

"Machine Learning and Computational Statistics"

7th Homework

Exercise 1:

Consider a two-class 1-dim. classification problem of two equiprobable classes ω_1 and ω_2 that are modeled by the normal distributions $N(0,1)$ and $N(0,5)$, respectively. Determine the decision regions R_1 and R_2 corresponding to the two classes.

Exercise 2:

(a) Consider a three-class 1-dim. problem where the classes ω_1 , ω_2 και ω_3 are modeled by the following uniform distributions

$$p(x|\omega_1) = \begin{cases} 1/5, & x \in (0,2) \cup (5,8) \\ 0, & \text{otherwise} \end{cases}$$

$$p(x|\omega_2) = \begin{cases} 1/9, & x \in (0,9) \\ 0, & \text{otherwise} \end{cases}$$

$$\text{και } p(x|\omega_3) = \begin{cases} 1, & x \in (3,4) \\ 0, & \text{otherwise} \end{cases}$$

(I) Assume that all classes are equiprobable.

- (i) Depict graphically in the same figure $P(\omega_i)p(x|\omega_i)$ (as functions of x) and identify the respective decision regions, as they are specified by the Bayes classifier.
- (ii) Compute the error classification probability of the Bayes classifier.
- (iii) Classify the point $x' = 3.5$ to one of the three classes using the Bayes classifier.

(II) Assume that the classes are not equiprobable.

- (i) Determine a set of values for the a priori probabilities of the three classes that guarantee that $x' = 3.5$ is assigned to class ω_2 . Justify briefly your choice.
- (ii) Is there any combination of the a priori probabilities that guarantees that $x' = 3.5$ will be assigned to ω_1 ? Explain.

Hints:

(H1) Focus only in the interval $[0,10]$ since all pdfs are zero out of this interval.

(H2) The error classification probability for the Bayes classifier is

$$P_e = \sum_{i=1}^M \int_{R_i} \left(\sum_{k=1, k \neq i}^M p(x|\omega_k)P(\omega_k) \right) dx$$

Exercise 3 (python code + text):

Consider a **three-class, four-dimensional** classification problem for which you can find attached two **sets**: one for **training** and one for **testing** (file [HW7.mat](#)). Each of these sets consists of pairs of the form (y_i, \mathbf{x}_i) , where y_i is the **class label** for vector \mathbf{x}_i . Let N_{train} and N_{test} denote the number of training and test sets, respectively. The data are given via the following arrays/matrices:

- ***train_x*** (a $N_{train} \times 4$ **matrix** that contains in its **rows** the **training** vectors \mathbf{x}_i)
- ***train_y*** (a N_{train} -dim. column **vector** containing the **class labels** (1, 2 or 3) of the corresponding **training** vectors \mathbf{x}_i included in ***train_x***).
- ***test_x*** (a $N_{test} \times 4$ **matrix** that contains in its **rows** the **test** vectors \mathbf{x}_i)
- ***test_y*** (a N_{test} -dim. column **vector** containing the **class labels** (1, 2 or 3) of the corresponding **test** vectors \mathbf{x}_i included in ***test_x***).

Adopt the **Bayes classifier** under the following two scenarios:

- (i) $p(\mathbf{x} | \omega_1)$, $p(\mathbf{x} | \omega_2)$ and $p(\mathbf{x} | \omega_3)$ are treated via the **parametric approach**
- (ii) $p(\mathbf{x} | \omega_1)$, $p(\mathbf{x} | \omega_2)$ and $p(\mathbf{x} | \omega_3)$ are treated via the **non-parametric k-NN density estimation approach**.

For each of the above cases use the training set to **estimate** $P(\omega_1)$, $P(\omega_2)$, $P(\omega_3)$, $p(\mathbf{x} | \omega_1)$, $p(\mathbf{x} | \omega_2)$, $p(\mathbf{x} | \omega_3)$. Then

(a) Classify the points \mathbf{x}_i of the test set, using the **Bayes classifier** (for each point apply the Bayes classification rule and keep the class labels, to an a N_{test} -dim. column **vector** , called ***Btest_y*** containing the **estimated class labels** (1, 2 or 3) of the corresponding **test** vectors \mathbf{x}_i included in ***test_x***) and

(b) Estimate the **error classification probability** based on the test set classification results.

Hint: After downloading the attached MATLAB file, use the attached python code to retrieve the above mentioned matrices and vectors: