



Curso: Mestrado Integrado em Engenharia Informática
U.C.: Descoberta do Conhecimento

Ficha de Exercícios 07	
Docente:	Hugo Peixoto José Machado
Tema:	RapidMiner - Regressão Linear
Turma:	PL
Ano Letivo:	2020-2021 – 2º Semestre
Duração da aula:	2 horas

1. Parte I

- [1] Que tipo de dados a regressão linear espera para todos os atributos? Qual o tipo de dados do atributo previsto quando este for calculado?
- [2] Porque é que os intervalos de atributos são tão importantes ao realizar data mining através de regressão linear?
- [3] O que são coeficientes de regressão linear? O que significa 'peso', neste contexto?
- [4] Qual é a fórmula matemática de regressão linear e como é organizada?
- [5] Como é que resultados da regressão linear podem ser interpretados/avaliados?

2. Parte II

- [1] Importe o dataset definido para a FE07, fe07-dataset.csv para o RapidMiner. Execute a componente de Data Understanding e DataPreparation.
- [2] Repita os passos no RapidMiner tal como descritos nos slides da aula e após executar o seu modelo, na secção dos resultados, examine os coeficientes dos atributos e as previsões para os custos das casas.
- [3] Relate seus resultados:
 - (a) Execute a avaliação e documentação dos seus resultados-
 - (b) Que atributos têm maior peso?
 - (c) Algum atributo foi removido do conjunto de dados por não ter uma boa capacidade de previsão? Em caso afirmativo, quais e por que você acha que eles não eram eficazes na previsão?
 - (d) Que outros atributos acha que ajudariam o seu modelo a prever melhor o preço da casa?



3. Parte III

[5] Encontre um dataset online que cumpra os requisitos para executar um modelo de regressão linear.

De forma sucinta execute novamente todo o processo executado até aos resultados, avaliando os modelos atingidos.