

Análise comparativa de funções de ativação em uma topologia de Rede Neural para o reconhecimento de dígitos escritos à mão

**Gabriela da Silva de Souza¹, Gustavo Damasceno da Silva¹, Luiza Marlene
Oliveira Dias²**

¹Instituto de Tecnologia – Faculdade de Engenharia da Computação e
Telecomunicações – Universidade Federal do Pará (UFPA) – Belém – PA – Brasil

²Instituto de Tecnologia – Faculdade de Engenharia Elétrica e Biomédica –
Universidade Federal do Pará (UFPA) – Belém – PA – Brasil

`gabriela.souza@itec.ufpa.br, gustavods07@gmail.com,
luiza8.marlene@gmail.com`

Abstract. *Computational models inspired by the human cognitive system, such as Artificial Neural Networks, represent excellent tools to assist in image processing. This article discusses the use of different neural networks for the accurate identification of handwritten digits available in the MNIST (Modified National Institute of Standards and Technology) database, comparing different activation functions with the TensorFlow tool and the Keras library in search of the best possible accuracy for the same network topology.*

Resumo. *Modelos computacionais inspirados pelo sistema cognitivo humano, como Redes Neurais Artificiais, representam excelentes ferramentas para auxiliar no processamento de imagens. Este artigo aborda a utilização de diferentes redes neurais para a identificação acurada de dígitos escritos à mão disponíveis na base de dados MNIST (do inglês, Modified National Institute of Standards and Technology database), comparando funções distintas de ativação através da ferramenta TensorFlow e da biblioteca Keras em busca da melhor acurácia possível em relação a mesma topologia de rede.*

1. Introdução

Redes Neurais surgiram no âmbito de estudos computacionais como uma tentativa de simular o funcionamento do cérebro humano através da reprodução simplificada de processos de aprendizado. A topologia de uma ANN (do inglês, *Artificial Neural Network*) está intrinsecamente relacionada ao poder de processamento associado, por conta do processo iterativo de múltiplas etapas que ocorre em tempo real atualizando pesos e bias ao longo da execução. Com o avanço e a diminuição de custo de processadores tornou-se possível utilizar-se de ANNs para pautar estudos em diversos campos de conhecimento, entre eles o reconhecimento de imagens.

Neste contexto, o aprendizado de máquina através de ANNs oferece a possibilidade de automatizar o reconhecimento de caracteres desenhados à mão, como dígitos numéricos.

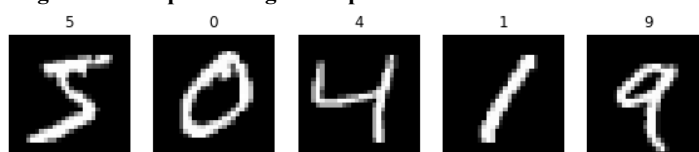
O estudo proposto implementa as funções de ativação *Sigmoidal*, *Softmax* e *ReLU* (do inglês, *Rectified Linear Unit*) para criar três versões de uma topologia simplificada de ANN (caracterizada por uma camada de entrada, uma camada escondida e uma camada de saída) voltada ao reconhecimento dos dígitos escritos à mão disponíveis no banco de dados MNIST e comparar a acurácia através de resultados estatísticos obtidos por processos de treinamento e teste.

2. Fundamentação teórica

2.1. Disposição e tratamento de dados

A base de dados MNIST dispõe de 70.000 imagens padronizadas (28 x 28 pixels) de dígitos escritos à mão e associadas ao valor numérico exibido na imagem. Por padrão, cada pixel exibe um valor entre 0 e 255 representando o nível de cinza. A Figura 1 exibe alguns dígitos escritos à mão dispostos no banco de dados utilizado.

Figura 1. Exemplos de dígitos dispostos no banco de dados MNIST.



Fonte: autor.

Para transformar cada imagem em parâmetros de entrada para a topologia da rede neural, os valores dos pixels (originalmente dispostos em uma matriz de duas dimensões) serão dispostos em arrays equivalentes de 784 números.

Cabe destacar que, para acelerar o processo de aprendizado e facilitar a convergência da ANN, também mostra-se importante normalizar os valores dos pixels, convertendo os valores para se adequarem ao intervalo de valores entre 0 e 1.

2.2. Topologia da rede neural artificial

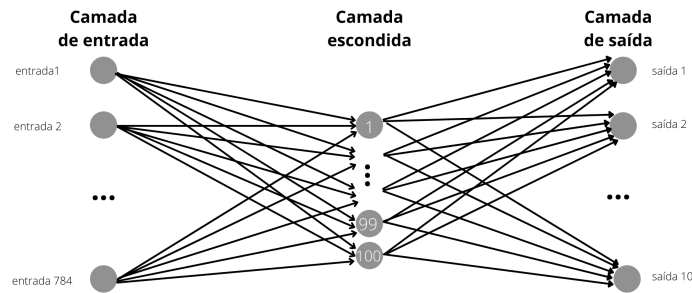
Para padronizar a estrutura de ANN e comparar com consistência a performance de cada função de ativação, optou-se por uma topologia simples: camada de entrada, camada escondida (única) e camada de saída.

A camada de entrada dispõe de 784 neurônios, um para cada valor associado ao pixel. A camada escondida, por sua vez, conta com 100 neurônios, enquanto a camada de saída é constituída por 10 neurônios, cada um associado a um dígito entre 0 e 9.

Cabe destacar que, cada neurônio da camada de saída será responsável por exibir um valor percentual relacionado à previsão do dígito disposto na imagem. Portanto, o maior percentual representará o resultado da predição estipulada pela rede.

A Figura 2 exibe a topologia básica para as redes neurais adotadas.

Figura 2. Topologia básica das ANN desenvolvidas.



Fonte: autor.

2.2. Funções de ativação

A função de ativação é a transformação não linear realizada ao longo do sinal de entrada. Para este estudo, são abordadas as funções *Sigmoidal* e *Softmax*.

2.2.1. Função Sigmoidal

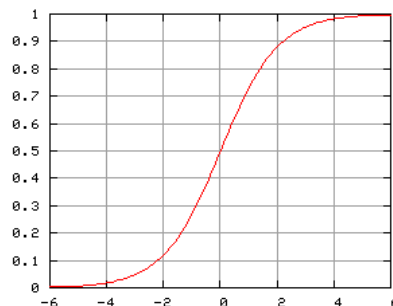
Considerando o funcionamento binário de um neurônio biológico, ativando ou não, a função sigmóide apresenta-se como uma eficiente técnica de modelagem para o aprendizado, assumindo valores apenas entre 0 e 1. Desta forma, aplicações de classificação binária são ideais para aplicação da função Sigmoidal.

A Equação 1 descreve a expressão matemática da função Sigmoidal, enquanto a Figura 3 exhibe o comportamento da função Sigmoidal ao lado da função descrita matematicamente.

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

Em que $\phi(x)$ representa a saída da função de ativação para o valor de entrada x .

Figura 3. Comportamento da função de ativação Sigmoidal.



Fonte: Adaptado da internet.

2.2.1. Função Softmax

A função de ativação Softmax é uma variação da função Sigmoidal e sua proposta é lidar com múltiplas classes na camada final, forçando a saída da ANN a

representar a probabilidade dos dados serem de uma das classes definidas. Desta forma, sem a função Softmax as saídas dos neurônios são apenas valores numéricos onde o maior indica a classe vencedora.

Na Equação 2 está descrita a expressão matemática da função Softmax.

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad \text{para } j = 1, \dots, K \quad (2)$$

Em que $\sigma(\mathbf{z})_j$ representa a saída da função de ativação, j representa o índice do neurônio de saída, k representa os índices dos neurônios da mesma camada e \mathbf{z} representa o vetor de neurônios da saída.

2.3. Propagação de erro e taxa de aprendizado

Durante as etapas de treino das ANNs, o erro das predições é calculado e posteriormente transmitido aos neurônios para correção dos valores de Peso e Bias através do algoritmo de *Backpropagation*.

Para garantir a estabilidade do aprendizado das redes neurais abordadas, a taxa de aprendizado é inicialmente fixada em 0,001. Entretanto, através da biblioteca *Keras* e da ferramenta *Adam* (derivado de *Adaptive Moment Estimation*), gradientes de primeira ordem com pouca necessidade de memória são utilizados para calcular taxas de aprendizagem adaptativa individuais para diferentes parâmetros de estimativas de primeiro e segundo momentos dos gradientes (parâmetros de adaptação da rede neural).

3. Metodologia de avaliação

A abordagem proposta levará em consideração 2 variações da topologia de rede neural descrita anteriormente (Tabela 1). Cabe destacar que, para a camada escondida, convencionou-se a utilização da função de ativação Sigmoidal, amplamente difundida para aplicações de redes neurais artificiais.

Tabela 1. Parâmetros das redes neurais propostas.

Rede neural	Função de ativação (camada escondida)	Função de ativação (camada de saída)
A	Sigmoidal	Sigmoidal
B	Sigmoidal	Softmax

Fonte: Autor.

3.1. Parâmetros de treinamento

As redes A e B foram submetidas aos mesmos parâmetros de execução da fase de treinamento, na qual utilizaram-se 60.000 imagens em ciclos de treino compostos por 200 épocas.

3.2. Parâmetros de teste

A fase de teste foi realizada através de um conjunto de 10.000 imagens não utilizadas na etapa de treinamento, e consiste na avaliação da predição dos dígitos dispostos.

3.2. Parâmetros de avaliação

A avaliação de performance das redes neurais A e B baseou-se no acompanhamento do parâmetro de acurácia, durante a fase de treinamento de cada rede, bem como na acurácia das predições durante a etapa de teste à qual foram submetidas.

Além disso, a função de perda relacionada à entropia cruzada, expressa matematicamente na Equação 3, também serviu de parâmetro para as comparações estipuladas, representando uma medida de erro eficaz, visto que as saídas da uma rede podem ser pensadas como hipóteses independentes (cada neurônio de saída representa um dígito diferente).

$$H(p, q) = - \sum_x p(x) \log q(x) \quad (3)$$

Em que $H(p, q)$ representa a entropia cruzada, $p(x)$ representa a probabilidade desejada e $q(x)$ representa a probabilidade real.

4. Resultados

Após submeter as ANNs A e B à fase de teste, foi possível obter os índices de acurácia dispostos na Tabela 2

Tabela 2. Acurácia das redes neurais A e B na fase de testes.

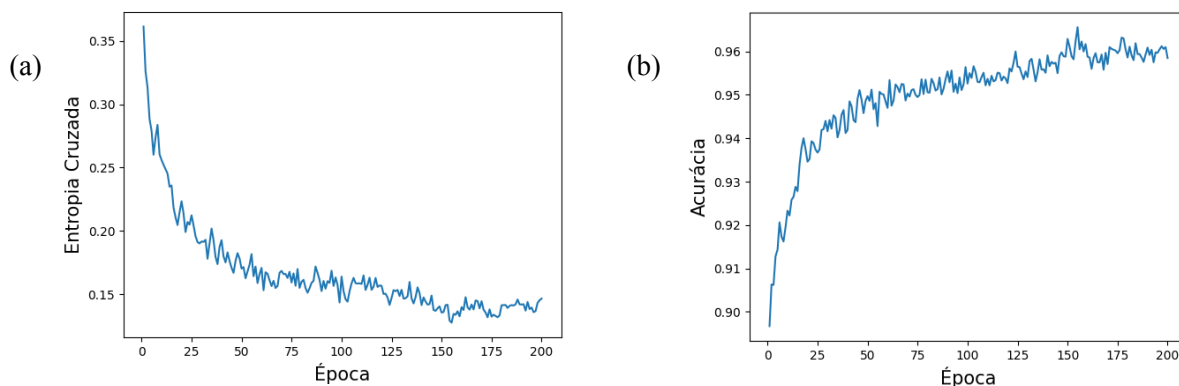
Rede neural	Acurácia na fase de testes %
A	0.9585000276565552
B	0.9606999754905701

Fonte: Autor.

A Figura 4 demonstra, para a rede neural A, a evolução do erro relacionado à entropia cruzada (a) e a evolução da acurácia durante o processo de treinamento, respectivamente.

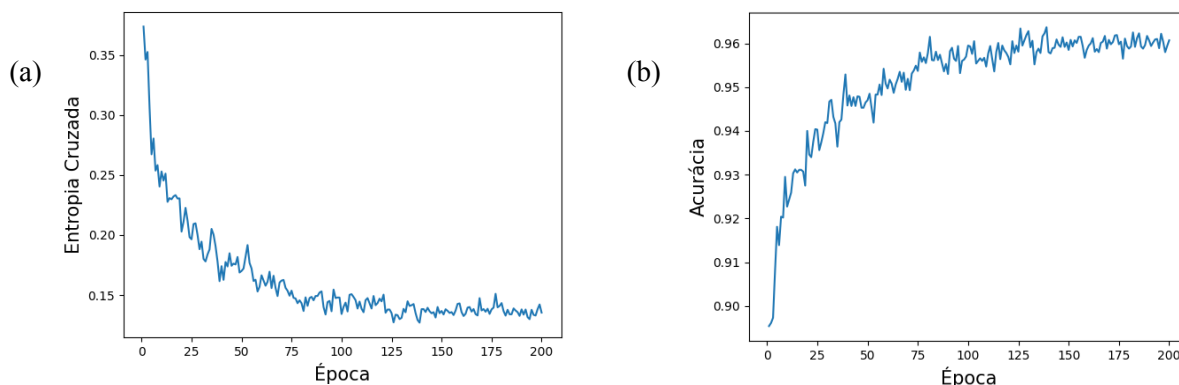
A Figura 5 demonstra, para a rede neural B, a evolução do erro relacionado à entropia cruzada (a) e a evolução da acurácia durante o processo de treinamento, respectivamente.

Figura 4. Acompanhamento da rede neural A: Evolução do erro relacionado à entropia cruzada (a) e a evolução da acurácia (b) durante o processo de treinamento, respectivamente.



Fonte: Autor.

Figura 5. Acompanhamento da rede neural B: Evolução do erro relacionado à entropia cruzada (a) e a evolução da acurácia (b) durante o processo de treinamento, respectivamente.



Fonte: Autor.

É possível observar que, apesar de oferecerem percentuais finais de acurácia muito semelhantes, a abordagem através da função de ativação *Softmax* implicou em um percentual de acerto maior e em níveis de estabilidade maiores ao longo das épocas. Esta estabilidade maior pode ser observada através dos gráficos dispostos para cada rede, atentando-se às variações menos flutuantes de entropia e acurácia, principalmente a partir da 75ª época de treinamento, se compararmos as redes A e B.

4. Conclusão

Através da aplicação de duas redes neurais, foi possível obter resultados satisfatórios de acurácia para a predição correta de dígitos escritos à mão, tomando como base um vasto banco de dados.

Além disso, foi possível comparar abordagens diferentes da mesma topografia de rede com a utilização de duas funções de ativação e compreender as nuances de performance das ANNs desenvolvidas através de um acompanhamento quantitativo das predições obtidas.

Por fim, os resultados dispostos neste trabalho reiteram o sucesso da aplicação do

aprendizado de máquina no reconhecimento de elementos visuais através da utilização de redes neurais.

5. Referências

1. K. T. Islam, G. Mujtaba, R. G. Raj and H. F. Nweke, "Handwritten digits recognition with artificial neural network," 2017 International Conference on Engineering Technology and Technopreneurship (ICE2T), 2017, pp. 1-4, doi: 10.1109/ICE2T.2017.8215993.
2. Estrutura de Dados e Algoritmos, **COS-121**, 2021. Disponível em: https://www.cos.ufrj.br/~rfarias/cos121/aula_09.html. Acesso em: 27 de maio de 2021.
3. LeCun, Y.; Cortes, C.; Burges, C.J.C. **The MNIST Database of Handwritten Digits**. 2012. Disponível em: <http://yann.lecun.com/exdb/mnist/> Acesso em: 05 de outubro de 2021.
4. TensorFlow. **MNIST for ML Beginners**. 2017. Disponível em: https://www.tensorflow.org/get_started/mnist/beginners. Acesso em: 05 de outubro de 2021.

