

Objetivo: Nesta atividade, você aplicará técnicas de aprendizado de máquina utilizando o algoritmo **SVM (Support Vector Machine)** para classificar tumores de mama como malignos ou benignos. Para isso, você usará o dataset Wisconsin Breast Cancer, amplamente utilizado na área médica para análise de dados diagnósticos. O objetivo é treinar e avaliar um modelo de SVM, explorando diferentes configurações de kernel e métricas de avaliação.

INSTRUÇÕES

Configuração do Ambiente no Google Colab

- Abra um novo notebook no Google Colab.
- Instale e importe as bibliotecas necessárias para o desenvolvimento do modelo:
 - `from sklearn.datasets import load_breast_cancer`

Carregamento e Exploração do Dataset

- Utilize o dataset Wisconsin Breast Cancer, que pode ser carregado diretamente da biblioteca sklearn
- Visualize as primeiras linhas do dataset para entender sua estrutura

Divisão dos Dados e Padronização:

- Divida os dados em conjuntos de treino (70%) e teste (30%) utilizando a função `train_test_split`
- Padronize os dados para melhorar o desempenho do modelo utilizando `StandardScaler`

Treinamento do Modelo SVM

- Utilize o SVM com o kernel linear para treinar o modelo com os dados padronizados

Avaliação do Modelo

- Após o treinamento, faça as previsões para os dados de teste
- Avalie o desempenho do modelo utilizando as métricas de precisão, recall, F1-score e a matriz de confusão

Exploração de Diferentes Kernels:

- Altere o kernel do modelo para explorar outras abordagens (ex.: rbf, poly) e compare os resultados com o modelo linear
- Altere os valores de c e γ para os mesmos utilizados em testes anteriores

Perguntas para Reflexão:

- Qual foi o desempenho do modelo SVM com o kernel linear? Compare as métricas de acurácia, precisão, recall e F1-score.
- Como o desempenho do modelo mudou ao utilizar um kernel não-linear (por exemplo, rbf ou poly)? Quais melhorias ou desafios surgiram?
- A padronização dos dados influenciou de alguma forma o desempenho do modelo? O que aconteceria se os dados não fossem padronizados?
- Quais outras estratégias de ajuste de hiperparâmetros ou pré-processamento poderiam ser aplicadas para melhorar o desempenho do SVM?

Generalização:

- Aplique esse mesmo processo em qualquer outro conjunto de dados do seu interesse que ainda não trabalhamos.