“Machine Learning and Computational Statistics”

7th Homework

Exercise 1:

Consider the case where the data at hand are modeled by a pdf of the form

$$p(\boldsymbol{x}) = \sum_{j=1}^{m} P_j\, p(\boldsymbol{x} \mid j), \quad \sum_{j=1}^{m} P_j = 1, \quad \int_{-\infty}^{+\infty} p(\boldsymbol{x} \mid j) = 1$$

where $P_j, j = 1, \dots, m$, are the a priori probabilities of the pdfs $p(\boldsymbol{x}|j)$, which involved in the definition of $p(\boldsymbol{x})$. In the “parameter updating” part of the EM-algorithm, which allows the estimation of the parameters of $p(\boldsymbol{x}|j)$’s as well as P_j’s, we need to solve the problem

$$[P_1, P_2, \dots, P_m] = argmax_{[P_1,P_2,\dots,P_m]} \sum_{i=1}^{N} \sum_{j=1}^{m} P(j|\boldsymbol{x_i}) \ln P_j\,, \ subject\ to\ \sum_{j=1}^{m} P_j = 1,$$

for fixed $P(j|\boldsymbol{x_i})$’s. Prove that, independently of the form adopted for each $p(\boldsymbol{x}|j)$, the solution of the above problem is

$$P_j = \frac{1}{N} \sum_{i=1}^{N} P(j|\boldsymbol{x_i})\,,\ j = 1, \dots, m.$$

<u>*Hint:*</u> In this case we have an equality constraint. Work as follows:

1. Define the Lagrangian function
   $L(P_1, P_2, \dots, P_m) = \sum_{i=1}^{N} \sum_{j=1}^{m} P(j|\boldsymbol{x_i}) \ln P_j + \lambda(\sum_{j=1}^{m} P_j - 1),$
2. Solve the equations $\frac{\partial L(P_1,P_2,\dots,P_m)}{\partial P_j} = 0, j = 1, \dots, m$, expressing each P_j in terms of λ.
3. Substitute P_j’s in the constraint equation $\sum_{j=1}^{m} P_j = 1$ and solve with respect to λ.
4. Compute P_j’s from the equations derived in step 2 above.

<u>*Note:*</u> In the case of equality constraints, the final solution **does not** involve the Lagrangian multipliers.

Exercise 2:

Consider again the setup of exercise 1, where now $p(\boldsymbol{x}|j)$’s are normal distributions with means $\boldsymbol{\mu}_j$ and **fixed** covariance matrices $\Sigma_j, j = 1, \dots, m$. Prove that the solution of the optimization problems

$$\boldsymbol{\mu}_j = argmax_{\boldsymbol{\mu}_j} \sum_{i=1}^{N} P(j|\boldsymbol{x_i}) \ln\left(p(\boldsymbol{x}_i|j;\boldsymbol{\mu}_j)\right), j = 1, \dots, m$$ [1]

is

$$\boldsymbol{\mu}_j = \frac{\sum_{i=1}^{N} P(j|\boldsymbol{x_i})\boldsymbol{x_i}}{\sum_{i=1}^{N} P(j|\boldsymbol{x_i})}, j = 1, \dots, m$$

<u>Hint:</u> Take the gradient of $\sum_{i=1}^{N} P(j|\boldsymbol{x_i}) \ln\left(p(\boldsymbol{x}_i|j;\boldsymbol{\mu}_j)\right)$ with respect to $\boldsymbol{\mu}_j$, set it equal to zero and solve for $\boldsymbol{\mu}_j$.

<u>Exercise 3</u> (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 onX_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 onX_2, estimate the values of $p_2(\boldsymbol{x})$ at the following points:
$\boldsymbol{x}_1 = (0.05, 0.15, -0.12, -0.08)$, $\boldsymbol{x}_2 = (7.18,\ 7.98,\ 9.12, 9.94)$, $\boldsymbol{x}_3 = (3.48, 4.01,\ 4.55,\ 4.96)$, $\boldsymbol{x}_4 = (20.78,\ -15.26,\ 19.38,\ -25.02)$.

<u>Hints:</u>

- 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.

[1] Note that, since Σ_j is fixed, it is $\boldsymbol{\theta}_j \equiv \boldsymbol{\mu}_j$.

<u>Exercise 4:</u> Consider a two-dimensional[2] three-class problem, where the classes ω_1, ω_2 and ω_3 are equiprobable[3] and are modeled by the normal distributions $p(\boldsymbol{x}|\omega_1) = N(\boldsymbol{\mu}_1, \Sigma_1)$, $p(\boldsymbol{x}|\omega_2) = N(\boldsymbol{\mu}_2, \Sigma_2)$ and $p(\boldsymbol{x}|\omega_3) = N(\boldsymbol{\mu}_3, \Sigma_3)$, respectively, where $\boldsymbol{\mu}_1 = [0, 0]^T$, $\boldsymbol{\mu}_2 = [0, 3]^T$, $\boldsymbol{\mu}_3 = [3, 0]^T$ and $\Sigma_1 = \Sigma_2 = \Sigma_3 = I$, with I being the 2x2 identity matrix.

(a) Utilizing the Bayes decision rule, classify each one of the data points $\boldsymbol{x}_1 = [1, 1]^T$, $\boldsymbol{x}_2 = [1, 2]^T$, $\boldsymbol{x}_3 = [2, 1]^T$ to one out of the three classes.

(b) Plot the points $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$ on a two-dim. axis coordinate system and

  i. Determine the point for which the Bayes decision rule cannot assign it to anyone of the three classes. What would be the result of the Bayes decision rule for this point if the a priori class probabilities were $P(\omega_1) = 0.5$, $P(\omega_2) = 0.3$ and $P(\omega_3) = 0.2$?

  ii. Determine the regions of the points in the plane that will be classified by the Bayes decision rule to each one of the three classes (give a short explanation).

<u>Hint:</u> (a) For each point $\boldsymbol{x}_i$ to be classified determine the a posteriori probabilities $P(\omega_j|\boldsymbol{x}_i)$, $j = 1,2,3$, utilizing the Bayes rule $P(\omega_j|\boldsymbol{x}_i) = \frac{p(\boldsymbol{x}_i|\omega_j) \cdot P(\omega_j)}{\sum_{k=1}^{m} p(\boldsymbol{x}_i|\omega_k) \cdot P(\omega_k)}$ and the apply the Bayes decision rule. Alternatively, the quantities $p(\boldsymbol{x}_i|\omega_j) \cdot P(\omega_j)$ could be considered.

(b) Geometrical arguments could be proved helpful for this question.

[2] That is, two features are used for characterizing each entity. In other words, the feature vector characterizing each entity is two-dimensional.

[3] This means that the a priori probabilities for all classes are equal to each other, that is, $P(\omega_1) = P(\omega_2) = P(\omega_3)$.