



Pontifícia Universidade Católica do Paraná
Escola Politécnica
Disciplina: Engenharia da Computação (EC)

TRABALHO ML – APRENDIZAGEM DE MÁQUINA – CLASSIFICADORES – CÓDIGO PYTHON

DESCRIÇÃO DA ATIVIDADE:

Exercício prático usando o Python, Scikit Learn e Google Colab.

Adapte o script em Python disponibilizado para treinamento de Árvores de Decisão e treine um classificador para a base de dígitos disponível no repositório da Scikit Learn. Compare os resultados utilizando as duas avaliações: *holdout* (na proporção de 70/30 – base de treinamento/base de testes) e validação cruzada (com 10 partições). Apresente a taxa de acerto e matriz de confusão de cada experimento, e realize uma análise dos resultados, com ênfase para os tipos de erros cometidos entre as classes.

Repita o procedimento usando uma segunda base de dados, à escolha (para classificação).

Link para a base Digits disponível no SCIKIT Learn

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_digits.html#sklearn.datasets.load_digits

Dicas:

O script fornecido foi originalmente implementado considerando a base Wine disponível no repositório da UCI. Este deverá ser alterado considerando a base de classificação sugerida. As equipes deverão ser capazes de adaptar o script, considerando mudar o esquema de leitura. A base sugerida tem função de leitura específica na Scikit Learn.

Após a adaptação do script para um novo problema, as equipes deverão avaliar o desempenho do classificador e realizar análise de erros via matriz de confusão.

Apresente os resultados dos testes:

- Taxa de acerto (geral e por classe);
- Matriz de confusão: Importante destacar as principais confusões na matriz e tentar encontrar uma linha de raciocínio que as explique (se possível).

Entregar o script e os resultados no Canvas, usando o link fornecido.