



Sistemas de Apoio à Decisão

PROVA SEM CONSULTA

DURAÇÃO 1,5 HORAS

2017/18

PARTE TEÓRICA

- 1) [3 Val] Explique sucintamente o que é um mecanismos de ELT - Extract, Load and Transform. Explique ainda porque estes mecanismos têm vindo a ser substituídos por mecanismos de ETL (Extraction, Transformation and Loading).
- 2) [3 Val] Distinga Regressão e Classificação na área de Machine Learning. Distinga ainda classificadores supervisionados de não supervisionados.
- 3) [3 Val] Explique a diferença entre Algoritmos Discriminativos e Algoritmos Generativos; a qual das classes pertence o método de Naive Bayes? Porque razão se chama Naive ao método de Naive Bayes?
- 4) [3.5 Val] Explique por palavras suas como funciona o Classificador Random Forrest e como funciona. Qual o problema principal que o Random Forrest vem resolver, face ao classificador Decision Tree?
- 5) [1.5 Val] A seguinte expressão que formula um problema de regressão linear

$$\hat{w} = \min_w \frac{1}{2} \sum_i^N (w_0 + w_1 x_i - y_i)^2$$

em que os w são os parâmetros e os y são os targets. Explique o problema formulado.

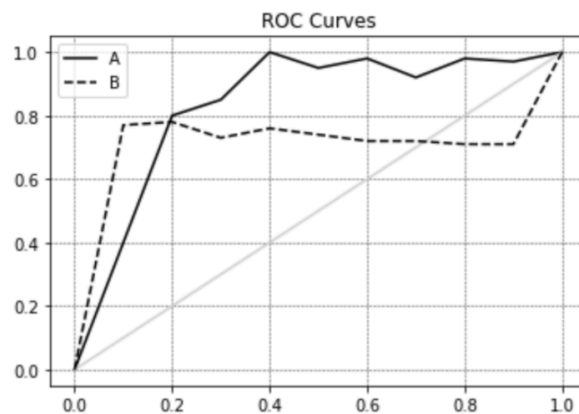
PARTE PRÁTICA

- 6) [3 Val] Considere os seguintes dados de um Data Set com a respetiva classe, que classifica o risco de devolução de um produto nos primeiros 10 dias.

Tipo	Marca	Tipo de Venda	Classe
Electrónica	Apple	Débito	Baixo
Electrónica	Apple	Crédito	Alto
Livros	Porto Editora	Débito	Baixo
Roupa	Levis	Débito	Baixo
Livros	Pearson	Débito	Baixo
Electrónica	Apple	Débito	Alto
Livros	Porto Editora	Débito	Baixo
Roupa	Levis	Crédito	Alto
Roupa	Levis	Crédito	Baixo

- a. Verifique se o melhor split é usando “Tipo” ou “Tipo de Venda”, calculando o gini Index do split resultante para cada um dos casos.

- 7) [3 Val] Considere as seguintes ROC curves:



- a. Diga o que é a “linha cinzenta” e porque razão as ROC curves são normalmente representadas acima dessa linha.
- b. Qual dos algoritmos usaria, se fossem proibitivas taxas de falsos positivos (FPR) elevadas?