Assignment 05

To be solved individually

Submit by email to jaime.cardoso@fe.up.pt

PART A: submit by 30 Dec, 2018

1. Consider again the dataset with heights and weights from previous exercises. Ignore the column with the male/female class.
a) Perform a clustering with k-means, setting K=2. How does the obtained clustering compare with the data partition as given by the labels male/female?
b) Perform a clustering with the x-means algorithm.
https://www.cs.cmu.edu/~dpelleg/download/xmeans.pdf
You may use any implementation of your choice of xmeans (include it in your submission). How many clusters are automatically learnt by xmeans? What's the criterion to choose the number of clusters?

2. Consider clustering 1D data with a mixture of 2 Gaussians using the EM algorithm. You are given the 1-D data points x = [3 10 30]. Suppose the output of the E step is the following matrix:

$$R = \begin{bmatrix} 1 & 0 \\ 0.4 & 0.6 \\ 0 & 1 \end{bmatrix}$$

where entry $r_{i,c}$ is the probability of observation x_i belonging to cluster c (the responsibility of cluster c for data point i). You just have to compute the M step. You may state the equations for maximum likelihood estimates of these quantities (which you should know) without proof; you just have to apply the equations to this data set. Show your work.

a) Write down the likelihood function you are trying to optimize.

b) After performing the M step for the mixing weights π_1, π_2, what are the new values?

c) After performing the M step for the means μ_1 and μ_2, what are the new values?

PART B: submit by 7 Jan, 2019

3. Consider a HMM with continuous outputs. Assume that the HMM only has two states; both states have the same initial probability; the probability of changing between different states is 0.1 (that is, the transition matrix is symmetric with 0.9 in the elements in the main diagonal). The emission density function for state 1 follows a **Laplace** distribution with mean 0 and standard deviation 0.2. For state 2 the emission density function is uniform in [0, 1]. Change the code provided in the class (file hmmTest.m / hmmTest.py) to compute the probability of the following sequence of length 10:

{0.7, 0.7, 0.1, 0.2, 0.3, 0.6, 0.2, 0.3, −0.1, 0.2}.

Send the modified file and write the computed probability.