“Machine Learning and Computational Statistics”

7th Homework

Exercise 1 (multiple choices question): Given an M-class classification problem, where the classes are denoted as ω_i, $i = 1, \ldots, M$, let $p(\boldsymbol{x}|\omega_i)$ be the probability density function describing class ω_i, $P(\omega_i)$ be the a priori class probability of ω_i and $P(\omega_i|\boldsymbol{x})$ be the posterior probability of ω_i, given $\boldsymbol{x}$. Consider a specific feature vector $\boldsymbol{x}$. Then, according to the general form of the Bayes classification rule, (choose the correct answer(s))

1. $\boldsymbol{x}$ is assigned to ω_i for which $P(\omega_i|\boldsymbol{x}) = max_{j=1,\ldots,M} P(\omega_j|\boldsymbol{x})$

2. $\boldsymbol{x}$ is assigned to ω_i for which $P(\omega_i) = max_{j=1,\ldots,M} P(\omega_j)$

3. $\boldsymbol{x}$ is assigned to ω_i for which $p(\boldsymbol{x}|\omega_i) = max_{j=1,\ldots,M} p(\boldsymbol{x}|\omega_j)$

4. $\boldsymbol{x}$ is assigned to ω_i for which $P(\omega_i) \cdot p(\boldsymbol{x}|\omega_i) = max_{j=1,\ldots,M} P(\omega_j) \cdot p(\boldsymbol{x}|\omega_j)$

Exercise 2 (multiple choices question): Given an M-class classification problem, where the classes are denoted as ω_i, $i = 1, \ldots, M$, let $p(\boldsymbol{x}|\omega_i)$ be the probability density function describing class ω_i, $P(\omega_i)$ be the a priori class probability of ω_i and $P(\omega_i|\boldsymbol{x})$ be the posterior probability of ω_i, given $\boldsymbol{x}$. Consider a specific feature vector $\boldsymbol{x}$. Then, according to the general form of the Bayes classification rule, (choose the correct answer(s))

1. It is possible for $\boldsymbol{x}$ to be assigned to a class ω_i for which $P(\omega_i) = 0$.

2. $\boldsymbol{x}$ is assigned to ω_i for which $P(\omega_i) = max_{j=1,\ldots,M} P(\omega_j)$, provided that all $p(\boldsymbol{x}|\omega_i)$’s coincide.

3. $\boldsymbol{x}$ is assigned to ω_i for which $p(\boldsymbol{x}|\omega_i) = max_{j=1,\ldots,M} p(\boldsymbol{x}|\omega_j)$, provided that all classes are equiprobable.

4. If $P(\omega_i) = max_{j=1,\ldots,M} P(\omega_j)$, $\boldsymbol{x}$ will be assigned to ω_i even if $p(\boldsymbol{x}|\omega_i) = 0$.

Exercise 3 (multiple choices question): Consider a two-class classification problem, which refers to the funs of two football teams (A and B) that are in the stadium to watch the match between the two teams. The colors of the two teams are blue and red, respectively. Class ω_1 is associated with the funs of team A and ω_2 with the funs of team B that are now in the stadium. Assume that the associated prior probabilities are $P(\omega_1) = 0.7$ and $P(\omega_2) = 0.3$. Which of the following statements is/are true?

1. If we randomly pick a fun from those that are in the stadium, then, we have 70% certainty that he/she is a fun of team A.

2. If we randomly pick a fun from those that are watching the match in tv, then, we have 70% certainty that he/she is a fun of team A.

3. If we randomly pick a fun from those that are in the stadium **and** wear a red scarf, then, we have more that 30% certainty that he/she is a fun of team A.

4. If we pick a fun from those that are in the stadium with red scarf, then, we are 100% sure that is a fun of team A.

Exercise 4 (multiple choices question): Given a two-class classification problem, where the classes are denoted as $\omega_i,\ i = 1{,}2$, let $p(\boldsymbol{x}|\omega_i)$ be the probability density function describing class ω_i, $P(\omega_i)$ be the a priori class probability of ω_i and $P(\omega_i|\boldsymbol{x})$ be the posterior probability of ω_i, given $\boldsymbol{x}$. Also, let $y \in \{1{,}2\}$ denotes the class label. Consider a specific feature vector $\boldsymbol{x}$. Which of the following statements is/are true?

1. The joint probability density function $p(y, \boldsymbol{x})$ is completely described by the $p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_2)$.

2. The joint probability density function $p(y, \boldsymbol{x})$ is completely described by the $P(\omega_1)$ and $P(\omega_2)$.

3. The joint probability density function $p(y, \boldsymbol{x})$ is completely described by the products $p(\boldsymbol{x}|\omega_1)P(\omega_1)$ **and** $p(\boldsymbol{x}|\omega_2)P(\omega_2)$.

4. The joint probability density function $p(y, \boldsymbol{x})$ is completely described by the products $p(\omega_1, \boldsymbol{x})P(\omega_1)$ and $p(\omega_2, \boldsymbol{x})P(\omega_2)$.

Exercise 5 (multiple choices question): Consider a three-class one-dimensional classification problem, where the probability density functions, which describe the three classes $\omega_1, \omega_2, \omega_3$, are unit variance normal distributions with mean values $\mu_1 = -3, \mu_2 = 0$ and $\mu_3 = 3$, respectively. Assume also that the classes are equiprobable (the class a priori probabilities are equal to each other). Then, the points $x_1 = -5, x_2 = -1$, $x_3 = 1$, $x_4 = 2$, will be classified by the Bayes decision rule as follows:

1. $x_1 \to \omega_1,\quad x_2 \to \omega_1,\quad x_3 \to \omega_2,\quad x_4 \to \omega_3$

2. $x_1 \to \omega_1,\quad x_2 \to \omega_2,\quad x_3 \to \omega_2,\quad x_4 \to \omega_3$

3. $x_1 \to \omega_1,\quad x_2 \to \omega_2,\quad x_3 \to \omega_3,\quad x_4 \to \omega_3$

4. $x_1 \to \omega_2,\quad x_2 \to \omega_2,\quad x_3 \to \omega_2,\quad x_4 \to \omega_3$

Exercise 6 (multiple choices question): Consider a one-dimensional three-class classification problem where the three classes ω_1, ω_2 and ω_3 are equiprobable (they have equal a priori probabilities) and they are modeled by uniform probability density functions as follows,

$$p(x|\omega_1) = \begin{cases} \frac{1}{9}, & x \in (0,9) \\ 0, & otherwise \end{cases}, \; p(x|\omega_2) = \begin{cases} \frac{1}{3}, & x \in (0,3) \\ 0, & otherwise \end{cases}, \; p(x|\omega_3) = \begin{cases} \frac{1}{4}, & x \in (5,9) \\ 0, & otherwise \end{cases}$$

The Bayesian classifier will assign the points $x_1 = 6$, $x_2 = 2$, $x_3 = 4$ as follows:

1. $x_1 \to \omega_3, \quad x_2 \to \omega_1, \quad x_3 \to \omega_2$

2. $x_1 \to \omega_1, \quad x_2 \to \omega_1, \quad x_3 \to \omega_2$

3. $x_1 \to \omega_3, \quad x_2 \to \omega_2, \quad x_3 \to \omega_1$

4. $x_1 \to \omega_1, \quad x_2 \to \omega_3, \quad x_3 \to \omega_2$

Exercise 7 (multiple choices question): Consider a three-class classification problem, for which 1000 data points are available. From those, 500 points stem from class ω_1, 300 points stem from class ω_2, 200 points stem from class ω_3. Then, the estimated a priori class probabilities $P(\omega_1), P(\omega_2)$ and $P(\omega_3)$ are

1. $P(\omega_1) = 0.3, \; P(\omega_2) = 0.5, P(\omega_3) = 0.2$

2. $P(\omega_1) = 0.5, \; P(\omega_2) = 0.3, P(\omega_3) = 0.2$

3. $P(\omega_1) = 0.2, \; P(\omega_2) = 0.5, P(\omega_3) = 0.3$

4. $P(\omega_1) = 0.3, \; P(\omega_2) = 0.2, P(\omega_3) = 0.5$

Exercise 8 (multiple choices question): Consider a two-class one-dimensional classification problem, with the observations $-0.2, -0.1, 0, 0.1, 0.2$ stemming from class ω_1 and the observations $4.8, 4.9, 5, 5.1, 5.2$ stemming from class ω_2. Assume that the probability density functions for both classes are unit variance normal distributions. Recall, also, that the maximum likelihood estimate of the mean μ of a normal distribution, based on a set $X = \{x_1, \dots, x_N\}$ of observations, is $\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$. Then, the estimated class a priori probabilities ($P(\omega_1)$ and $P(\omega_2)$) and class probability density functions ($p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_2)$) are:

1. $P(\omega_1) = 0.5, P(\omega_2) = 0.5, p(\boldsymbol{x}|\omega_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), p(\boldsymbol{x}|\omega_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-5)^2}{2}\right)$

2. $P(\omega_1) = 0.1, P(\omega_2) = 0.9, p(\boldsymbol{x}|\omega_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-5)^2}{2}\right), p(\boldsymbol{x}|\omega_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

3. $P(\omega_1) = 0.5, P(\omega_2) = 0.5, p(\boldsymbol{x}|\omega_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x+0.1)^2}{2}\right), p(\boldsymbol{x}|\omega_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-4.9)^2}{2}\right)$

4. $P(\omega_1) = 0.5, P(\omega_2) = 0.5, p(\boldsymbol{x}|\omega_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-0.1)^2}{2}\right), p(\boldsymbol{x}|\omega_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-5.1)^2}{2}\right)$

Exercise 9 (multiple choices question): Consider a one-dimensional three-class classification problem where the three classes ω_1, ω_2 and ω_3 are equiprobable (they have equal a priori probabilities) and they are modeled by uniform probability density functions as follows

$$p(x|\omega_1) = \begin{cases} \frac{1}{9}, & x \in (0,9) \\ 0, & otherwise \end{cases}, \; p(x|\omega_2) = \begin{cases} \frac{1}{3}, & x \in (0,3) \\ 0, & otherwise \end{cases}, \; p(x|\omega_3) = \begin{cases} \frac{1}{4}, & x \in (5,9) \\ 0, & otherwise \end{cases}$$

The regions R_1, R_2 and R_3 on which the Bayes classifier decides in favor of ω_1, ω_2 and ω_3, respectively, are

1. $R_1 = (0,9), R_2 = (0,3), R_3 = (5,9)$

2. $R_1 = (3,5), R_2 = (0,3), R_3 = (5,9)$

3. $R_1 = (0,5), R_2 = (0,3), R_3 = (5,9)$

4. $R_1 = (3,9), R_2 = (0,3), R_3 = (5,9)$

Exercise 10 (multiple choices question): Consider a one-dimensional two-class classification problem where the two classes ω_1 and ω_2 are modeled by uniform probability density functions as follows

$$p(x|\omega_1) = \begin{cases} \frac{1}{10}, & x \in (0,10) \\ 0, & otherwise \end{cases}, \; p(x|\omega_2) = \begin{cases} \frac{1}{4}, & x \in (8,12) \\ 0, & otherwise \end{cases}$$

Which choices for the class a priori probabilities ensure that the Bayes classification rule will assign the point $x' = 9$ to class ω_1?

1. $P(\omega_1) = 0.5, P(\omega_2) = 0.5$

2. $P(\omega_1) = 0.3, P(\omega_2) = 0.7$

3. $P(\omega_1) = 0.7, P(\omega_2) = 0.3$

4. $P(\omega_1) = 0.8, P(\omega_2) = 0.2$

Exercise 11 (multiple choices question): Consider an l-dimensional two-class classification problem, which involves two classes, ω_1 and ω_2, and let R_1 and R_2 be the regions in the R^l where we decide in favor of ω_1 and ω_2, respectively. Then, the probability of classification error P_e is

1. $P_e = P(\boldsymbol{x} \in R_1, \boldsymbol{x} \in \omega_1) + P(\boldsymbol{x} \in R_2, \boldsymbol{x} \in \omega_2)$

2. $P_e = P(\boldsymbol{x} \in \omega_2 | \boldsymbol{x} \in R_1) + P(\boldsymbol{x} \in \omega_1 | \boldsymbol{x} \in R_2)$

3. $P_e = P(\boldsymbol{x} \in R_1 | \boldsymbol{x} \in \omega_2) + P(\boldsymbol{x} \in R_2 | \boldsymbol{x} \in \omega_1)$

4. $P_e = P(\boldsymbol{x} \in R_1, \boldsymbol{x} \in \omega_2) + P(\boldsymbol{x} \in R_2, \boldsymbol{x} \in \omega_1)$

Exercise 12 (multiple choices question): Consider a one-dimensional two-class classification problem where the two classes ω_1, and ω_2 are equiprobable (they have equal a priori probabilities) and they are modeled by uniform probability density functions as follows

$$p(x|\omega_1) = \begin{cases} \frac{1}{10}, & x \in (0,10) \\ 0, & otherwise \end{cases}, \quad p(x|\omega_2) = \begin{cases} \frac{1}{4}, & x \in (0,4) \\ 0, & otherwise \end{cases}.$$

Then, the probability of classification error of the Bayes classifier is

1. $P_e = 0.2$

2. $P_e = 0.5$

3. $P_e = 0.3$

4. $P_e = 0.4$

Exercise 13: Consider a two-class 1-dim. problem where the classes ω_1 and ω_2 are modeled by the following uniform distributions

$$p(x|\omega_1) = \begin{cases} 1/5, & x \in (0,1) \cup (4,8) \\ 0, & \text{otherwise} \end{cases} \qquad p(x|\omega_2) = \begin{cases} 1/9, & x \in (0,9) \\ 0, & \text{otherwise} \end{cases}$$

(I) Assume that the classes are equiprobable ($P(\omega_1) = P(\omega_2)$).

(i) Depict graphically in the same figure $P(\omega_i)p(x|\omega_i)$, $i = 1,2$, (as functions of x) and identify the respective decision regions, as they are specified by the Bayes classifier.
(ii) Classify the point $x' = 3.5$ to one of the two 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 two classes that guarantee that $x' = 5$ is assigned to class ω_2. Justify briefly your choice.

(ii) Is there any combination of the class priori probabilities that guarantees that $x' = 3$ 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. Note that $P(\omega_1) + P(\omega_2) = 1$.

(*H2*) For (II-i): Consider the inequality $P(\omega_1)p(x'|\omega_1) < P(\omega_2)p(x'|\omega_2)$ and the fact that $P(\omega_1) + P(\omega_2) = 1$. Work similarly for (II-ii).

Exercise 14: Consider a two-class 1-dim. classification problem of two equiprobable classes ω_1 and ω_2 ($P(\omega_1) = P(\omega_2)$) that are modeled by the normal distributions $N(-1,1)$ and $N(1,4)$, respectively. Depict the quantities $P(\omega_j)p(x|\omega_j)$ for $j = 1,2$, in the same graph and determine the decision regions R_1 and R_2 corresponding to the two classes, according to the Bayes classification rule.

Exercise 15: Consider a two-dimensional[1] two-class problem, where the classes ω_1 and ω_2 are equiprobable[2] and are modeled by the normal distributions $p(\boldsymbol{x}|\omega_1) = N(\boldsymbol{\mu}_1, \Sigma_1)$ and $p(\boldsymbol{x}|\omega_2) = N(\boldsymbol{\mu}_2, \Sigma_2)$, respectively, where $\boldsymbol{\mu}_1 = [6, 0]^T$, $\boldsymbol{\mu}_2 = [0, 6]^T$ and $\Sigma_1 = \Sigma_2 = 2 \cdot I$, with I being the 2x2 identity matrix.

(a) Utilizing the Bayes decision rule, classify each one of the data points $\boldsymbol{x}_1 = [2, 4]^T$, $\boldsymbol{x}_2 = [4, 2]^T$, $\boldsymbol{x}_3 = [2, 2]^T$ to one out of the three classes.

(b) Determine the line that separates the class regions of the two classes

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

(b) Consider the equation $P(\omega_1)p(\boldsymbol{x}|\omega_1) = P(\omega_2)p(\boldsymbol{x}|\omega_2)$ and keep in mind that $\boldsymbol{x} = [x_1, x_2]^T$.

---

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

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

Exercise 16: Consider a two-class one-dimensional classification problem, where the pdfs of the two equiprobable classes ω_1 and ω_2 are

$p(x|\omega_1) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{x^2}{2}\right)$ and $p(x|\omega_2) = \begin{cases} \frac{1}{2\sqrt{2\pi}}, & x \in \left[-\sqrt{2\pi}, \sqrt{2\pi}\,\right] \\ 0, & otherwise \end{cases}$, respectively.

Determine the classifier that minimizes the probability of classification error and the write down the decision regions associated with each class.

Exercise 17: Consider a two-class two-dimensional classification problem for which the data vectors $\boldsymbol{x}_1 = [0,0]^T$, $\boldsymbol{x}_2 = [3,0]^T$, $\boldsymbol{x}_3 = [0,3]^T$, $\boldsymbol{x}_4 = [3,3]^T$, $\boldsymbol{x}_5 = [9,9]^T$, $\boldsymbol{x}_6 = [12,9]^T$, $\boldsymbol{x}_7 = [9,12]^T$, $\boldsymbol{x}_8 = [12,12]^T$ belong to class ω_1, while the data vectors $\boldsymbol{x}_9 = [9,0]^T$, $\boldsymbol{x}_{10} = [12,0]^T$, $\boldsymbol{x}_{11} = [9,3]^T$, $\boldsymbol{x}_{12} = [12,3]^T$ belong to class ω_2.

(a) Classify the data points $\boldsymbol{x} = [4, 1.5]^T$ and $\boldsymbol{x}' = [8, 1.5]^T$ to one of the two classes ω_1 and ω_2, utilizing the Bayesian classification rule, where the values of the pdfs associated with ω_1 and ω_2 at points $\boldsymbol{x}$ and $\boldsymbol{x}'$ will be estimated via the Parzen windows nonparametric approach (with $h = 1$).

(b) What would be the correct hypothesis about the pdfs that model the two classes (this would be of interest if a parametric approach were to be used)?

Hint: For (a): The a priori class probabilities should also be estimated.

For (b): Unimodal or multimodal pdfs and of what type.

Exercise 18 (python code + text):

Consider a two-class, two-dimensional classification problem for which you can find attached two sets: one for training and one for testing (file *HW8.mat*). Each of these sets consists of pairs of the form $(y_i, \boldsymbol{x}_i)$, where y_i is the class label for vector $\boldsymbol{x}_i$. 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 $N_{train} \times 2$ **matrix** that contains in its rows the training vectors $\boldsymbol{x}_i$)

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

Assume that the two classes, ω_1 and ω_2 are modeled by normal distributions.

Adopt the **Bayes classifier**.

i. Use the training set to **estimate** $P(\omega_1), P(\omega_2), p(\boldsymbol{x}|\omega_1), p(\boldsymbol{x}|\omega_2)$ (Since $p(\boldsymbol{x}|\omega_j)$ is modeled a normal distribution, it is completely identified by $\boldsymbol{\mu}_j$ and Σ_j. Use the ***ML*** **estimates** for them as given in the lecture slides).

ii. **Classify** the points $\boldsymbol{x}_i$ 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 or 2) of the corresponding test vectors $\boldsymbol{x}_i$ included in ***test_x.***).

iii. Estimate the error classification probability ((1) **compare** ***test_y*** and ***Btest_y*** , (2) **count** the positions where both of them have the same class label and (3) **divide** with the total number of test vectors).