$ e $E_{25,65}$, seja no ponto $E_{25,30}$ da rede híbrida treinada com as *cenas naturais*. Na segunda linha, expondo a 22ª cena, percebemos um dos casos de desvio da média, onde a eficiência da rede cai em todas as reconstruções.

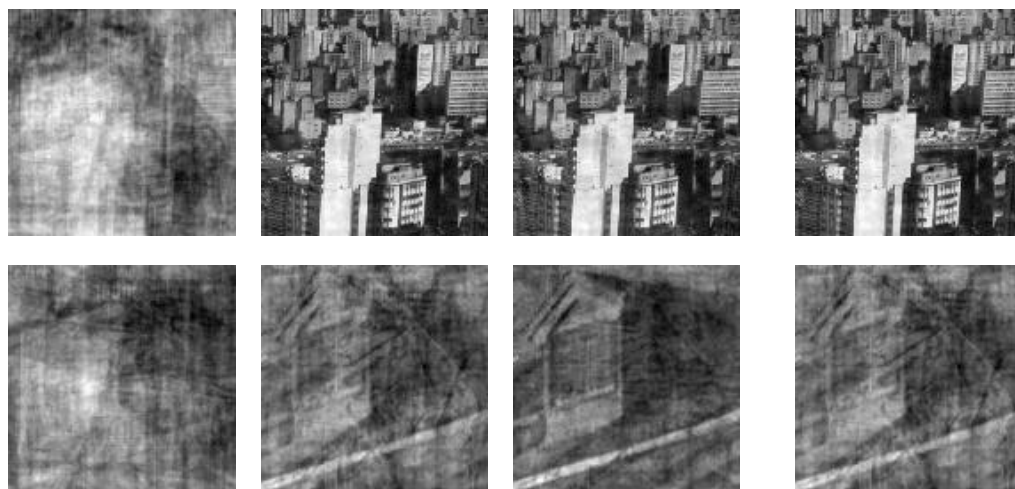


Fig. 4.8: As linhas superior e inferior contêm respectivamente as reconstruções das 6ª e 22ª cenas artificiais do ensemble $D.2$. Da esquerda para a direita, as imagens representam os pontos $E_{4,65}$, $E_{25,30}$ e $E_{25,65}$. A última coluna (em destaque), é a reconstrução das mesmas imagens por uma rede híbrida treinada com cenas naturais usando os parâmetros do ponto $E_{25,30}$.

4.2.3 Cenas Geradas por um Caminhante Aleatório

Apesar de a propriedade de *orientação* não se destacar facilmente nas imagens formadas por um caminhante aleatório bidimensional, a *auto-similaridade* é uma característica bastante marcante, predominando em todos os exemplos. A predominância de uma propriedade em detrimento da outra já traz uma certa distinção deste grupo em relação ao das cenas naturais e ao das artificiais – como poderá ser confirmado a partir da resposta fornecida pela rede.

O desempenho de uma rede treinada com a população de imagens geradas por um caminhante aleatório – de mesma natureza do exemplo mais à esquerda da figura D.3 – pode ser visto nos gráficos da figura 4.9, similares aos já analisados nas seções anteriores. De uma forma global, este modelo de rede mostra mais dificuldade para tratar com caminhantes aleatórios, tendo em vista que o erro médio de reconstrução manteve-se mais alto – fato este provavelmente explicado também pela existência de grande número de pixels com valor zero. A comparação entre estas curvas e suas análogas de mesma arquitetura dos casos anteriores mostra que a diferença é maior para M menor – o que vem confirmar novamente a dependência em relação ao número de unidades de saída.

A presença de ruído gaussiano nos quadrantes (A), (B) e (C) é essencialmente inócua na evolução das curvas, de modo que a rede não apresenta qualquer alteração de comporta-

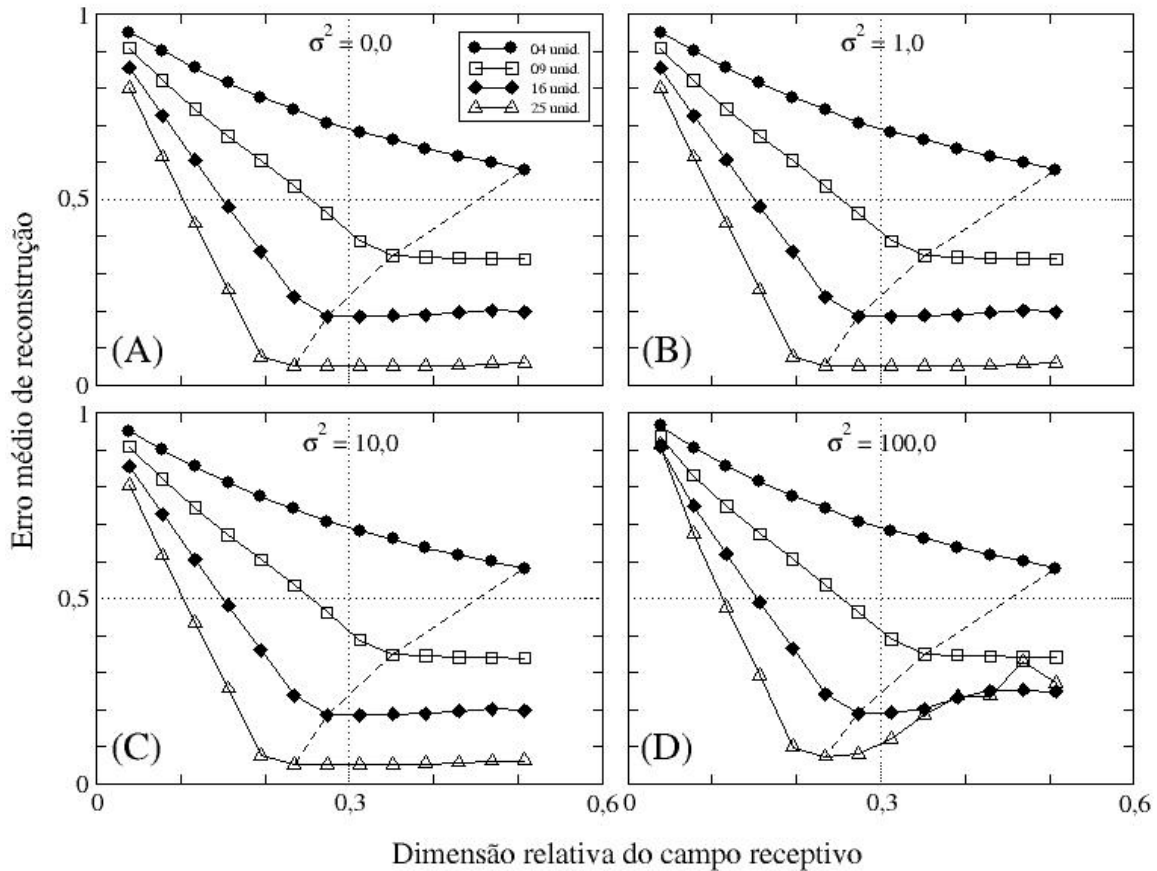


Fig. 4.9: Erro médio de reconstrução E pela variação do campo receptivo r_o . A rede foi treinada com cenas de um caminhante aleatório, e os gráficos (A), (B), (C) e (D) representam seu desempenho quando submetida a ruídos de natureza gaussiana com variância 0, 1, 10 e 100, respectivamente. A curva tracejada separa os pontos obtidos de campos sem sobreposição dos pontos vindos de campos com sobreposição. A escala dos eixos está normalizada em $E \times r_o/N$.

mento. No quadrante (D), os únicos destacamentos verificados estão nos pontos extremos das curvas para $M = 16$ e $M = 25$, como nas demais imagens. O gráfico 4.10, por sua vez, não apresenta novidades senão nas linhas da retropropagação de erro, cujo desempenho deixou de ser tão parecido com a arquitetura de campos receptivos. Se B_4 teve um resultado visivelmente melhor, B_{25} mostrou piora na sua eficiência quando comparada à dos *ensembles* anteriores.

A rede híbrida na figura 4.13 (C), treinada com cenas naturais e empregada para reconstruir cenas de caminhantes aleatórios, se comportou como a rede 4.9 (A), como se os vetores pesos, treinados sob cenas de natureza, tivessem aprendido com estas um conjunto adequado

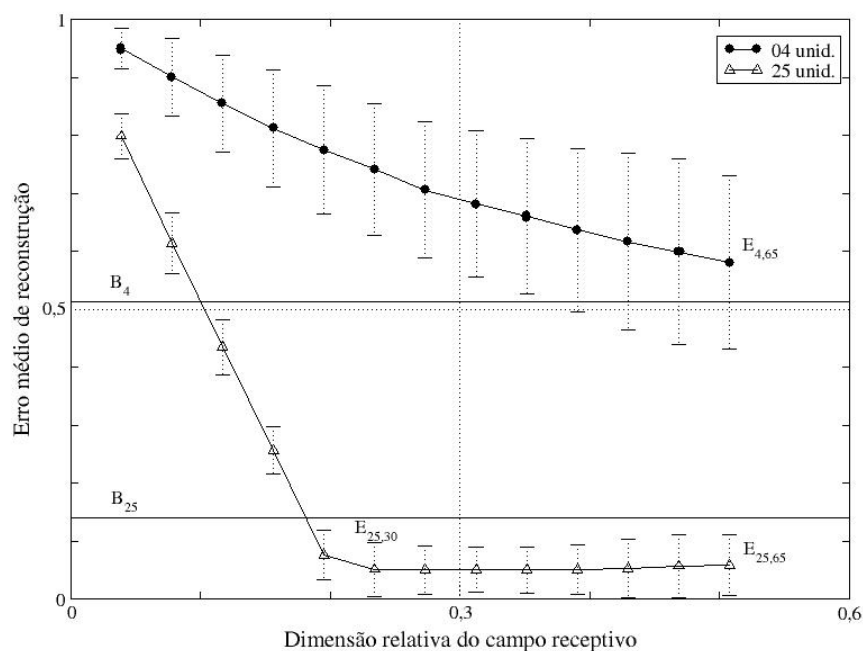


Fig. 4.10: Erro médio de reconstrução sobre caminhantes aleatórios para uma rede com campos receptivos restritos (as duas curvas com barras de erro indicadas pela legenda) e para outra treinada com retropropagação de erro (as duas retas paralelas indicadas por B_M).

e suficiente de características para simular a rede dos caminhantes aleatórios bidimensionais.

4.2.4 Cenas Fractais

Fractais, como foi apresentado na seção 3.3.1, são estruturas bastante sofisticadas que também apresentam propriedades bem específicas. Apesar de nem todos os tipos de fractais serem dotados de todas as propriedades que os caracterizam, normalmente são estigmatizados por sua *subdivisão infinita* e por sua *auto-similaridade estatística*. E por mais que a *orientação* não seja uma marca preponderante na natureza deles, são considerados como o meio artificial mais prático de construir imagens naturais [17].

A imagem central da figura D.3 corresponde a um exemplo de fractal cujo processo de geração é discutido no apêndice C. O *ensemble* de treinamento da rede neural conteve 30 padrões distintos de dimensão 128×128 pixels, com dimensão fractal $d_F = 2,5$, criados através do processo de geração de movimento browniano pelo método de substituição do ponto médio para duas dimensões.

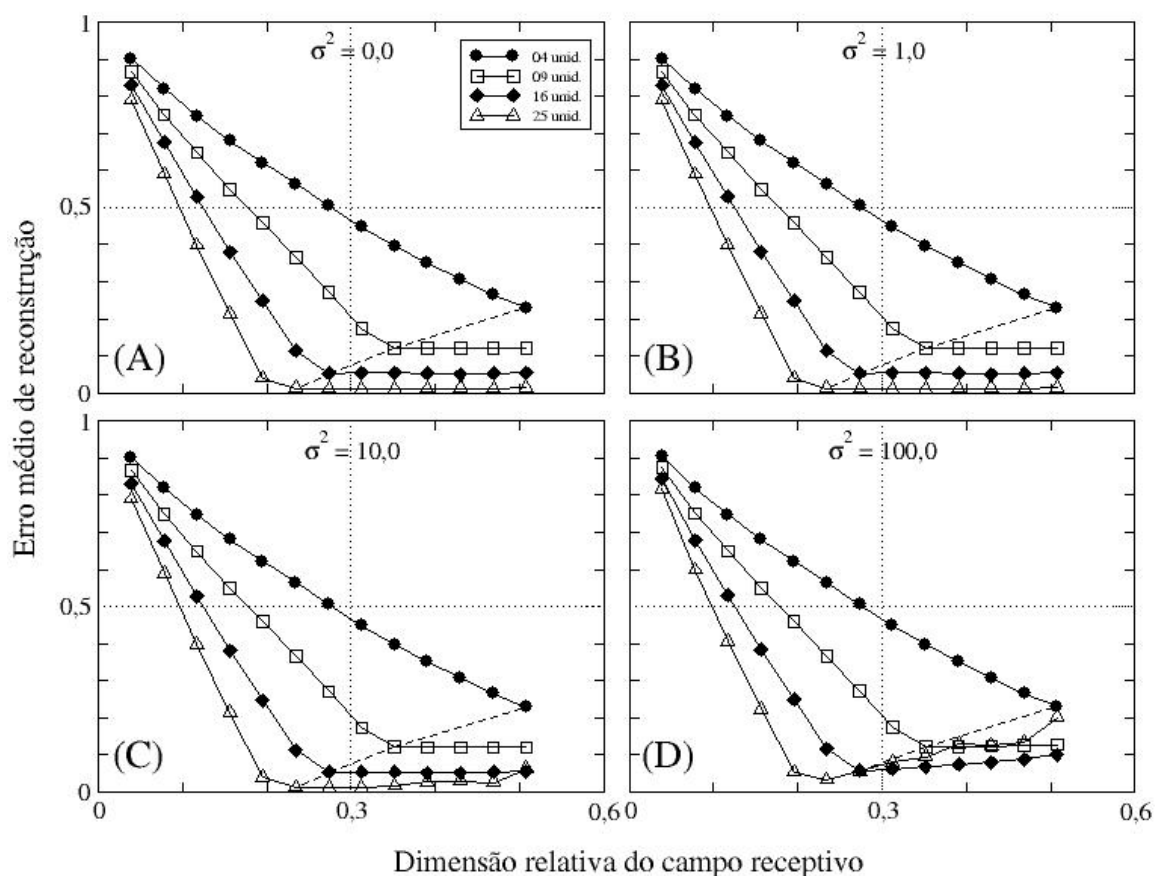


Fig. 4.11: Erro médio de reconstrução E pela variação do campo receptivo r_o . A rede foi treinada com cenas fractais, os gráficos (A), (B), (C) e (D) representam seu desempenho quando submetida a ruídos de natureza gaussiana com variância 0, 1, 10 e 100, respectivamente. A curva tracejada separa os pontos obtidos de campos sem sobreposição dos pontos vindos de campos com sobreposição. A escala dos eixos está normalizada em $E \times r_o/N$.

Nos 4 gráficos da figura 4.11, podemos perceber que uma rede treinada para reconstruir imagens fractais obtém melhores respostas do que os demais casos analisados. Além de as curvas de erros serem globalmente inferiores, também são mais refratárias à interferência do ruído. A origem disto pode estar no fato de que o *ensemble* fractal é composto por padrões mais homogêneos entre si do que os do *ensemble* natural ou os do artificial.

As curvas da figura 4.12 não acrescentam novidades para este caso, exceto pelo fato de os desvios da média da arquitetura para $M = 25$ revelarem-se os mais baixos. O aprendizado com retropropagação de erro também foi bastante eficiente e com valores bastante próximos, tanto para B_4 quanto para B_{25} . A rede híbrida do gráfico 4.13 (D), treinada com cenas

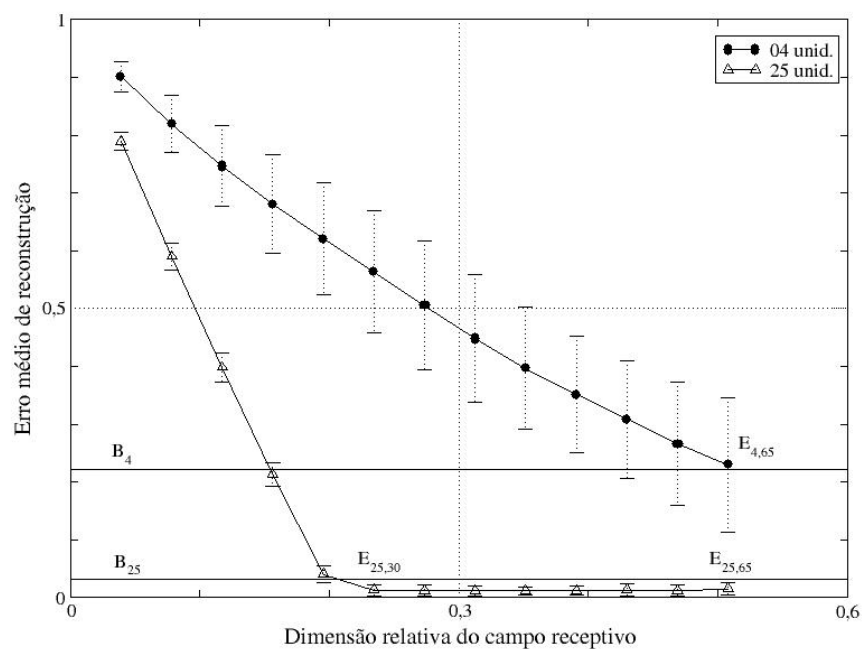


Fig. 4.12: Erro médio de reconstrução sobre cenas fractais para uma rede com campos receptivos restritos (as duas curvas com barras de erro indicadas pela legenda) e para outra treinada com retropropagação de erro (as duas retas paralelas indicadas por B_M).

naturais e reconstruindo fractais, não apresentou diferenças ao ser comparada com o gráfico 4.11 (A), tal qual o caso dos caminhantes aleatórios.

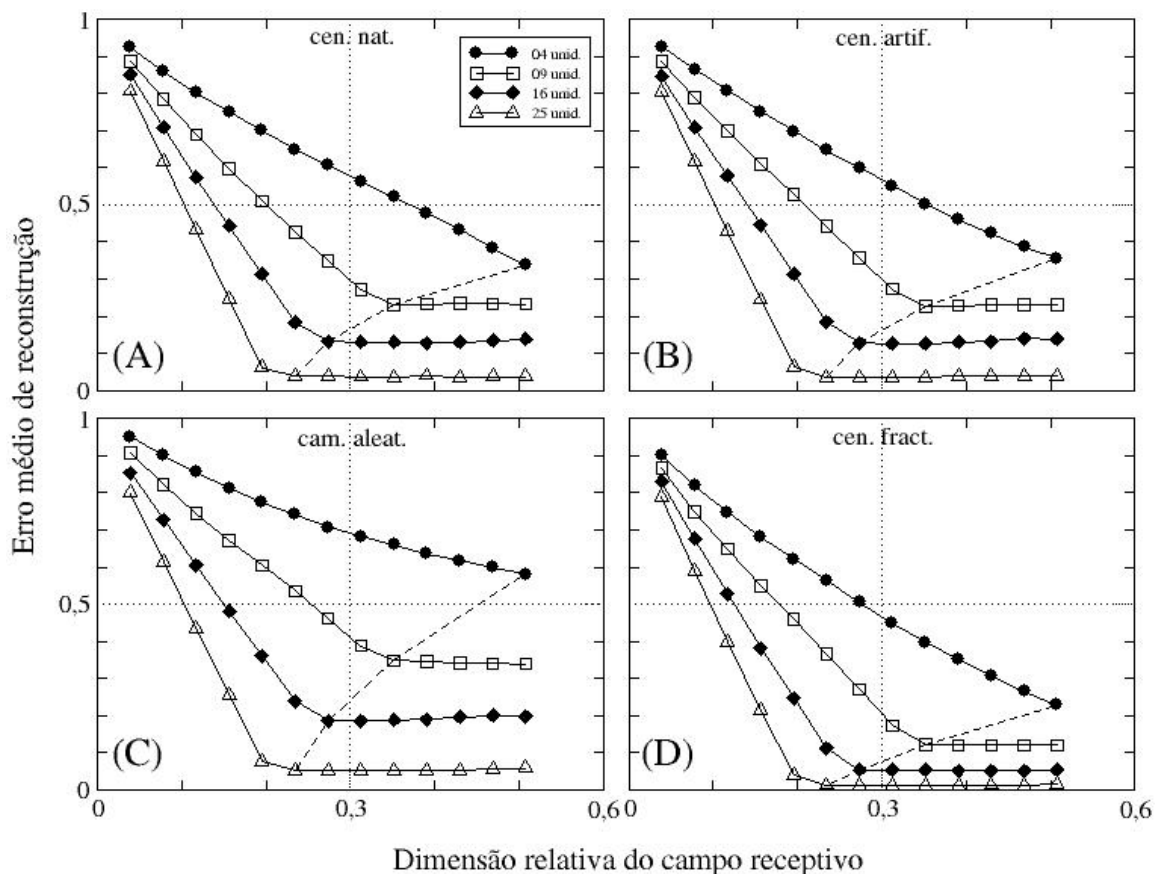


Fig. 4.13: Erro médio de reconstrução E pela variação do campo receptivo r_o . Em todos os casos, a rede foi treinada com cenas naturais, e os diversos quadrantes representam o desempenho da rede a partir da reconstrução de imagens de naturezas distintas e sem ruído: o gráfico (A) contém cenas naturais, o gráfico (B) contém cenas artificiais, o (C) contém cenas do caminhante aleatório, e o (D) contém cenas fractais.

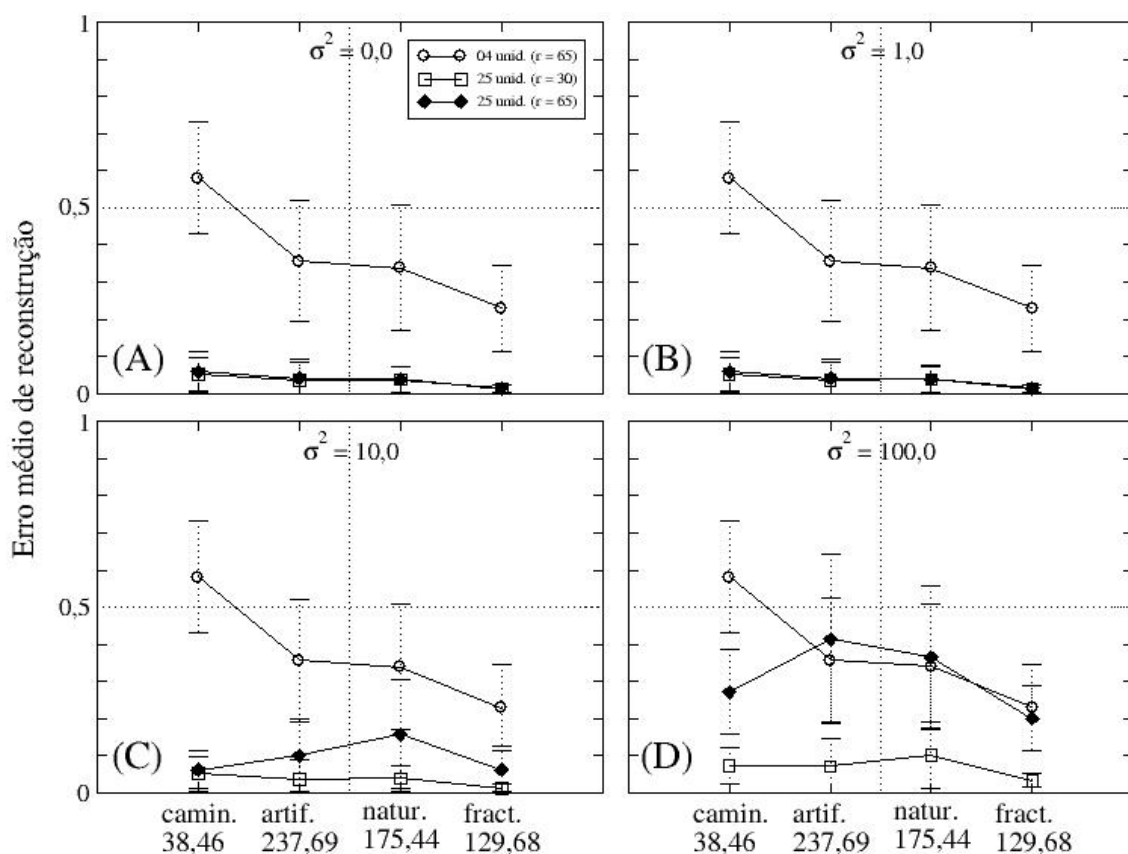


Fig. 4.14: Cada curva em cada gráfico apresenta o erro médio de reconstrução para uma arquitetura de rede neural distinta. O eixo das abscissas marca a natureza das imagens do ensemble trabalhado pela rede. Na escala do eixo é exibido também o valor médio da variância dos pixels.

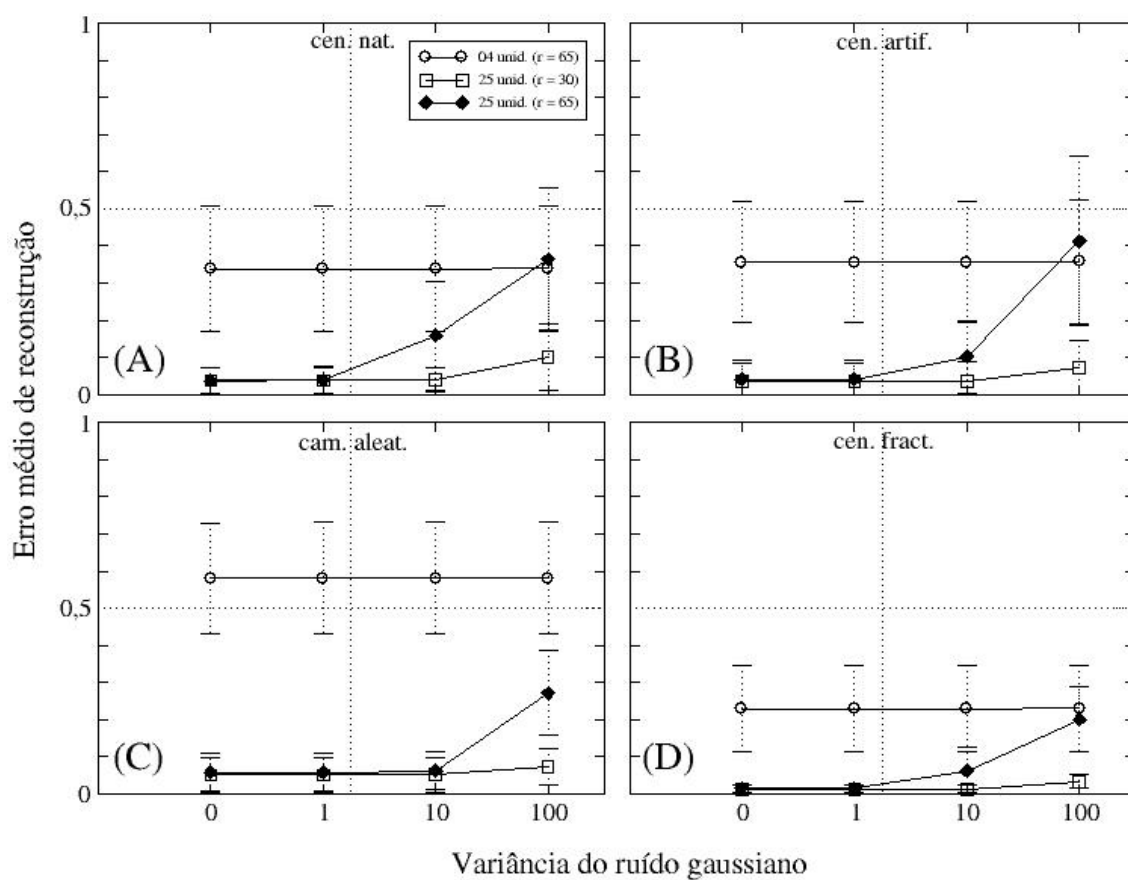


Fig. 4.15: Cada curva em cada gráfico apresenta o erro médio de reconstrução para uma arquitetura de rede neural distinta. O eixo das abscissas marca a variância σ^2 do ruído gaussiano ao qual a rede foi submetida no treinamento.

Conclusão

Conclusão

Ao longo de todo o texto, procuramos enfatizar muito as vantagens que o sistema biológico leva sobre os artificiais e os incontáveis métodos utilizados pelos pesquisadores no sentido de tentar aproximar-se deles através de técnicas artificiais. Vemos grande relevância neste estudo sob o ponto de vista científico, pois este tipo de sistema serve como modelo para processamento e codificação de estímulos sensoriais, tanto para sistemas artificiais quanto para os próprios sistemas biológicos. Em se tratando do cérebro, nossa máquina biológica mais poderosa, cremos que a simulação por redes neurais é uma das ferramentas mais fáceis e práticas, apresentando resultados muito promissores.

Muitos problemas de identificação de familiaridade ou de análise de componentes principais de um conjunto de padrões somente podem ser resolvidos computacionalmente por redes de neurônios quando tratados a partir de uma rede de múltiplas unidades de saída. Estas arquiteturas, contudo, normalmente operam através de regras de aprendizagem não-locais (como Sanger e Oja para múltiplas unidades), o que as torna pouco realistas do ponto de vista biológico. Os codificadores lineares N - M - N treinados pelo método de retropropagação de erro também são muito eficientes apresentando excelente desempenho. Apesar de também extraírem as M primeiras componentes principais da população de imagens, funcionam através de um aprendizado supervisionado não-local (diferente dos sistemas biológicos) e exigem muito tempo computacional.

Partindo do trabalho de Idiart *et al* [12], empregamos uma solução analítica para a determinação dos vetores pesos ótimos \vec{D}^a dos sinais reconstruídos \vec{e}^τ fazendo uso, para tanto, de uma abordagem local na qual os padrões de treinamento \vec{e}^τ são restringidos com campos receptivos. Neste modelo, cada unidade de saída $S^{a\tau}$ tem conexões limitadas sobre os terminais de entrada impedindo seu acesso completo. Podemos imaginar que cada unidade de saída representa uma rede independente, cada qual agindo sobre uma fração

de \vec{e}^T e aprendendo a partir de uma regra de aprendizagem local – como a regra de Oja para uma unidade, por exemplo. Este modelo não somente é uma alternativa às regras não-locais de aprendizado não-supervisionado para múltiplas unidades, como também agrega a vantagem de exigir menos tempo de processamento computacional em seu treinamento.

Os resultados provaram que a eficiência deste modelo está diretamente ligada à quantidade de informação que ele é capaz de enxergar. Esta quantidade de informação depende do número $M = M' \times M'$ de unidades de saída e da dimensão $\vec{r} = (r_x, r_y)$ dos campos receptivos restritos. Maior número de unidades de saída pressupõe mais informação e, por conseguinte, menor erro. O mesmo pode ser dito a respeito da dimensão do campo receptivo até o limite em que inicia sua sobreposição, ou seja, enquanto é verificada a relação $M'r_o \leq N$ (onde r_o pode representar tanto a orientação r_x quanto a r_y). Após este ponto, não existe acréscimo significativo de dados que leve a um aprendizado maior. A presença de ruído gaussiano, todavia, desde que com variância adequadamente escolhida, permite determinar um mínimo global no desempenho mais otimizado da rede para o caso em que $M'r_o \geq N$.

Trabalhamos com arquiteturas de rede que utilizaram 4 valores diferentes de unidades de saída: $M = 4, 9, 16$ e 25 . Para $M = 25$, a configuração mais eficiente, alcançamos desempenhos qualitativo e quantitativo muito bons. Isto foi verificado numericamente nos gráficos e visualmente nas reconstruções das imagens. Os erros calculados são bastante baixos, da ordem de 3%, e as imagens reconstruídas são bastante fiéis, o que parece provar que a estatística do sistema foi extraída eficientemente. Além disto, a confrontação com os resultados de um codificador com mesma arquitetura mostrou-se equivalente nas duas análises.

A configuração de 25 unidades também revelou-se suficiente do ponto de vista estatístico quando verificamos que no nosso conjunto de cenas naturais os primeiros 29 autovalores mais representativos da matriz correlação comportam aproximadamente 100% de seu traço. Isto, por si só, basta para se realizar uma estatística de dois pontos eficiente.

Os *ensembles* natural e artificial, os mais heterogêneos, exibiram erros de reconstrução com grandes desvios da média para M pequeno com r grande. O incremento de M para os mesmos valores de r proporcionou mais informação e permitiu que o erro se tornasse mais

representativo. O conjunto de cenas fractais, por sua vez, cujo processo de geração lhes concedeu uma homogeneidade toda particular, teve os menores desvios verificados.

As redes híbridas – redes treinadas com cenas naturais e empregadas na reconstrução de imagens de outra natureza – não manifestaram desempenho diferenciado das redes normais. Isto sugere que os pesos sinápticos das primeiras extraíram certas características que lhes permitiram generalizar seu aprendizado e reconstruir com a mesma eficiência esperada imagens de outros *ensembles*.

Em trabalhos futuros, esperamos explorar mais o *ensemble* de cenas naturais com o modelo de rede de campos receptivos restritos com o objetivo de testar seu comportamento em casos de maior número de padrões de imagens e com características menos heterogêneas entre si. Os desvios da média certamente diminuirão, e a extração de características se tornará mais representativa. Além disto, parece-nos interessante investigar se a rede é capaz de aprender mais incrementando M para valores superiores a 25, ou seja, 36 e 49. O propósito disto vem do fato de que a quantidade de componentes principais necessária, segundo o número de autovalores significativos da matriz correlação dos padrões de entrada, foi aqui 29 para todos os casos testados. A confirmação disto tem de ser verificada para novos *ensembles*.

Além disto, temos também interesse em investigar com mais cuidado o comportamento de redes híbridas. Sabemos que são eficientes na generalização de pesos obtidos com imagens que possuam características similares às das cenas naturais com as quais foram treinadas, mas desconhecemos seu comportamento se fossem submetidas a padrões absolutamente diversos.

Apêndice A

Componentes Principais: um Exemplo Simples

Com o propósito de ilustrar o processo de análise de componentes principais abordado na seção 2.2, aplicaremos neste apêndice um exemplo simplificado para um sistema bidimensional.

Seja ϵ^2 o espaço bidimensional dos padrões $\vec{e}^\tau = (e_1^\tau, e_2^\tau)$, como apresentado na figura A.1 (A). Este espaço está preenchido com um vetor sináptico $\vec{w}^0 = (1, 1)$ e 4 exemplos de treinamento com coordenadas $\vec{e}^1 = (1/2, 1/2)$, $\vec{e}^2 = (-1/2, 1)$, $\vec{e}^3 = (-1/2, -1/2)$ e $\vec{e}^4 = (1/2, -1)$. Os pontos foram escolhidos de tal forma a se aproximarem de uma distribuição gaussiana centrada (dados com média zero).

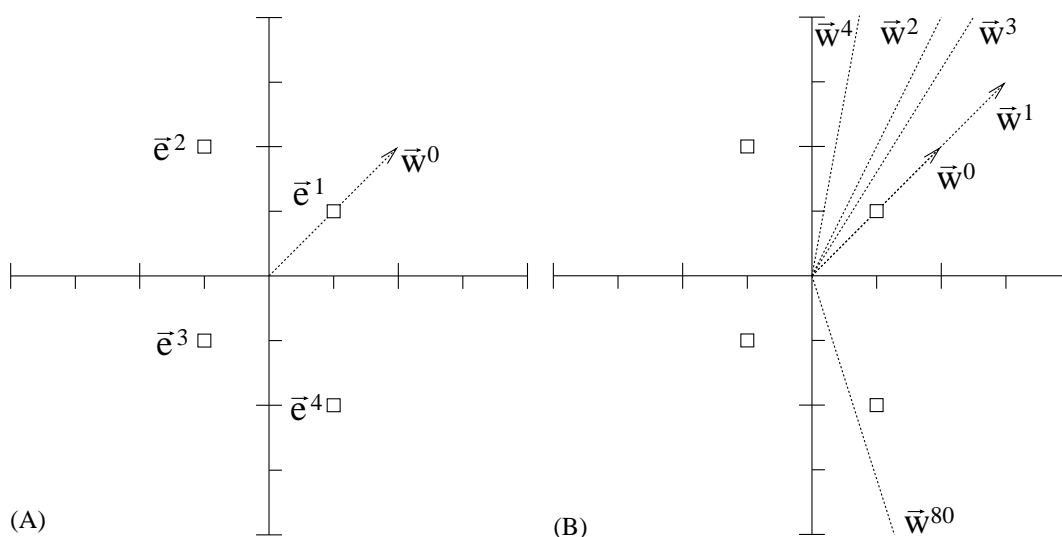


Fig. A.1: Espaço de ϵ^2 padrões de treinamento para um percéptron simples linear com regra de aprendizagem de Hebb. No lado (A), os padrões (quadrados vazados) e o vetor sináptico w^0 , que está em sua posição inicial; no lado (B), a evolução do crescimento do vetor sináptico ao longo das iterações do processo de aprendizagem da rede. Somente w^0 e w^1 aparecem com a magnitude original.

Extrairemos a primeira componente principal desta distribuição empregando um método iterativo sobre o *percéptron* simples linear

$$S^\tau = \vec{w}^{\tau-1} \cdot \vec{e}^\tau \quad (\text{A.1})$$

e a regra de aprendizagem de Hebb

$$\Delta \vec{w}^\tau = S^\tau \vec{e}^\tau, \quad (\text{A.2})$$

tal que a taxa de aprendizagem foi assumida com valor 1. O processo de treinamento da rede se dá do modo como está descrito na seção 2.1, ou seja, alternando a aplicação das expressões (A.1) e (A.2).

Os valores pertinentes às 4 primeiras iterações da rede são exibidos na tabela A.1. Em sua quarta coluna, encontramos a evolução quantitativa do vetor peso, ou seja, \vec{w}^1 , \vec{w}^2 , \vec{w}^3 e \vec{w}^4 . Esta evolução está graficamente representada na figura A.1 (B). A quinta coluna apresenta os respectivos módulos.

τ	S^τ	$\Delta \vec{w}^\tau$	\vec{w}^τ	w^τ	θ^τ
1	1,000	(0,500; 0,500)	(1,500; 1,500)	2,121	45,000°
2	0,750	(-0,375; 0,750)	(1,125; 2,250)	2,516	63,435°
3	-1,688	(0,844; 0,844)	(1,969; 3,094)	3,667	57,529°
4	-2,109	(-1,055; 2,109)	(0,914; 5,203)	5,283	80,036°

Tab. A.1: Valores correspondentes às 4 primeiras iterações no processo de treinamento utilizando-se a regra de aprendizagem de Hebb. A última coluna apresenta a evolução do ângulo do vetor peso à medida em que o aprendizado é realizado.

Finalizando 5 iterações para cada um dos exemplos de treinamento (o que resulta em 20 iterações), o módulo do vetor peso alcança o valor 5.928,4 com inclinação de $-72,616^\circ$; finalizando 10, o módulo mede aproximadamente 4×10^7 fazendo ângulo de $-72,272^\circ$; por fim, ao finalizar 20 (somando 80 iterações), temos um vetor peso com comprimento de $1,823 \times 10^{15}$ e ângulo de $-72,271^\circ$.

O vetor \vec{w}^{80} posicionou-se de modo a indicar a orientação que melhor representa a direção de máxima variância da distribuição dos padrões de treinamento. A ausência de alteração expressiva na inclinação do vetor indica que ele está oscilando em torno de uma posição ótima e expressa a finalização do aprendizado da rede. Este aprendizado já se deu desde pelo menos a vigésima iteração; no entanto, como a regra de Hebb não informou à rede

que o treinamento foi suficiente, as iterações seguiram. Em outras palavras, o vetor peso continuou sendo incrementado chegando a um valor da ordem de 10^{15} .

Apêndice B

Erro Quadrado Médio

Neste apêndice, nos utilizaremos da definição do erro quadrado médio E introduzida na seção 2.3 pela expressão

$$E = \frac{1}{2} \left\langle \left(\frac{\vec{e}}{|\vec{e}|} - \frac{\tilde{e}}{|\tilde{e}|} \right)^2 \right\rangle \quad (\text{B.1})$$

objetivando determinar vetores pesos ótimos \vec{D}^a para a reconstrução \tilde{e} dos padrões de entrada segundo o modelo proposto por Idiart *et al* [12]

$$\tilde{e} = \sum_{a=1}^M S^a \vec{D}^a, \quad (\text{B.2})$$

tal que S^a são as unidades de saída da rede. Uma solução analítica é encontrada quando assumimos que trabalhamos com um sistema *auto-mediante*. Neste caso, é possível simplificar a expressão (B.1) na razão entre duas médias dada por

$$E = \frac{\langle |\vec{e} - \tilde{e}|^2 \rangle}{\langle |\vec{e}|^2 \rangle}. \quad (\text{B.3})$$

Inicialmente, expandimos o numerador de (B.3) em

$$\langle |\vec{e} - \tilde{e}|^2 \rangle = \langle \vec{e} \cdot \vec{e} \rangle - 2\langle \vec{e} \cdot \tilde{e} \rangle + \langle \tilde{e} \cdot \tilde{e} \rangle. \quad (\text{B.4})$$

O primeiro termo de (B.4), que é igual ao denominador de (B.3), pode ser desenvolvido de modo a provarmos que corresponde ao traço da matriz correlação (2.11) – que por sua vez, devido ao *teorema da conservação do traço*, é igual à soma dos autovalores da matriz

$$\langle \vec{e} \cdot \vec{e} \rangle = \sum_{i=1}^N \langle e_i e_i \rangle = \sum_{i=1}^N C_{ii} = \text{Tr } \mathbf{C} = \sum_{i=1}^N \lambda_i. \quad (\text{B.5})$$

O termo central fica

$$-2\langle \vec{e} \cdot \tilde{e} \rangle = -2 \sum_a \langle \vec{e} \cdot \vec{D}^a S^a \rangle = -2 \sum_a \vec{D}^a \cdot \langle \vec{e} S^a \rangle = -2 \sum_a \vec{D}^a \cdot \vec{L}^a, \quad (\text{B.6})$$

ao adotarmos a definição $\vec{L}^a = \langle \vec{e} S^a \rangle$, enquanto o último termo fica

$$\begin{aligned} \langle \vec{e} \cdot \vec{e} \rangle &= \left\langle \sum_i \tilde{e}_i \tilde{e}_i \right\rangle = \sum_i \left\langle \sum_a S^a D_i^a \sum_b S^b D_i^b \right\rangle \\ &= \sum_{a,b} \sum_i D_i^a D_i^b \langle S^a S^b \rangle = \sum_{a,b} \vec{D}^a \cdot \vec{D}^b Q^{ab}. \end{aligned} \quad (\text{B.7})$$

Aqui, também definimos $Q^{ab} = \langle S^a S^b \rangle$ como a correlação das unidades de saída. A escolha do valor ótimo de \vec{D} é feita através da minimização do erro quadrado médio (B.3) derivando-o em relação a D^a

$$\frac{\partial E}{\partial D^a} = \frac{1}{\langle \vec{e} \cdot \vec{e} \rangle} \left[-2 \left\langle \vec{e} \cdot \frac{\partial \vec{e}}{\partial D^a} \right\rangle + 2 \left\langle \vec{e} \cdot \frac{\partial \vec{e}}{\partial D^a} \right\rangle \right] = 0. \quad (\text{B.8})$$

Inserimos (B.2)

$$\left\langle \left(\sum_b S^b \vec{D}^b \right) S^a \right\rangle - \langle \vec{e} S^a \rangle = 0 \quad (\text{B.9})$$

e empregamos as definições de Q^{ab} e de \vec{L}^a . Então,

$$\sum_b \vec{D}^b Q^{ab} = \vec{L}^a, \quad (\text{B.10})$$

de modo que o erro quadrado médio será minimizado quando escolhermos

$$\vec{D}^a = \sum_{b=1}^N [Q^{-1}]^{ab} \vec{L}^b. \quad (\text{B.11})$$

Aplicamos (B.11) em (B.7)

$$\langle \vec{e} \cdot \vec{e} \rangle = \sum_a \vec{D}^a \cdot \sum_{b,c} Q^{ab} [Q^{-1}]^{bc} \vec{L}^c = \sum_a \vec{D}^a \cdot \sum_{b,c} \delta^{ac} \vec{L}^c = \sum_a \vec{D}^a \cdot \vec{L}^a. \quad (\text{B.12})$$

Este resultado e os de (B.5) e (B.6) são juntados em (B.3), de modo que o erro quadrado médio se torna

$$E = 1 - \frac{\sum_{a=1}^M \vec{D}^a \cdot \vec{L}^a}{\sum_{i=1}^N \lambda_i}. \quad (\text{B.13})$$

A representação mais redundante se dá quando cada campo receptivo abarca todas as entradas (as componentes do vetor \vec{e}), ou seja, quando $r = N$. Neste caso, todas as unidade desenvolvem pesos sinápticos \vec{w}_a iguais entre si e orientados ao longo da primeira componente principal. Para qualquer unidade a , $\vec{w}_a = \vec{v}_1$, onde \vec{v}_1 é o autovetor relacionado ao maior autovalor da matriz correlação completa dos padrões de entrada em (2.11). Da equação (B.13), obtemos

$$E = 1 - \frac{M \lambda_1^2}{(M \lambda_1 + \sigma^2) \sum_{i=1}^N \lambda_i}, \quad (\text{B.14})$$

que se aproxima de um limite finito à medida que $M \rightarrow \infty$ devido à redundância da representação.

Consideramos, agora, uma representação reduzida otimizada de dimensão M de uma rede que usa como componentes principais os M autovetores com o maior autovalor, ou seja, $\vec{w}_a = \vec{v}_a$ para $a = 1, 2, \dots, M$ – onde \vec{w}_a , vale lembrar, é um vetor N -dimensional. O erro quadrado médio correspondente é

$$E = 1 - \frac{\sum_a^M \lambda_a^2 / (\lambda_a + \sigma^2)}{\sum_i^N \lambda_i}. \quad (\text{B.15})$$

Aqui, o erro depende da percentagem do traço da matriz correlação representada pelos M maiores autovalores. Se o ruído for nulo, temos que $\sigma^2 = 0$, e a expressão (B.15) se reduz para

$$E = 1 - \frac{\sum_a^M \lambda_a}{\sum_i^N \lambda_i}. \quad (\text{B.16})$$

Apêndice C

Modelos Matemáticos de Fractais

A descrição a seguir foi retirada do livro de Mandelbrot *et al* dado na referência [17], material com proposta meramente informativa onde não encontramos grande preocupação com rigor matemático. Nossa intenção também está distante do detalhamento do formalismo, restringindo nosso escopo tão somente à implementação computacional dos algoritmos que nos foram oferecidos.

Seja $P(m, L)$ a probabilidade de m pontos dentro de um hipercubo (ou hiperesfera) no espaço E -dimensional de tamanho L centrado em um ponto arbitrário S de coordenada $\vec{x} = (x_1, \dots, x_E)$. $P(m, L)$ é normalizada, ou seja,

$$\sum_{m=1}^N P(m, L) = 1 \quad \text{para todo } L. \quad (\text{C.1})$$

As grandezas de interesse são obtidas dos momentos da distribuição: a *dimensão de massa*

$$M(L) = \sum_{m=1}^N mP(m, L), \quad (\text{C.2})$$

o *número de caixas* de tamanho L necessários para cobrir S ,

$$N_{\text{caixas}}(L) = \sum_{m=1}^N \frac{1}{m} P(m, L), \quad (\text{C.3})$$

e a *entropia configuracional* quando o espaço é dividido em cubos de tamanho L

$$S(L) = \sum_{m=1}^N \log(m) P(m, L). \quad (\text{C.4})$$

Para um *fractal uniforme* ou *conjunto fractal* com $N \rightarrow \infty$, a distribuição toma a forma $P(m, L) \rightarrow P(m/L^D)$, e todos os momentos resultam no mesmo D . Para um *fractal não uniforme* (o mapa de Poincarè, por exemplo), os momentos podem ter diferentes valores.

Quando amostras e formas exigem mais de uma dimensão, são chamadas *multifractais*. Além disto, um elemento aleatório adicional pode ser incluído permitindo a simulação de fenômenos naturais.

O *movimento browniano unidimensional* (MB1 ou “*brown noise*”) é o fractal aleatório mais simples, sendo o coração das generalizações que seguem. Seja, pois, $X(t)$ o processo aleatório dependente do tempo t , a variável independente e real. Sendo assim, definimos o *incremento do processo aleatório* com distribuição gaussiana como

$$X_{21} = X(t_2) - X(t_1), \quad (\text{C.5})$$

assim como definimos também a proporcionalidade

$$\sigma_{X_{2-1}}^2 \propto |t_2 - t_1|, \quad (\text{C.6})$$

tal que $\sigma_{X_{2-1}}^2$ é a variância do incremento (C.5).

Podemos escrever também que

$$X(t_0 + t) - X(t_0) \doteq \frac{1}{\sqrt{r}} [X(t_0 + rt) - X(t_0)], \quad (\text{C.7})$$

onde o operador \doteq indica que os incrementos de X são *estatisticamente auto-similares* por terem a mesma função distribuição de dimensão finita¹ para todo o t_0 e $r > 0$. Em outras palavras, são *estatisticamente indistinguíveis*.

Comentamos a seguir dois métodos para gerar *movimento browniano unidimensional*.

Integral do Ruído Branco

A integral do ruído branco descorrelacionado W é um método que satisfaz (C.5) e (C.6).

$$X(t) = \int_{-\infty}^t W(s) ds. \quad (\text{C.8})$$

As variáveis aleatórias W são descorrelacionadas e têm a mesma distribuição normal. Além disto, o gráfico de um trecho do movimento browniano $X(t)$ tem uma dimensão fractal de 1,5, e a intersecção da curva com uma linha horizontal tem uma dimensão de 0,5.

¹ Do inglês, “*finite dimensional joint distribution functions*”.

Substituição do Ponto Médio

Sejam o intervalo temporal $t \in [0, 1]$ e o processo aleatório $X(t)$ ajustado para $X(0) = 0$ e $X(1) = \chi$, onde χ é um valor aleatório com distribuição gaussiana, com média zero e variância $\sigma_{X_{1-0}}^2 \equiv \sigma^2$. Assim, esperamos que

$$\sigma_{X_{t_2-t_1}}^2 = |t_2 - t_1| \sigma^2 \quad (\text{C.9})$$

contanto que $0 \leq t_1 \leq t_2 \leq 1$.

Ajustamos $X(1/2)$ como a média aritmética entre $X(0)$ e $X(1)$ mais um valor gaussiano aleatório D_1 com média 0 e variância Δ_1^2

$$X(1/2) - X(0) = \frac{1}{2}[X(1) + X(0)] + D_1. \quad (\text{C.10})$$

Tendo que todo o lado direito tem média zero, o lado esquerdo também terá. Por conseguinte, o incremento $X(1) - X(1/2)$ também terá média zero.

De (C.9) e (C.10)

$$\sigma_{X_{1/2-0}}^2 = \left| \frac{1}{2} - 0 \right| \sigma^2 = \frac{1}{2} \sigma^2 \quad (\text{C.11})$$

$$= \left(\frac{1}{2} \right)^2 \sigma_{X_{1+0}}^2 + \Delta_1^2, \quad (\text{C.12})$$

de modo que $\Delta_1^2 = 1/4 \sigma^2$. A construção feita em (C.12) é somente possível quando $X(t)$ para um t qualquer for descorrelacionado.

Segue-se com o próximo nível $X(1/4) - X(0) = 1/2[X(1/2) + X(0)] + D_2$, onde, novamente, $[X(1/2) - X(1/4)]$ e $[X(1/4) - X(0)]$ são gaussianos com média zero. De

$$\sigma_{X_{1/4-0}}^2 = \left(\frac{1}{2} \right)^2 \sigma_{X_{1/2+0}}^2 + \Delta_2^2 = 1/4 \sigma^2,$$

tiramos $\Delta_2^2 = 1/8 \sigma^2$. A idéia volta a ser aplicada em $X(3/4)$.

Refinando este processo, obtemos

$$\Delta_n^2 = \frac{1}{2^{n+1}} \sigma^2 \quad (\text{C.13})$$

para a variância de D_n . Então, para diferenças de tempo $\Delta t = 2^{-n}$, adicionamos um elemento aleatório de variância $2^{-(n+1)} \sigma^2$, que é, como esperado, proporcional a Δt .

A generalização para parâmetros $0 < H < 1$ é chamada *movimento browniano fracional*. O parâmetro H descreve a “rugosidade” da função a pequenas escalas. Para o movimento

fractal browniano (fBm²), teremos a dimensão fractal $d_F = D = 2 - H$, tal que podemos identificar três categorias de fBm de acordo com os valores de H . Se $H > 0,5$, existe correlação positiva entre os incrementos, isto é, se o gráfico de X está aumentando para algum t_0 , então aumentará mais para $t > t_0$. Se $H = 0,5$, o movimento browniano tem incrementos independentes no sentido probabilístico, ou seja, tem a correlação nula. Se $H < 0,5$, a correlação dos incrementos é negativa, e as curvas parecem oscilar mais erráticamente.

Analogamente ao caso anterior da equação (C.7) onde $H = 0,5$, dizemos que os incrementos de X são estatisticamente auto-similares com H

$$X(t_0 + t) - X(t_0) \doteq \frac{1}{r^H} [X(t_0 + rt) - X(t_0)]. \quad (\text{C.14})$$

Assim, fBm “acelerados” $X(rt)$ são apropriadamente reescalados com a divisão pela amplitude r^H . As substituições de ponto médio D_n têm variâncias

$$\Delta_n^2 = \frac{\sigma^2}{(2^n)^{2H}} (1 - 2^{2H-2}). \quad (\text{C.15})$$

Por outro lado, a relação entre as variâncias σ^2 e $\sigma_{X_{t_2-t_1}}^2$ para um intervalo de tempo qualquer $t_2 - t_1$ fica

$$\sigma_{X_{t_2-t_1}}^2 = |t_2 - t_1|^{2H} \sigma^2, \quad (\text{C.16})$$

que infelizmente *não* aproxima adequadamente os pontos gerados aos do movimento browniano fracional. Ou seja, este processo não tem incrementos estacionários; os tempos t não são todos equivalentes estatisticamente. Para corrigir isto, usamos o método da *adição aleatória sucessiva*, no qual o D_n é adicionado para todos os pontos e não somente os pontos médio.

² Do inglês, “*fractal Brownian motion*”.

Apêndice D

Ensembles de Imagens

Este apêndice foi reservado para a descrição dos *ensembles* de imagens utilizados como sinal de entrada, para treinamento e reconstrução, de uma rede neural linear sujeita a campos receptivos restritos. Ainda que o foco deste trabalho seja a estatística de cenas naturais, cinco populações adicionais foram incluídas com o propósito de oferecer resultados comparativos.

As cenas de natureza foram escolhidas de forma a não conter objetos ou estruturas feitas pelo homem, a evitar horizontes e também grandes espaços vazios, como o céu ou o mar. Procuramos utilizar diversas escalas do ambiente, desde florestas vistas à distância a até uma única planta em destaque. Com isto, esperávamos abranger mais explicitamente a propriedade de auto-similaridade. O *ensemble*, apresentado na figura D.1, é constituído por 30 imagens centradas com dimensão 128×128 pixels e 256 níveis de cinza. Podem ser observados exemplos singulares intencionalmente incluídos para testar a robustez da rede e permitir que ela aprenda com casos particulares. De cima para baixo e da esquerda para a direita, podemos citar a imagem 4, na qual se verifica a lateral do tronco de uma árvore dominando praticamente todo o lado direito da mesma. Nos exemplos 10 e 15, uma porção bastante grande da imagem é tomada pela água, enquanto o 16 é totalmente composto por nuvens.

O primeiro *ensemble* comparativo é o de *cenas artificiais*, assim definidas por serem essencialmente constituídas por estruturas e objetos construídos pelo homem. Também é composto por 30 padrões centrados diferentes entre si com dimensão 128×128 pixels e 256 níveis de cinza, como pode ser visto na figura D.2.

O conceito de *aproximações de imagens naturais* introduzido na seção 3.3 viabiliza o



Fig. D.1: *População de 30 imagens composta por cenas naturais com dimensão 128×128 pixels cada.*

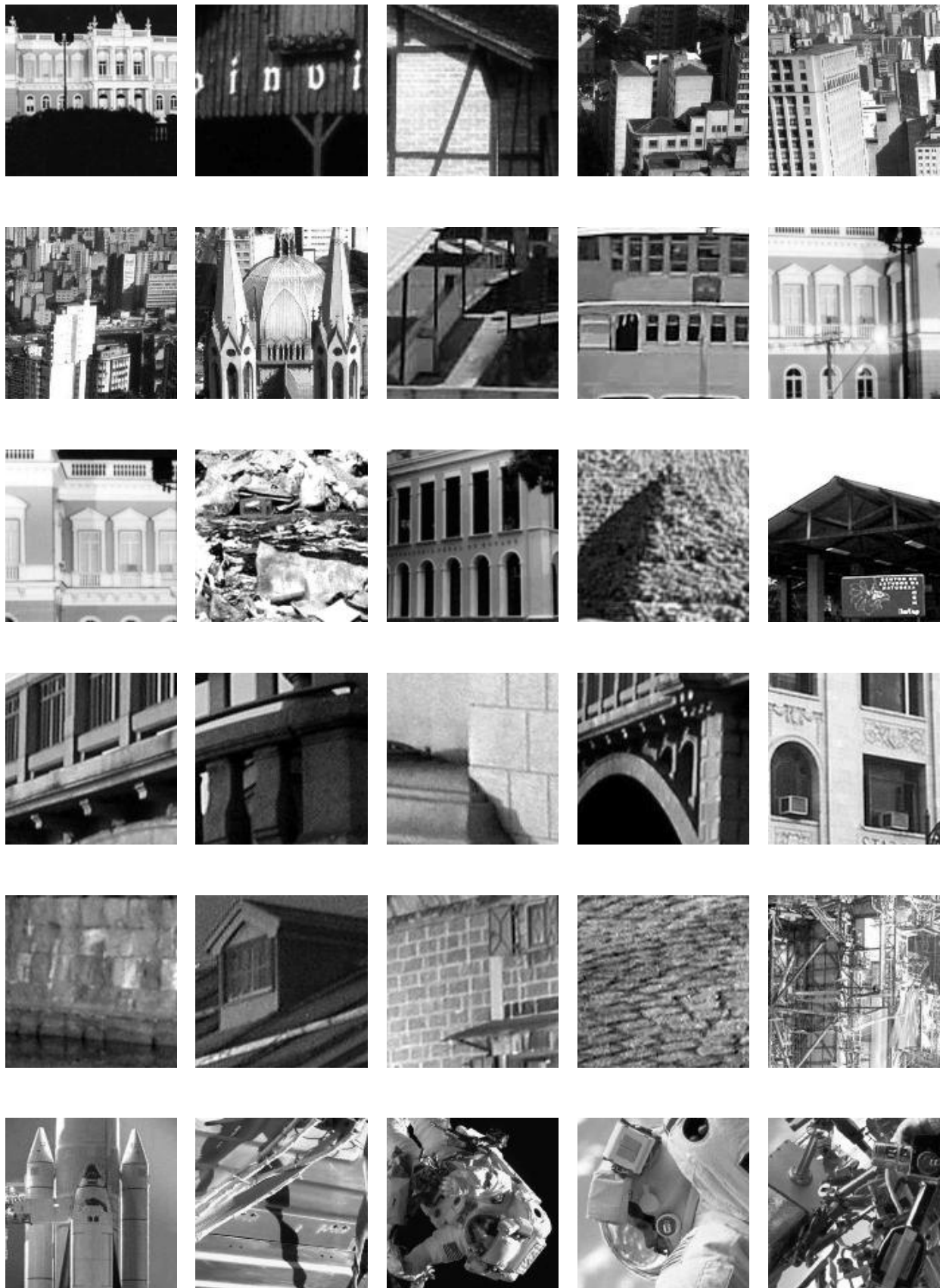


Fig. D.2: População de 30 imagens composta por cenas artificiais (estruturas criadas pelo homem) com dimensão 128×128 pixels cada.

estudo e a análise de duas populações de imagens que, ainda que geradas a partir de um algoritmo computacional – sendo, portanto, *sintéticas* –, assemelham-se visual e estatisticamente às cenas naturais.

Na seção 3.3.2, mostramos que é possível utilizar um procedimento simples como o *caminhante aleatório bidimensional* para gerar imagens nada elementares dentro da geometria euclidiana. Partindo de um ponto do plano decidido ao acaso, uma partícula avança, um passo de cada vez, e com igual probabilidade, para qualquer um dos quatro pontos que compõem diretamente sua vizinhança, ou seja, norte, sul, leste ou oeste. Cada vez que ocupa uma casa, à sua coordenada é incrementado um valor constante tal que seja possível mapear o rastro de sua passagem. Se a partícula alcança uma das bordas limites, as condições de contorno garantem a continuidade de sua caminhada a partir da borda oposta. O resultado deste mapeamento é uma matriz numérica que pode ser visualizada como uma imagem complexa análoga ao primeiro exemplo encontrado na figura D.3. Utilizamos nesta pesquisa imagens centradas com dimensão de 128×128 pixels em tons de cinza.

As cenas fractais tratadas na seção 3.3.1 são mais sofisticadas, seja na parametrização necessária à sua criação, seja na descrição de suas propriedades. 30 modelos de imagens centradas com dimensão 128×128 pixels e tons de cinza, como ilustrado no segundo exemplo da figura D.3, foram empregados no treinamento da rede. Todos os exemplos têm dimensão $d_F = 2,5$ e foram criados sem somas aleatórias, com desvio padrão igual a 1 e através do processo de geração de movimento browniano pelo método de substituição do ponto médio para duas dimensões. Detalhes estão disponíveis no apêndice C.

A terceira imagem em D.3 é um exemplo do resultado da filtragem logarítmica aplicada sobre o primeiro padrão do *ensemble* D.1. O filtro aplicado é a função de reescalamto logarítmico da intensidade do pixel sugerida por Ruderman [25] através da equação (3.5). Conquanto tenhamos utilizado este filtro para criar um novo conjunto de imagens a partir das cenas naturais, estas não foram utilizadas para o treinamento das redes, mas tão somente para o estudo de seu espectro de potência.

A imagem mais à direita da figura D.3 exemplifica um *ruído branco*. O ruído branco é um sinal que é absolutamente aleatório, ou seja, não privilegia qualquer frequência. Para gerar esta imagem, criamos uma matriz com 128 linhas e 128 colunas contendo elementos

númericos entre 0 e 1 descorrelacionados uns dos outros. Para tanto, faz-se necessário o uso de um algoritmo computacional que simule a geração de números aleatórios. A validade estatística deste sinal, para todos os propósitos práticos de nossos estudos, também restringiu-se unicamente à análise do espectro de potência em comparação com o dos demais *ensembles*.

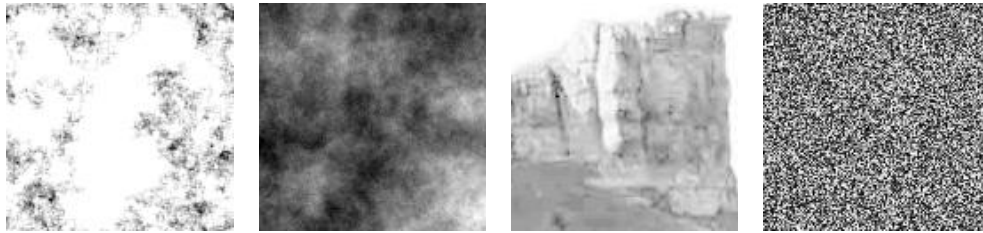


Fig. D.3: *Exemplos de 4 ensembles: estas populações possuem 30 padrões distintos cada com a mesma dimensão 128×128 pixels. A primeira é um exemplo da trajetória de um caminhante aleatório bidimensional, a segunda, um fractal, a terceira é uma cena natural filtrada pelo logaritmo de Ruderman, e a última é um ruído branco.*

A variância determinada pixel a pixel para cada conjunto de imagens é exibida na figura D.4. Mais do que revelar texturas bastante particulares, prova que as distribuições ponto a ponto são, em média, suficientemente homogêneas.

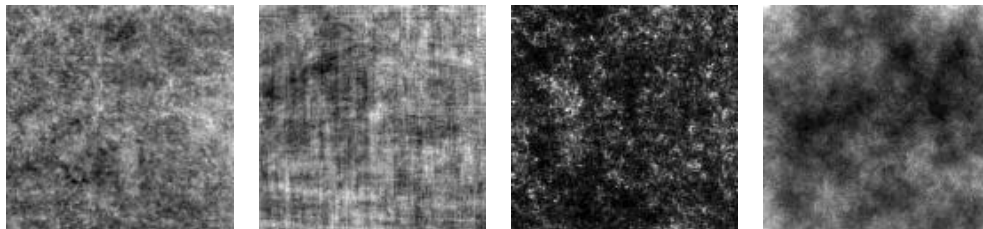


Fig. D.4: *Imagens de dimensão 128×128 cujas intensidades representam as variâncias dos pixels correspondentes aos ensembles em estudo. Da esquerda para a direita, temos cenas naturais, cenas artificiais, caminhante aleatório e cenas fractais.*

Referências Bibliográficas

- [1] Abbott, L. F., e Dayan, P. 2001. *Theoretical Neuroscience*. The MIT Press, Cambridge, Massachusetts, Londres.
- [2] Barlow, H. B. 1989. Unsupervised Learning. *Neural Computation*. **1**, 295-311.
- [3] Bear, M. F., Connors, B. W., e Paradiso, M. A. 1996. *Neuroscience, Exploring the Brain*. Williams & Wilkins, Baltimore.
- [4] Butkov, E. 1988. *Física Matemática*. Livros Técnicos e Científicos Editora S.A., Rio de Janeiro.
- [5] Connor, F. R. 1978. *Sinais - Tópicos de Introdução à Electrónica e às Telecomunicações*. Interciência Editora Lda., Lisboa.
- [6] Field, D. J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*. **4**, 2379-2394.
- [7] Field, D. J. 1994. What is the goal of sensory coding? *Neural Computation*. **6**, 559-601.
- [8] Hancock, P. J. B., Baddeley, R. J. e Smith, L. S. 1992. The principal components of natural images. *Network*. **3**, 61-70.
- [9] Hertz, J., Krogh, A., e Palmer, R. G. 1991. *Introduction to the Theory of Neural Computation*. Addison-Wesley Publishing Company, Nova Iorque.
- [10] Haykin, S., e Veen, B. V. 2001. *Sinais e Sistemas*. Bookman, Porto Alegre.
- [11] Idiart, M. A. P. 2000. Performances in Supervised Learning. *Physica A*. **285**, 566-578.

- [12] Idiart, M. A. P., Berk, B., e Abbott, L. F. 1995. Reduced Representation by Neural Networks with Restricted Receptive Fields. *Neural Computation*. **7**, 507-517.
- [13] Lazzaro, J., e Wawrzynek, J. 1999. JPEG Quality Transcoding Using Neural Networks Trained with a Perceptual Error Measure. *Neural Computation*. **11**, 267-296.
- [14] Linsker, R. 1988. Self-Organization in a Perceptual Network. *Computer*. **21(3)**, 105-117.
- [15] Linsker, R. 1992. Local Synaptic Learning Rules Suffice to Maximize Mutual Information in a Linear Network. *Neural Computation*. **4**, 691-702.
- [16] Liu, Y., e Shouval, H. 1994. Localized principal components of natural images - an analytic solution. *Network: Comput. Neural Syst.* **5**, 317-324.
- [17] Mandelbrot, B. B. 1983. *The Fractal Geometry of Nature*. W. H. Freeman and Company, Nova Iorque.
- [18] Mandelbrot, B. B. *et al.* 1988. *The Science of Fractal Images*. Heinz-Otto Peitgen Dietmar Saupe Editors.
- [19] Nevado, A., Turiel, A., e Parga, N. 2000. Scene dependence of the non-Gaussian scaling properties of natural images. *Network: Comput. Neural Syst.* **11**, 131-152.
- [20] Oja, E. 1989. Neural Networks, Principal Components, and Subspaces. *Int. J. of Neural Systems*. **1**, 61-68.
- [21] Olshausen, B. A., e Field, D. J. 1996. Natural image statistics and efficient coding. *Network: Comput. Neural Syst.* **7**, 333-339.
- [22] Oppen, M., e Kinzel, W. 1995. *Physics of Neural Network: Models of Neural Networks III Association, Generalization and Representation*. Springer.
- [23] Press, W. H., Teukolsky, S. A., Vetterling, W. T., e Flannery, B. P. 1994. *Numerical Recipes in C: The art of scientific computing*. Cambridge University Press, Cambridge, Nova Iorque, Port Chester, Melbourne, Sidnei.
- [24] Révész, P. 1990. *Random Walk in Random and Non-random Environments*. World Scientific Publishing Co. Pte. Ltd., Londres.

- [25] Ruderman, D. L. 1994. The statistics of natural images. *Network: Comput. Neural Syst.* **5**, 517-548.
- [26] Salinas, R. A. 1999. *Introdução à Física Estatística*. Editora da USP, São Paulo.
- [27] Sanger, T. D. 1989. Optimal Unsupervised Learning in a Single-Layer Linear Feed-forward Neural Network. *Neural Networks.* **2(6)**, 459-473.
- [28] Shannon, C. E. 1948. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **27(3)**, 379.
- [29] Shouval, H. 1995. *Formation and organization of receptive fields, with an input environment composed of natural scenes*. Brown University.
- [30] Shouval, H., Blais, B. S., Intrator, N., e Cooper, L. N. 1998. Receptive field formation in natural scene environments: comparison of single-cell learning rules. *Neural Computation.* **10**, 1797-1813.
- [31] Shouval, H., e Liu, Y. 1996. Principal component neurons in a realistic visual environment. *Network: Comput. Neural Syst.* **7**, 501-515.
- [32] Smith, E., Atkinson, R., Atkinson, R., e BEN, D. 1995. *Introdução à Psicologia*. Artes Médicas, Porto Alegre.
- [33] Turiel, A., e Parga, N. 2000. The multi-fractal structure of contrast changes in natural images: from sharp edges to textures. *Neural Computation.* **12**, 763-793.
- [34] Turiel, A., Parga, N., e Mato, G. 1998. The self-similarity properties of natural images resemble those of turbulent flows. *Physical Review Letters.* **80(5)**, 1098-1101.
- [35] Van Hateren, J. H., e Van Der Schaaf, A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. B* **265**, 359-366.
- [36] Wegner, T., e Tyler, B. 1995. *Criando Fractais*. Axcel Books, Rio de Janeiro.