Aprendizagem 2023

Lab 4: Linear Regression and *k*NN

Prof. Rui Henriques

Practical exercises

I. Lazy learning

1. Consider the following data:

	input		output	
	y_1	y_2	y_3	y_4
$\mathbf{x}_1$	1	1	A	1.4
$\mathbf{x}_2$	2	1	B	0.5
$\mathbf{x}_3$	2	3	B	2
$\mathbf{x}_4$	3	3	B	2.2
$\mathbf{x}_5$	2	2	A	0.7
$\mathbf{x}_6$	1	2	A	1.2

Assuming a *k*-nearest neighbor with *k*=3 applied within a leave-one-out schema:

a) Let y_3 be the output variable (*categoric*). Considering an Euclidean (*l*2) distance, provide the classification estimates for x_1.

b) Let y_4 be the output variable (*numeric*). Considering cosine similarity, provide the mean regression estimate for x_1.

c) Consider a weighted-distance *k*-nearest neighbor with Manhattan (l_1) distance, identify the:

i. weighted mode estimate of x_1 for y_3 outcome

ii. weighted mean estimate of x_1 for y_4 outcome

II. Linear regression

1. Considering the following data to learn a model $z = w_1 y_1 + w_2 y_2 + \varepsilon$, where $\varepsilon \sim N(0, 0.1)$

	y_1	y_2	output
x_1	3	-1	2
x_2	4	2	1
x_3	2	2	1

Compare:

a) $\boldsymbol{w} = [w_1\ w_2]^T$ using the maximum likelihood approach

b) $\boldsymbol{w}$ using the Bayesian approach, assuming $p(\boldsymbol{w}) = N\left(\boldsymbol{w} \mid \boldsymbol{u} = [0\ 0], \boldsymbol{\sigma} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}\right)$