



## Redes Neurais Aprendendo via *Backpropagation* e Neuroevolução Uma Comparação visual

Eduardo Bassani Chandelier<sup>1</sup>; Márcio Dorn<sup>1</sup>  
<sup>1</sup>Universidade Federal do Rio Grande do Sul - UFRGS

### Introdução

Redes neurais são modelos computacionais inspirados no cérebro humano. Elas são usadas para o reconhecimento de padrões e problemas de classificação, possuindo, muitas vezes, dezenas de milhares de neurônios, o que dificulta a interpretação do que foi aprendido. Otimizadores são algoritmos que realizam o treinamento de redes neurais. Existem duas principais categorias de otimizadores, as baseadas no algoritmo *backpropagation* e as que utilizam algoritmos genéticos, chamadas de neuroevolução. Quando cientistas de dados procuram otimizadores para treinar suas redes neurais, normalmente alguma variação do (SGD) é escolhida, devido ao desempenho tanto em tempo quanto em espaço (memória) quando comparados com métodos de neuroevolução como LEEA. Atualmente estão sendo testados e desenvolvidos diversos algoritmos para interpretação de redes neurais que nos permitem parar de enxergá-las como uma “caixa-preta” e começar a interpretar e comparar o que foi aprendido. A pesquisa se encaixa nesse contexto e tenta responder à seguinte pergunta: Duas redes iguais (mesma topologia), que atingem uma alta acurácia no conjunto de teste, treinadas com algoritmos diferentes (SGD e LEEA), aprendem os mesmos pesos?

### Métodos

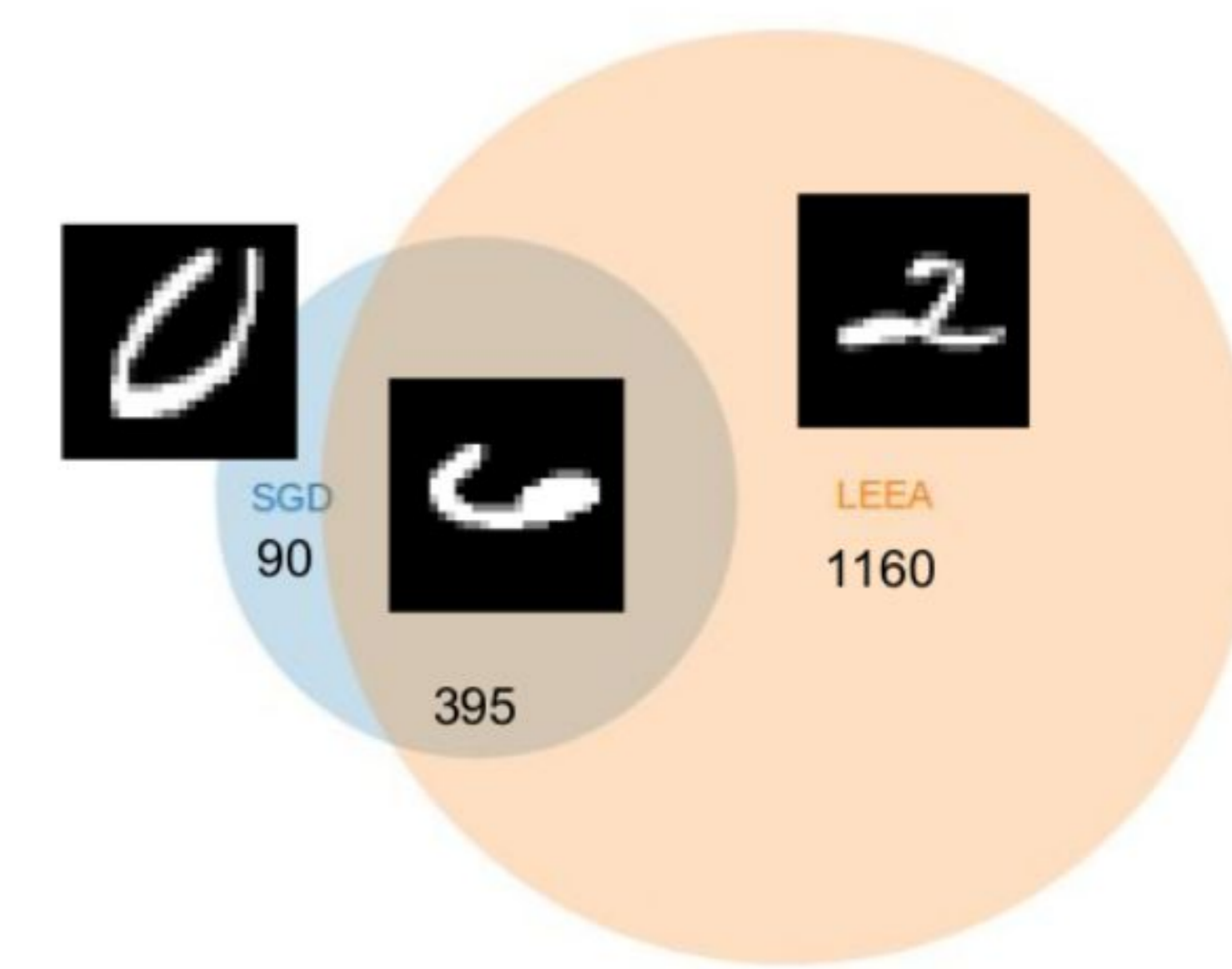
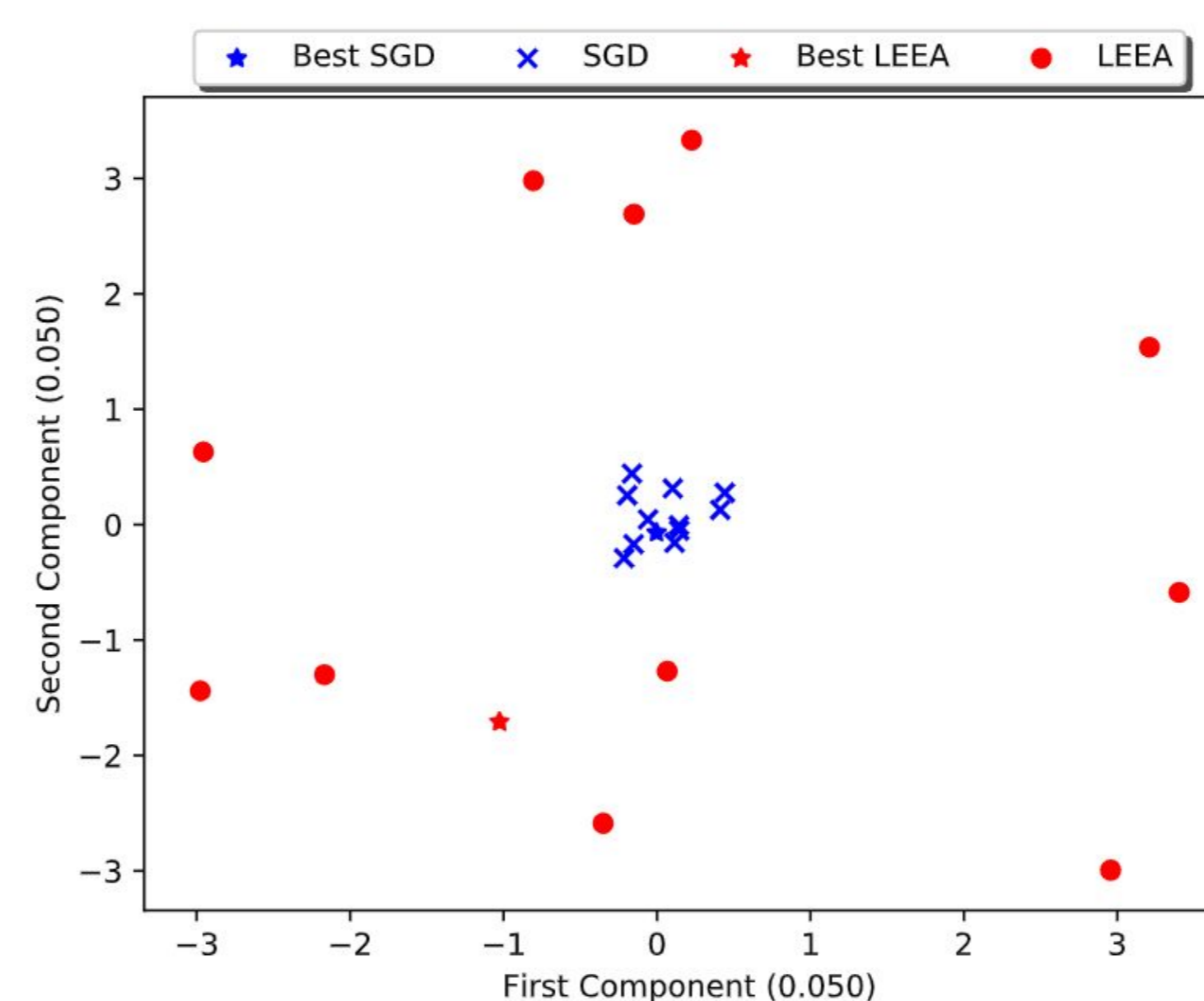
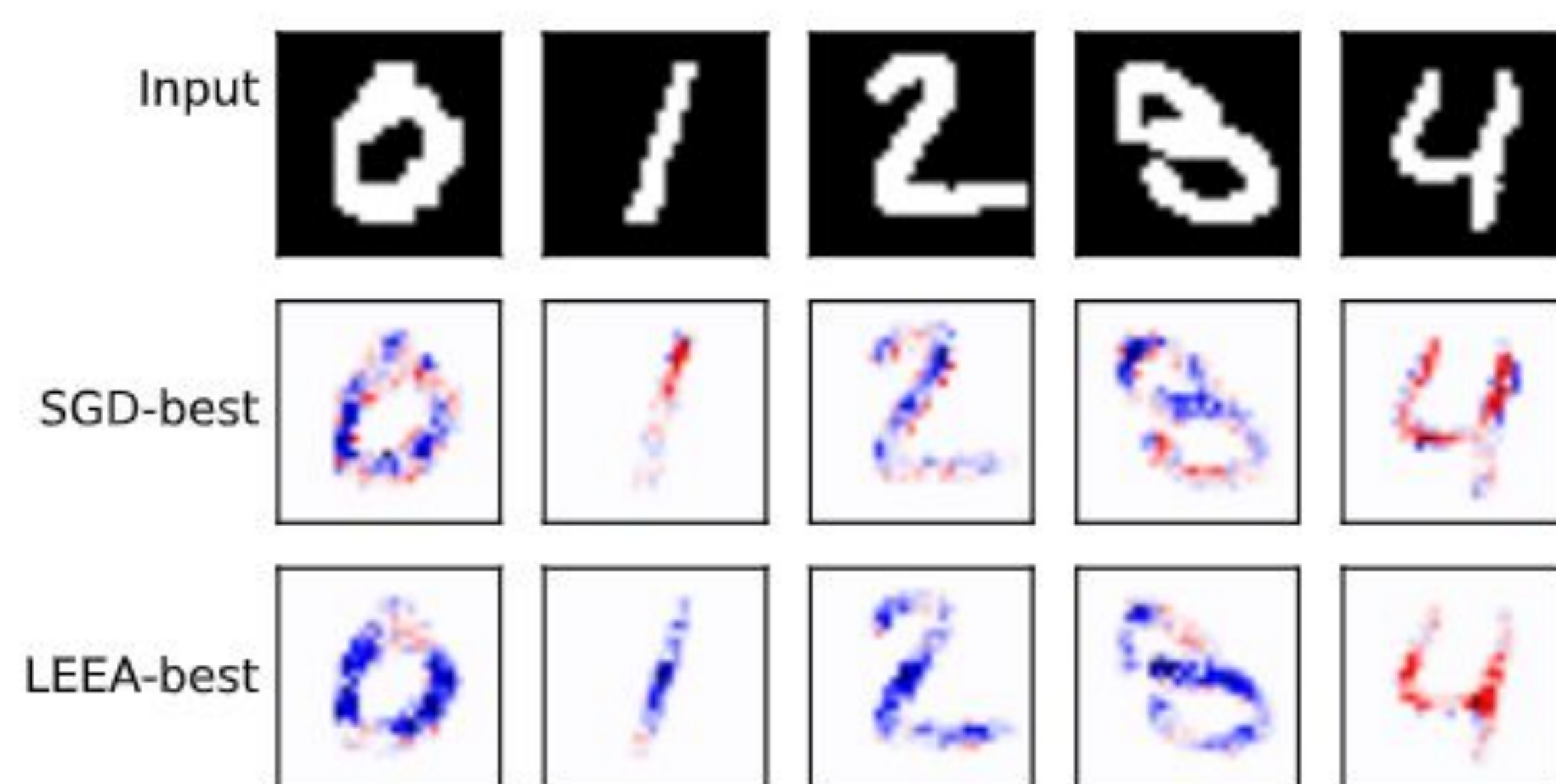
Para realizar a pesquisa utilizamos o *keras*, um *framework* de aprendizado de máquina que oferece uma API de alto nível para treinamento de redes neurais. Ela nos disponibiliza diversos otimizadores, entre eles o SGD, que foi utilizado nos experimentos. O LEEA não estava na lista de otimizadores disponíveis e teve que ser implementado utilizando operações de matrizes nos pesos do modelo, assim utilizando instâncias do mesmo modelo para os diferentes algoritmos, garantindo que estamos comparando somente o otimizador. Para as comparações utilizamos Diagrama de Venns para representar a intersecção dos erros de ambos os algoritmos, PCA (*Principal Component Analysis*) para desenhar as redes em uma gráfico 2D, distância euclidiana dos pesos das redes e por último o LRP (*Layer-wise Relevance Propagation*) e *Max Activation*, que são algoritmos usados para interpretação visual dos pesos de uma rede. Os algoritmos serão testados no banco de dados MNIST, uma coleção de 70 mil imagens de dígitos escritos a mão, no banco Íris, que possui 150 amostras e 3 categorias para classificação de íris, no CIFAR-10, um *dataset* com 60 mil imagens coloridas e CuMiDa, um banco de dados de expressão gênica para classificação de pacientes com câncer disponibilizado pelo próprio grupo de pesquisa.

### Objetivo

Comparar algoritmos dessas duas categorias para analisar suas diferenças ou semelhanças de forma visual. Os algoritmos escolhidos foram o *Stochastic Gradient Descent* (SGD), um otimizador que utiliza o algoritmo *backpropagation* e LEEA (*Limited Evaluation Evolutionary Algorithm*), que utiliza algoritmos genéticos.

### Resultados Parciais

Até o momento temos implementado os dois algoritmos, eles foram comparados utilizando o Diagrama de Venns, PCA, Distâncias Euclidiana e LRP, as acurácias estão em 98% e 94% para o conjunto de teste usando o SGD e LEEA, respectivamente Todos os testes foram feitos utilizando somente o MNIST. Faltam ainda a análise do Max Activation, todas as análises para os outros datasets para termos mais representatividade e principalmente aumentar a acurácia do LEEA para não termos ruído nas comparações devido a diferença de acurácia.



LRP: Saída de um algoritmo que mostra a relevância dos pixels na classificação. Neste caso uma comparação entre SGD e LEEA no dataset MNIST. A cor vermelha indica que o pixel foi importante para classificar a imagem como sendo 4, e os azuis, como sendo não 4.

PCA: Plot dos pesos encontrados pelas redes no espaço. As rede possui mais de 40 mil pesos/bias, mas com o PCA podemos representar esses pesos em um gráfico 2D.

Diagrama de Venns: Representa os conjuntos das imagens que cada algoritmo errou, indicando que, mesmo LEEA tendo errado um número maior de exemplos, existem algumas (90), que ela acertou, comparando com SGD.