Exercise 5 (**python code**):

Consider the two data sets $X_1$and X_2 contained in the attached file "Dataset.mat", each one of them containing 4-dimensional data vectors, in its rows. The vectors of X_1 stem from the pdf $p_1(\boldsymbol{x})$, while those of X_2 stem from the pdf $p_2(\boldsymbol{x})$.

(a) Based on X_1, estimate the values of $p_1(\boldsymbol{x})$ at the following points:
$\boldsymbol{x}_1 = (2.01, 2.99,\ 3.98,\ 5.02)$, $\boldsymbol{x}_2 = (20.78,\ -15.26,\ 19.38,\ -25.02)$,
$\boldsymbol{x}_3 = (3.08, 3.88,\ 4.15,\ 6.02)$.

(b) Based on X_2, estimate the values of $p_2(\boldsymbol{x})$ at the following points:
$\boldsymbol{x}_1 = (0.05, 0.15, -0.12, -0.08)$, $\boldsymbol{x}_2 = (7.18,\ 7.98,\ 9.12, 9.94)$, $\boldsymbol{x}_3 = (3.48, 4.01,\ 4.55,\ 4.96)$, $\boldsymbol{x}_4 = (20.78,\ -15.26,\ 19.38,\ -25.02)$.

In both the above cases use a parametric and the non-parametric approach.

Hints:

- To load the data sets use the script "HW6.ipynb".
- Use the Sklearn.mixture.GaussianMixture class (https://scikit-learn.org/stable/modules/generated/sklearn.mixture.GaussianMixture.html), if you are willing to use Gaussian mixtures modelling.

It could be proved useful for the modelling of each pdf to compute the mean of each data set and then to consider the distances of the data vectors from it. However, other methods can also be applied.