

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA

UNIDADE ACADÊMICA DE SISTEMAS E COMPUTAÇÃO

DISCIPLINA: Inteligência Artificial

PERÍODO: 2024.1

DOCENTE: Joseana Macêdo Fechine Régis de Araújo

DISCENTES: Caroline de Oliveira Cordeiro 121111059

Francisco Antonio Dantas de Sousa 121110408

Guilherme Alberto Dutra Camelo 121210159

Marcos Vinícius Fragoso de Oliveira 122210469

Campina Grande

Outubro de 2024

# Introdução

Neste trabalho, foram desenvolvidos três modelos de Aprendizagem de Máquina (Machine Learning) com Redes Neurais, sendo um deles uma Rede Neural Convolucional. Redes Neurais são métodos de IA que fazem parte do paradigma Conexionista, onde é utilizado como base da estrutura do modelo redes neurais humanas, como são feitas as conexões entre os neurônios para o compartilhamento de informação.

No desenvolvimento dos modelos foi utilizado o [*TensorFlow*](https://www.tensorflow.org/?hl=pt-br), biblioteca de código aberto da Google, que permite a criação, treinamento e implantação de modelos de Machine Learning e Deep Learning, permitindo a criação de redes complexas e eficientes.

# Modelos utilizados

Para este projeto, foram treinados dois modelos de Redes Neurais gerais e um modelo de Redes Neurais Convolucionais (CNN). Redes Neurais são modelos computacionais inspirados no funcionamento do sistema nervoso central, principalmente na forma que as informações são processadas. O modelo vai utilizar um sistema de grafos onde cada neurônio é representado por nós e suas conexões por arestas, onde cada aresta vai ter um peso associado que equivale a importância do sinal emitido naquela conexão. A partir desses sistemas, é possível treinar a rede para reconhecer padrões e aprender o funcionamento do fluxo de informações de dado contexto.

As Redes Neurais Convolucionais são uma especificação das redes neurais que processam dados com estrutura de grades, onde vão existir várias camadas de convolução (operação matemática onde é feito o somatório de vários produtos) para extrair as características visuais da imagem, organizando isso em uma hierarquia. Este modelo é altamente eficaz e muito utilizado para tarefas que necessitam de reconhecimento de objetos e detecção de padrões em imagens.

# Base de Dados

O projeto conta com a utilização de 3 datasets para o treinamento dos modelos, esses datasets são:

## MNIST (Modified National Institute of Standards and Technology Database):

O **MNIST** é um banco de dados de dígitos manuscritos, ele é comumente utilizado para treinamento de processamentos de imagem. Ele foi criado por meio da "remixagem" das amostras dos conjuntos de dados originais do NIST. Ele contém 70.000 imagens em escala de cinza de dígitos manuscritos (de 0 a 9), sendo 60.000 para treino e 10.000 para teste.

## Fashion MNIST:

O **Fashion MNIST** é um dataset alternativo ao **MNIST** tradicional, voltado para a classificação de peças de roupa. Assim como o **MINIST** possui 60.000 imagens para treino e 10.000 para teste que estão divididas em 10 classes de vestuário, como camisetas, sapatos e bolsas.

## SVHN (Street View House Numbers):

O **SVHN (Street View House Numbers)** é um conjunto de dados de imagens de dígitos que serve como referência para a classificação de dígitos. O conjunto de dados é composto por mais de 600.000 imagens de dígitos impressos (de 0 a 9) extraídas de fotos de placas de números de casas.

# Modelos Treinados

## Classificação de Dígitos

O primeiro modelo treinado tem como objetivo principal detectar e classificar dígitos manuscritos, levanto em consideração aqui apenas caracteres numerais de 0 até 9. Foi utilizado para tal o *dataset* MNIST, descrito na seção 3.1. Os dados importados já estão divididos em treino e teste, de forma que essa divisão não foi necessária. Após a importação, os conjuntos de treino e teste são transformados em vetores de 784 valores, para poderem assim serem processados pela camada densa/totalmente conectada da rede, e normalizados para o intervalo [0,1], melhorando assim a eficiência e estabilidade do treinamento.

Modelo vai ter 3 camadas, onde:

* **Primeira camada densa:** tem 512 neurônios, a função de ativação ReLU (Rectified Linear Unit) – função computacionalmente simples que permite o aprendizado de padrões complexos e o comportamento não-linear do modelo. Essa camada recebe todas as entradas e captura os padrões complexos.
* **Segunda camada densa**: tem 256 neurônios e também ativação ReLU.
* **Camada de saída**: tem 10 neurônios responsáveis por classificar os dígitos de 0 a 9 e não tem função de ativação.

O modelo é então compilado: o otimizador escolhido foi o Adam, ele vai ajustar os pesos do treinamento utilizando uma taxa de aprendizado de 0.001; a função de perda escolhida foi a Entropia Cruzada Categórica Sparse, já que estava sendo trabalhada 10 classes (os 10 dígitos), essa função ajuda na diminuição de memória gasta; a acurácia do modelo é registrada durante o treinamento, para servir como métrica de avaliação dos resultados tardiamente.

Por fim, foi utilizado um *batch size* de 32 e 7 *epochs*. Cada época teve 1875 passos executados no treinamento, e demorou em média 23s para cada execução. A acurácia vai de 94,34% até 99,29%, indicando a melhoria na capacidade de classificação dos dígitos à medida que o modelo aprende com os dados. Já a perda caiu de 0.1869 para 0.0222, indicando também a melhoria na classificação.

O teste foi executado em 1 segundo, levando em média 3 milissegundos por passo, teve uma acurácia de 0.9782 e uma perda de 0.0897.

Em resumo, o modelo teve um desempenho de aprendizagem muito bom, com uma acurácia alta desde o início do treinamento, faixa que se manteve no teste, e a sua perda também foi muito boa, embora tenha sido um pouco mais alta que o melhor valor do treinamento. Assim, o modelo final é adequado para a classificação de dígitos manuscritos.

## Classificação de Roupas

Esse modelo tem como objetivo classificar imagens de roupas em 10 categorias diferentes, utilizando um treinamento muito semelhante ao do modelo anterior. Novamente foi utilizado um *dataset* MNIST, dessa vez o dataset *Fashion*, descrito na seção 3.2. Por mais que não tenha sido “setado”, a taxa de aprendizado do Adam continua sendo 0.001, que é seu valor *default*.

No treinamento, novamente tivemos 1875 passos executados, e a média do tempo de execução foi de 16s. A acurácia vai de 0.8284 até 0.9053, demonstrando a melhoria da classificação ao longo do treinamento do modelo. A perda caiu de 0.4702 para 0.2511, indicando novamente a melhoria na classificação.

O teste foi executado em 1 segundo, levando em média 3 milissegundos por passo, e tendo uma acurácia de 0.8780 e uma perda de 0.3393.

Por fim, o modelo teve um desempenho bem razoável, porém com valores de acurácia e perda no teste que indicam a possibilidade de melhoria.

## Classificação de Dígitos - *dataset* SVHN

Neste modelo, o objetivo é realizar a classificação de dígitos utilizando o dataset SVHN, que contém imagens de dígitos extraídos de placas de casas (Street View House Numbers). Ao contrário dos modelos anteriores utilizados em datasets como o Fashion MNIST, que dependiam apenas de camadas densas, este modelo faz uso de uma arquitetura de rede neural convolucional (CNN), aproveitando as características espaciais das imagens para uma extração de padrões mais eficiente.

Os dados são pré-processados para serem compatíveis com a rede convolucional. Cada imagem no formato de 32x32 pixels com três canais RGB é normalizada para o intervalo [0,1]. Esse formato tridimensional preserva a informação de cor das imagens, essencial para capturar a complexidade dos dígitos em diferentes contextos visuais.

A arquitetura da rede neural convolucional é composta por três camadas convolucionais, cada uma seguida de uma camada de pooling:

1. **Primeira camada convolucional**: Contém 32 filtros com kernel de tamanho 3x3 e utiliza a função de ativação ReLU. Essa camada realiza a detecção inicial de bordas e formas simples.
2. **Segunda camada convolucional**: Com 64 filtros e ativação ReLU, essa camada aprofunda a análise dos padrões detectados, sendo capaz de capturar detalhes mais complexos das imagens.
3. **Terceira camada convolucional**: Possui 128 filtros, também com ativação ReLU, para identificar padrões ainda mais abstratos.

Entre essas camadas convolucionais, são aplicadas camadas de **MaxPooling 2x2**, que reduzem a dimensionalidade das características, ajudando a diminuir o custo computacional e evitando o overfitting.

Após essas camadas convolucionais, a rede "achata" as saídas das camadas anteriores em um vetor unidimensional, que é passado para as camadas densas. A arquitetura densa contém uma camada intermediária de 256 neurônios, seguida de uma camada de saída de 10 neurônios, representando as classes de dígitos (0 a 9). A função de ativação ReLU também é utilizada na camada densa para aprender padrões complexos.

O modelo foi compilado com o otimizador **Adam**, mantendo sua taxa de aprendizado padrão de 0,001, e a função de perda **Entropia Cruzada Categórica Sparse** foi aplicada para classificação multi-classes. A métrica de acurácia foi usada para monitorar o desempenho.

Esse modelo possui um desempenho aceitável com uma acurácia de 89.06% após 10 etapas de treinamento.

# Conclusão

Considerando os resultados obtidos neste trabalho, podemos concluir que as Redes Neurais, em particular as Convolucionais (CNNs), são ferramentas poderosas para a classificação de imagens. Os três modelos treinados demonstraram alta precisão, alcançando acurácias em média de 98% na classificação de dígitos manuscritos (MNIST), 88% para imagens de roupas (Fashion MNIST) e 87% para dígitos de placas de rua (SVHN).

A arquitetura das redes neurais, composta por camadas densas e funções de ativação adequadas, permitiu a extração de características relevantes das imagens. O otimizador Adam, combinado com a função de perda de entropia cruzada categórica, mostrou-se eficaz em ajustar os pesos da rede e minimizar o erro de classificação. Além disso, o pré-processamento dos dados, como a normalização dos pixels, foi fundamental para melhorar o desempenho dos modelos. Adicionalmente, a divisão dos dados em conjuntos de treinamento e teste permitiu avaliar a capacidade de generalização dos modelos para novos dados.

Diante disso, os resultados obtidos neste trabalho são promissores e abrem caminho para futuras pesquisas. A exploração de diferentes arquiteturas de redes neurais, o uso de técnicas de aumento de dados e o ajuste fino dos hiperparâmetros podem levar a melhorias ainda maiores no desempenho dos modelos.

Em suma, este estudo reforça a capacidade das Redes Neurais de se adaptarem a diferentes contextos de classificação de imagens e destaca a importância de ajustes de hiperparâmetros e pré-processamento adequado dos dados. Através da aplicação em datasets variados, comprovou-se a versatilidade e o potencial das Redes Neurais, inclusive Convolucionais, para resolver problemas complexos de reconhecimento de padrões visuais com altos níveis de precisão.