



Programa de Pós-Graduação em Engenharia de Computação e Sistemas - PECS/UEMA

Professor: Omar Andres Carmona Cortes

Aluno: Adrielson Ferreira Justino

Atividade II: Regularizadores e Otimizadores

Descrição:

- Usando o código previamente desenvolvido, verifique se a mudança de otimizadores produz melhores resultados.
- Teste pelo menos 2 otimizadores além do utilizado na Atividade I.
- Em seguida adicione regularizadores ao experimento (Teste dropout e penalidades).
- Use testes estatísticos para mostrar se a diferença é significativa ou não.

1. DESCRIÇÃO DA BASE DE DADOS

Para esta atividade foi escolhida a base de dados **Breast Cancer Wisconsin (Diagnostic)**¹ que consiste em um conjunto de dados utilizado em tarefas de diagnóstico de câncer de mama. Ela contém informações sobre características extraídas de imagens de células tumorais obtidas de biópsias de tecido mamário, sendo ideal para tarefas de classificação binária, no treinamento de modelos de aprendizado de máquina para prever se um tumor é maligno ou benigno com base nas características descritas.

As características da base foram computadas a partir de imagens de aspiração com agulha fina (FNA) de diferentes massas de mama. A base contém um total de 569 amostras e 31 atributos, conforme descrito a seguir:

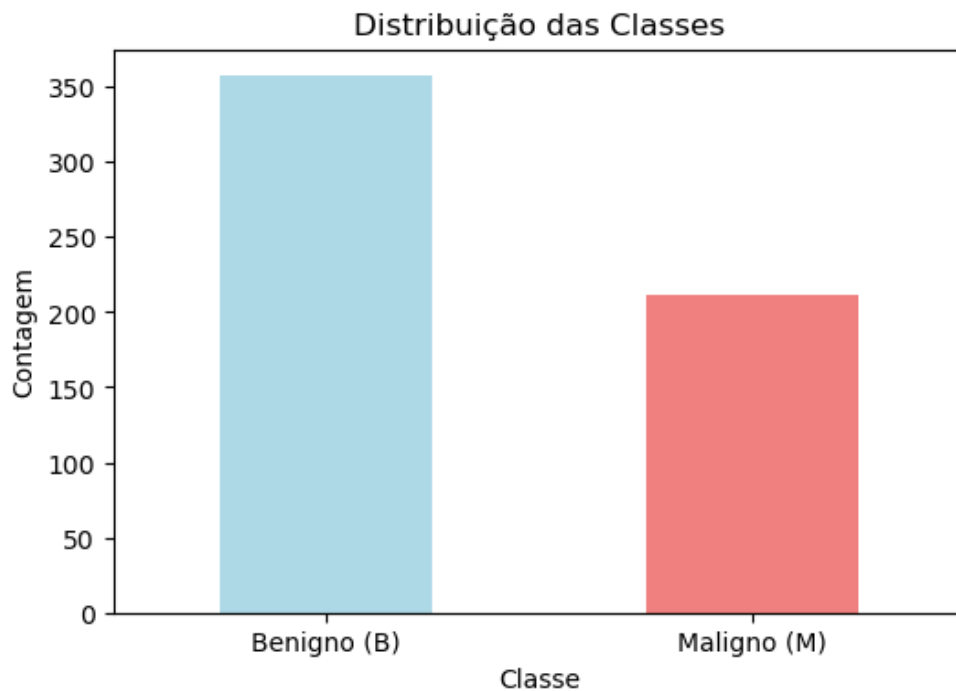
- ID number: Cada entrada no conjunto de dados é identificada por um número de ID único.
- Diagnosis (M = maligno, B = benigno): Esta é a variável de destino que indica se o tumor é maligno (câncer de mama) ou benigno (não canceroso).
- Dez características de valor real são calculadas para cada núcleo celular:
 - *radius* (média das distâncias do centro aos pontos na circunferência)
 - *texture* (desvio padrão dos valores em escala de cinza)
 - *perimeter*
 - *area*

¹ <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>

- *smoothness* (variação local nos comprimentos dos raios)
- *compactness* ($\text{perímetro}^2 / \text{área} - 1.0$)
- *concavity* (gravidade das porções côncavas do contorno)
- *concave points* (número de porções côncavas do contorno)
- *symmetry*
- *fractal dimension* (aproximação à "linha costeira" - 1)
- Para cada imagem, foram calculados os valores médios, erros padrão e "pior" (maior valor das três maiores) dessas características, resultando em um total de 30 características. Por exemplo, o campo 3 é o Raio Médio, o campo 13 é o Erro Padrão do Raio e o campo 23 é o Pior Raio.
- Não há valores ausentes neste conjunto de dados.

Na distribuição das classes no conjunto de dados estão dispostas 357 amostras são benignas (B) e 212 amostras são malignas (M), conforme a Figura 1.

Figura 1. Distribuição das Classes



2. PRÉ-PROCESSAMENTO

Inicialmente, os dados foram preparados para o treinamento do modelo. Para isso foram removidas duas colunas que não são necessárias para a análise: a coluna "id" e a coluna "Unnamed: 32". Em seguida, foi realizada uma codificação dos rótulos das classes. Ele mapeia a classe "M" (Maligno) para o valor 1 e a classe "B" (Benigno) para o valor 0, tornando os rótulos binários. As características foram selecionadas utilizando a indexação de colunas do DataFrame. As características (atributos preditivos) foram armazenadas em uma variável X, enquanto os rótulos (diagnósticos malignos ou benignos) foram armazenados em uma variável y. As

características foram selecionadas a partir da coluna 1 até o final do DataFrame, excluindo a primeira coluna que continha o diagnóstico. Finalmente, os dados são divididos em conjuntos de treinamento (X_train, y_train) e teste (X_test, y_test), 30% dos dados foram reservados para o conjunto de teste, enquanto 70% foram usados para o conjunto de treinamento.

3. ARQUITETURA DOS MODELOS

O principal objetivo desta atividade foi realizar testes visando a otimização de hiperparâmetros em um modelo de rede neural voltado para a classificação binária. Para isso, foi criado um modelo *baseline* simples que serviu como ponto de partida. Posteriormente, foram conduzidos novos testes com diferentes otimizadores e valores de penalidade para investigar como esses fatores afetam o desempenho do modelo. As configurações do baseline são descritas a seguir.

O modelo inicial trata-se de uma rede neural Perceptron Multilayer (MLP) com quatro camadas. A camada de entrada recebe um vetor de entrada de 30 dimensões e é seguida por duas camadas ocultas. A primeira camada densa possui 256 neurônios com função de ativação *ReLU*, enquanto a segunda camada possui 128 neurônios também com ativação *ReLU*. Por fim, a camada de saída tem 2 neurônios com função *softmax*, adequada para classificação binária. A função de perda usada foi *categorical_crossentropy*. O otimizador *RMSprop* foi empregado para ajustar os pesos da rede, o *callback* de parada antecipada que monitora a perda durante o treinamento foi definido igual 0.

Para identificar os otimizadores e os valores de penalidade mais adequados, foi implementado uma função (*test_Model*) que cria modelos com diferentes configurações, incluindo otimizadores *Adaptive Moment Estimation* (ADAM), *Stochastic Gradient Descent* (SGD) e *Root Mean Square Propagation* (RMSprop), taxas de *dropout* de 0.0 (sem dropout), 0.2 (dropout de 20%) ou 0.4 (dropout de 40%), bem como penalidades L2 de 0.0, 0.01 ou 0.1.

A cada configuração, o modelo foi treinado e avaliado usando métricas de desempenho, como a perda (*loss*) e a precisão (*accuracy*). Os resultados foram registrados para análise posterior. Para isso foi utilizado o teste estatístico de Análise de Variância (ANOVA).

4. RESULTADOS

O resultado obtido com o modelo *baseline* foi uma perda de aproximadamente 3.20 e uma acurácia de 0.8655. Por meio da função *test_Model* foi possível testar 27 combinações de configurações para os modelos usando os 3 otimizadores, 3 taxas de *dropout* e 3 valores de penalidade previamente definidos, conforme a **Figura 2**.

Figura 2. Configurações e pontuações dos modelos testados

```

Resultados para o otimizador adam:
Configuração: ('adam', 0.0, 0.0), Loss: 0.4814, Accuracy: 0.9298
Configuração: ('adam', 0.0, 0.01), Loss: 0.9330, Accuracy: 0.9357
Configuração: ('adam', 0.0, 0.1), Loss: 4.0226, Accuracy: 0.9298
Configuração: ('adam', 0.2, 0.0), Loss: 0.2240, Accuracy: 0.9240
Configuração: ('adam', 0.2, 0.01), Loss: 1.0873, Accuracy: 0.8596
Configuração: ('adam', 0.2, 0.1), Loss: 5.4820, Accuracy: 0.8421
Configuração: ('adam', 0.4, 0.0), Loss: 0.2539, Accuracy: 0.9123
Configuração: ('adam', 0.4, 0.01), Loss: 1.1491, Accuracy: 0.9123
Configuração: ('adam', 0.4, 0.1), Loss: 6.3856, Accuracy: 0.8655
Resultados para o otimizador sgd:
Configuração: ('sgd', 0.0, 0.0), Loss: 0.6440, Accuracy: 0.6901
Configuração: ('sgd', 0.0, 0.01), Loss: 507562624.0000, Accuracy: 0.7018
Configuração: ('sgd', 0.0, 0.1), Loss: 602100.2500, Accuracy: 0.6842
Configuração: ('sgd', 0.2, 0.0), Loss: 0.6358, Accuracy: 0.6901
Configuração: ('sgd', 0.2, 0.01), Loss: 53144092.0000, Accuracy: 0.6784
Configuração: ('sgd', 0.2, 0.1), Loss: 4321551.0000, Accuracy: 0.6842
Configuração: ('sgd', 0.4, 0.0), Loss: 1.4278, Accuracy: 0.6842
Configuração: ('sgd', 0.4, 0.01), Loss: 71139224.0000, Accuracy: 0.6959
Configuração: ('sgd', 0.4, 0.1), Loss: 265339.4062, Accuracy: 0.6842
Resultados para o otimizador rmsprop:
Configuração: ('rmsprop', 0.0, 0.0), Loss: 1.2466, Accuracy: 0.8596
Configuração: ('rmsprop', 0.0, 0.01), Loss: 2.5515, Accuracy: 0.7427
Configuração: ('rmsprop', 0.0, 0.1), Loss: 4.6018, Accuracy: 0.8070
Configuração: ('rmsprop', 0.2, 0.0), Loss: 0.7066, Accuracy: 0.5965
Configuração: ('rmsprop', 0.2, 0.01), Loss: 1.0583, Accuracy: 0.7485
Configuração: ('rmsprop', 0.2, 0.1), Loss: 2.7522, Accuracy: 0.9181
Configuração: ('rmsprop', 0.4, 0.0), Loss: 0.3494, Accuracy: 0.9123
Configuração: ('rmsprop', 0.4, 0.01), Loss: 0.9358, Accuracy: 0.9123
Configuração: ('rmsprop', 0.4, 0.1), Loss: 4.2739, Accuracy: 0.9123

```

As 3 configurações dos modelos testados que alcançaram maior acurácia foram selecionadas para avaliar se as diferenças entre os modelos é significativa ou não. A comparação entre os modelos foi realizada por meio do teste ANOVA.

Tabela 1. Pontuações por modelo

	Otimizador	Dropout	Penalidade L2	loss	accuracy
baseline	x	x	x	3.20	0.8655
Configuração 1	adam	0.0	0.01	0.9330	0.9357
Configuração 2	adam	0.0	0.0	0.4814	0.9298
Configuração 3	adam	0.0	0.1	4.0226	0.9298

A **Tabela 1** apresenta o resumo da configuração 3 modelos selecionados. Por meio, do teste estatístico ANOVA foi identificado um valor-p igual 0.4462. Isso indica que, com base nestas pontuações, não há diferença estatisticamente significativa entre os modelos testados em diferentes configurações em relação à perda e acurácia alcançadas.