

## Análise de Dados e Aprendizagem Automática

### Trabalho Prático 1

O objetivo deste trabalho é utilizar algoritmos de regressão para treinar modelos preditivos relativos ao conjunto de dados denominado **fishcatch.csv**, segundo as questões abaixo. As informações sobre os dados e a sua origem encontram-se no ficheiro denominado **fishcatch.pdf**.

#### 0. Análise exploratória (1.5 valores)

Comece por ler o ficheiro de dados e analisar o seu conteúdo, extraíndo toda a informação que lhe parecer relevante. Não se limite a usar as instruções que aprendeu, comente os resultados e torne a sua análise o mais rica possível.

#### 1. Regressão Linear 1 (6.0 valores)

- Construir um modelo de regressão linear para prever a variável “Width” usando todas as características numéricas disponíveis.
- Obter os coeficientes do modelo e identificar as características mais/menos relevantes para o modelo. Comentar face à análise inicial.
- Obter as métricas do modelo (treino e teste) e comentar a qualidade do mesmo.

#### 2. Regressão Linear 2 (2.5 valores)

- Construir um novo modelo de regressão linear depois de excluir uma das características numéricas disponíveis. Escolher criteriosamente a característica a excluir e justificar a escolha.
- Obter as métricas do modelo e comentar com os valores obtidos em 1. É o que se esperava? Justifique.

#### 3. Regressão Polinomial (5.0 valores)

- Escolher uma única característica para criar um modelo de regressão polinomial simples. Justificar a escolha da característica.
- Registar as métricas do modelo e comparar com o que se obteria num modelo linear apenas com essa característica.
- Construir alguns modelos fazendo variar o grau do polinómio. Qual o grau de polinómio que melhor se adequa ao problema?
- Construir um gráfico de dispersão (variável “Width” em função da característica escolhida) e representar cada um dos modelos treinados neste ponto. Comentar o gráfico.

#### 4. Regressão Linear e Polinomial

- Construir um modelo misto - linear e polinomial. Para a componente polinomial usar a característica selecionada no ponto anterior, e o grau de polinómio considerado óptimo. Para a componente linear, usar as restantes características numéricas.
- Registar as métricas do modelo e comparar com o modelo inicial (ponto 1).

O trabalho deve ser entregue pelo PAE no formato Jupyter Lab (\*.ipynb), onde se inclui o código e toda a análise que permita responder às questões propostas.