

Licenciatura em Engenharia Informática – DEI/ISEP
Análise de Dados em Informática

Ficha Teórico-Prática 7

Classificação: Rede Neuronal

Objetivos:

- Modelos de Redes Neurais, usando Python;
- Análise e discussão dos resultados.

1. O betão é um dos materiais mais importantes em engenharia civil. A resistência à compressão do betão é uma função altamente não linear da sua idade e dos seus componentes. O objetivo é desenvolver modelos que prevejam a resistência (*strength*) do betão em função dos seguintes atributos previsores:

cement	cimento (kg / m3)
slag	cinza volante (kg / m3)
ash	escória de alto-forno (kg / m3)
water	água (kg / m3)
superplastic	superplastificante (kg / m3)
coarseagg	agregado grosso (kg / m3)
fineagg	agregado fino (kg / m3)
age	idade do teste (dias)

- a) Comece por carregar o ficheiro ("`concrete.csv`"), verifique a sua dimensão e obtenha um sumário dos dados.
- b) Usando os gráficos apropriados explore os vários atributos do conjunto de dados.
- c) Verifique se os dados precisam ser normalizados. Em caso positivo elabore uma função para realizar a normalização min-max que mapeia os valores das variáveis no intervalo [0-1]:

$$y' = \frac{y - \min_y}{\max_y - \min_y}$$

- d) Separe o conjunto de dados em dois subconjuntos treino e teste, segundo o critério *holdout* (70% treino/30% teste).
 - e) Treine uma rede neuronal usando a função *MLPRegressor* da biblioteca **Scikit-Learn** para a prever a resistência (*strength*) do betão. Comece por considerar como configuração da rede apenas um nó na camada interna. Avalie as previsões da rede usando o conjunto de teste.

 - f) Repita a alínea anterior considerando as seguintes configurações da rede:
 - i) 1 nível interno, com 3 nós
 - ii) 2 níveis internos, com 6 e 3 nós, respetivamente
 - g) Usando o método de treino *k-fold cross validation* obtenha modelos de previsão de *strength* com:
 - i) Uma rede neuronal com a configuração anterior com melhor desempenho
 - ii) Um modelo de regressão linear múltipla
 - iii) Um modelo árvore de regressãoE obtenha a média e o desvio padrão do RMSE de cada modelo.
 - h) Verifique se a diferença de desempenho entre os dois melhores modelos obtidos anteriormente é estatisticamente significativa.
-
- 2. Qual a diferença entre uma rede **feed-forward** e uma **rede recorrente**?
 - 3. Considerando o dataset "*BreastCancer*" utilizado na Ficha PL6, carregue o dataset e realize a análise descrita no exercício 1, utilizando a biblioteca *Keras* para a construção dos modelos das Neural Networks.