Assignment 02

To be solved INDIVIDUALLY or in GROUP of at most two elements

Submit by nov 3, 2019, 23h59 by email to jaime.cardoso@fe.up.pt

1. Regression.

Consider the following data

x_1	x_2	y (output)
368	15	1.7
340	16	1.5
665	25	2.8
954	40	5
331	15	1.3

a) What's the regression solution for $f(y)=w_1x_1+w_2x_2$?

b) Trying to improve the fitting, we collect another feature x_3:

x_1	x_2	x_3	y (output)
368	15	383	1.7
340	16	356	1.5
665	25	690	2.8
954	40	994	5
331	15	346	1.3

What's now the solution for $f(y)=w_1x_1+w_2x_2+w_3x_3$? Is it unique?

c) In some contexts, it is interesting to introduce different costs per example in the error function: $L(w) = \frac{1}{2}\sum_{n=1}^{N} c_n(y_n - w_0 - w^t x_n)^2$

with $x_n, w \in R^d$ and $c_n \in R^+, w_0 \in R$

Generalize the Probabilistic Interpretation as given in slide 11 of lecture 2 to motivate the given loss function.

2. **Classification.** Consider the data in 'heightWeightData.txt'. The first column is the class label (1=male, 2=female), the second column is height, the third weight.

a) Write a Matlab/Python function to model each class data as a bi-dimensional Gaussian distribution with the mean and variance matrix learnt from the data using maximum likelihood estimation. The function should receive as input the training data and the test data, making prediction (male/female) for the test point, using the **maximum a posterior probability**.

b) What's the estimated $p([165\ 80]^t \mid \text{male})$?

c) Use the previous function to make predictions (male / female) for the following test points:

$[165\ 80]^t$, $[181\ 65]^t$, $[161\ 57]^t$ and $[181\ 77]^t$.

d) Repeat b) and c) using as features D=H-W and S= (H+W)/2, where H and W are the height and weight, respectively. (Note: use the same function from a) but changing the input data).

e) Find a pair of features Z_1 and Z_2, linearly related with H and W, with a diagonal covariance matrix.