<u>“Machine Learning and Computational Statistics”</u>

<u>7a</u>

Exercise 1 (multiple choices question): Consider an M-class classification problem, where the involved classes are denoted as $\omega_1, \dots, \omega_M$, and let $R_1, \dots, R_M$ be the regions in the feature space where the Bayes classification rule decides in favor of $\omega_1, \dots, \omega_M$, respectively. Which of the following statements regarding the Bayes classification rule is/are true?

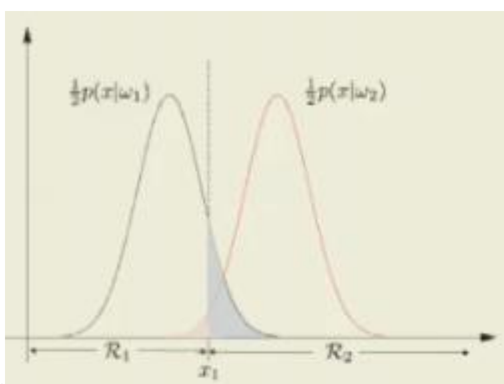
1. The Bayes classification rule minimizes the probability of classification error.

2. The Bayes classification rule only defines the regions $R_1, \dots, R_M$ in the feature space for which it decides in favor of the respective classes $\omega_1, \dots, \omega_M$.

3. The Bayes classification rule, adjusts the probability density functions describing the classes $\omega_1, \dots, \omega_M$.

4. The Bayes classification rule necessarily minimizes the least squares error.

Exercise 2 (multiple choices question): Consider a two-class one-dimensional classification problem (classes ω_1, ω_2). Note that the goal of any classification rule is to partition the feature space in two regions R_1 and R_2, and based in which one of them a pattern (feature) lies, it decides in favor of ω_1 and ω_2, respectively (for an example see figure below). How many classification rules can we define for the above classification problem?



,

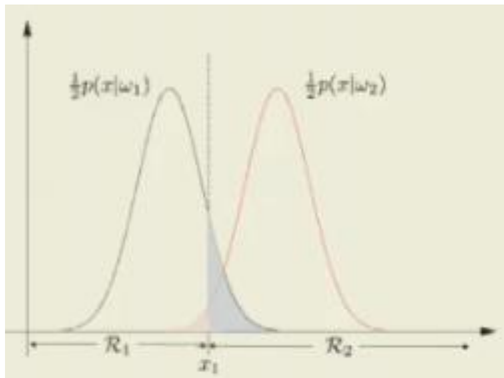
1. Two

2. None

3. One

4. Infinite

Exercise 3 (multiple choices question): Consider a two-class one-dimensional classification problem (classes ω_1, ω_2). Note that any classification rule, actually splits the feature space in two regions R_1 and R_2, and based in which one of them a pattern (feature) lies, it decides in favor of ω_1 and ω_2, respectively (for an example see figure below). How many optimal, with respect to the probability of classification error criterion, classification rules can we define for the above classification problem?



1. Two

2. None

3. One

4. Infinite

Exercise 4 (multiple choices question): Consider a production line that manufactures the brakes that are used in the slowing down aircraft system of a specific type of civil airplanes. Consider the two-class classification task, where the entities are the produced brakes items, which have to be classified as “non-defective” (class ω_1) or “defective” (class ω_2). Which of the following scenarios would probably have very unpleasant consequences?

1. A brake item $\boldsymbol{x}$, is assigned by the classifier to class ω_2, while it stems from class ω_1.

2. A brake item $\boldsymbol{x}$, is assigned by the classifier to class ω_1, while it stems from class ω_1.

3. A brake item $\boldsymbol{x}$, is assigned by the classifier to class ω_2, while it stems from class ω_2.

4. A brake item $\boldsymbol{x}$, is assigned by the classifier to class ω_1, while it stems from class ω_2.

Exercise 5 (multiple choices question): Consider a two-class classification task, and let $L = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$, where λ_{ij} is the relative weight of the error committed when a data vector $\boldsymbol{x}$ is assigned by a classifier to class ω_j (that is, it lies in the region R_j that corresponds to ω_j), while it stems from class ω_i, $i, j = 1,2$. Which of the following statements is/are true?

1. L is a diagonal matrix.

2. Typically, L is an anti-diagonal matrix.

3. L has at least one zero column.

4. L has at least one zero row.

Exercise 6 (multiple choices question): Consider a three-class classification task, and let $L = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{bmatrix}$, where λ_{ij} is the relative weight of the error committed when a data vector, $\boldsymbol{x}$, is assigned by a classifier to class ω_j (that is, it lies in the region R_j that corresponds to ω_j), while it stems from class ω_i, $i, j = 1, 2, 3$. Which of the following statements is/are true?

1. L is a diagonal matrix.

2. Typically L is an anti-diagonal matrix.

3. Typically, the diagonal elements of L are zero.

4. L has at least one zero row.

Exercise 7 (multiple choices question): Consider an one-dimensional two-class classification problem where the involved classes are denoted as ω_1 and ω_2, and let R_1 and R_2 be their associated regions in the feature space, as they are defined by a specific classifier. The classes are modeled by the probability density functions $p(x|\omega_1)$ and $p(x|\omega_2)$, respectively, while the associated prior probabilities are $P(\omega_1)$ and $P(\omega_2)$. Which of the following statements is/are true?

1. The error probability associated with class ω_1 equals to the area under the curve of $P(\omega_1) \cdot p(x|\omega_1)$ within R_2.

2. The error probability associated with class ω_1 equals to the area under the curve of $P(\omega_2) \cdot p(x|\omega_2)$ within R_2.

3. The error probability associated with class ω_1 equals to the area under the curve of $P(\omega_1) \cdot p(x|\omega_1)$ within R_1.

4. The error probability associated with class ω_2 equals to the area under the curve of $P(\omega_2) \cdot p(x|\omega_2)$ within R_1.

Exercise 8 (multiple choices question): Consider a two-class classification problem where the involved classes are denoted by ω_1 and ω_2 and let R_1 and R_2 be their associated regions in the feature space, as they are defined by a specific classifier. The classes are modeled by the probability density functions $p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_2)$, respectively. Also, let $L = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$, where λ_{ij} is the relative weight of the error committed when a data vector $\boldsymbol{x}$ is assigned by a classifier to class ω_j, while it stems from class ω_i, $i, j = 1{,}2$. Assuming that L is anti-diagonal, which of the following expressions is/are valid for the risk r_1, associated with class ω_1?

1. $r_1 = \lambda_{11} \int_{R_2} p(\boldsymbol{x}|\omega_1) d\boldsymbol{x}$

2. $r_1 = \lambda_{12} \int_{R_2} p(\boldsymbol{x}|\omega_1) d\boldsymbol{x}$

3. $r_1 = \lambda_{12} \int_{R_1} p(\boldsymbol{x}|\omega_2) d\boldsymbol{x}$

4. $r_1 = \lambda_{12} \int_{R_1} p(\boldsymbol{x}|\omega_1) d\boldsymbol{x}$

Exercise 9 (multiple choices question): Consider a two-class classification problem where the involved classes are denoted as ω_1 and ω_2 and let R_1 and R_2 be their associated regions in the feature space, as they are defined by a specific classifier. The classes are modeled by the probability density functions (pdfs) $p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_2)$, respectively. Also, let $L = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$, where λ_{ij} is the relative weight of the error committed when a data vector $\boldsymbol{x}$ is assigned by a classifier to class ω_j, while it stems from class ω_i, $i, j = 1{,}2$. Finally, let r be the average risk and P_e be the probability of classification error. Which of the following statements is/are true?

1. When L is anti-diagonal with equal anti-diagonal elements, then $r \equiv P_e$.

2. In the case where $p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_2)$ are both uniform pdfs with disjoint supports, then $r = P_e = 0$.

3. When at least three entries of L are not zero, then $r \equiv P_e$.

4. When all the elements of L are equal to each other, then $r \equiv P_e$

Exercise 10 (multiple choices question): Consider a two-class classification problem where the involved classes are denoted by ω_1 and ω_2. The classes are modeled by the probability density functions (pdfs) $p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_2)$, respectively, while $P(\omega_1)$, $P(\omega_2)$ and $P(\omega_1|\boldsymbol{x})$, $P(\omega_2|\boldsymbol{x})$ denote the prior and the posterior probabilities that are associated with the two classes, respectively. Also, let $L = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$, where λ_{ij} is the relative weight of the error committed when a data vector $\boldsymbol{x}$ is assigned by a classifier to class ω_j, while it stems from class ω_i, $i,j = 1,2$. Finally, let r be the average risk and P_e be the probability of classification error. Which of the following statements is/are true?

1. In general, the Bayes classification rule minimizes r.

2. The classifier that minimizes r, assigns the region $R_1 = \{x \in R^l: \lambda_{12}P(\omega_1|\boldsymbol{x}) > \lambda_{21}P(\omega_2|\boldsymbol{x})\}$ to class ω_1.

3. The classifier that minimizes r, assigns the region $R_2 = \{x \in R^l: \lambda_{12}P(\omega_1)p(\boldsymbol{x}|\omega_1) > \lambda_{21}P(\omega_2)p(\boldsymbol{x}|\omega_2)\}$ to class ω_2.

4. The classification rule that minimizes r is equivalent to the Bayes classification rule when the class priors are defined as $P'(\omega_1) = \frac{\lambda_{12}P(\omega_1)}{\lambda_{12}P(\omega_1)+\lambda_{21}P(\omega_2)}$ and $P'(\omega_2) = \frac{\lambda_{21}P(\omega_2)}{\lambda_{12}P(\omega_1)+\lambda_{21}P(\omega_2)}$.

Exercise 11 (multiple choices question): Consider an M-class classification problem where the involved classes are denoted as $\omega_i, i = 1, \dots, M$. The classes are modeled by the probability density functions (pdfs) $p(\boldsymbol{x}|\omega_i)$, $i = 1, \dots, M$. $P(\omega_i)$ and $P(\omega_i|\boldsymbol{x})$ denote the prior and the posterior probabilities, respectively, associated with class $\omega_i, i = 1, 2, \dots, M$. Also, let $L = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1M} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{M1} & \lambda_{M2} & \cdots & \lambda_{MM} \end{bmatrix}$, where λ_{ij} is the relative weight of the error committed when a data vector $\boldsymbol{x}$ is assigned by a classifier to class ω_j, while it stems from class ω_i, $i,j = 1, \dots, M$. Finally, let r be the average risk and P_e be the probability of classification error. Which of the following statements is/are true?

1. The classifier that minimizes r, assigns the region $R_1 = \left\{x \in R^l: \sum_{k=1}^{M} \lambda_{k1}P(\omega_k)p(\boldsymbol{x}|\omega_k) < \sum_{k=1}^{M} \lambda_{kj}P(\omega_k)p(\boldsymbol{x}|\omega_k), for\ j = 2, \dots, M\right\}$ to class ω_1.

2. The classifier that minimizes r, assigns the region $R_M = \left\{x \in R^l: \sum_{k=1}^{M} \lambda_{kM}P(\omega_k|\boldsymbol{x}) > \sum_{k=1}^{M} \lambda_{kj}P(\omega_k|\boldsymbol{x}), for\ j = 1, \dots, M-1\right\}$ to class ω_M.

3. If all $p(\boldsymbol{x}|\omega_i)$'s are uniform with disjoint supports, S_i, then the class regions defined by the classifier that minimizes r are $R_i = S_i$, $i = 1, \dots, M$.

4. If all $p(\boldsymbol{x}|\omega_i)$'s are uniform with disjoint supports, S_i, then class regions defined by the classifier that minimizes P_e are $R_i = S_i$, $i = 1, \dots, M$.

Exercise 12 (multiple choices question): Consider a two-class one-dimensional classification task where the two classes, ω_1 and ω_2, are equiprobable (they have equal a priori probabilities) and they are modeled by the uniform distributions $p(x|\omega_1) = \begin{cases} 1/5, & x \in (0,5) \\ 0, & otherwise \end{cases}$, and $p(x|\omega_2) = \begin{cases} 1/2, & x \in (4,6) \\ 0, & otherwise \end{cases}$, respectively. Consider the loss matrices $L_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $L_2 = \begin{bmatrix} 0 & 0.75 \\ 0.25 & 0 \end{bmatrix}$ and let r^1 and r^2 be the respective associated average risks (clearly, r^1 coincides with the probability of classification error). The regions ${R_1}^1$ and ${R_2}^1$ that are assigned to each one of the two classes, by the classifier that minimizes r^1, and the regions ${R_1}^2$ and ${R_2}^2$ that are assigned to each of the two classes by the classifier that minimizes r^2, are:

1. ${R_1}^1 = (0,4)$, ${R_2}^1 = (4,6)$ and ${R_1}^2 = (0,4)$, ${R_2}^2 = (4,6)$

2. ${R_1}^1 = (0,5)$, ${R_2}^1 = (5,6)$ and ${R_1}^2 = (0,4)$, ${R_2}^2 = (4,6)$

3. ${R_1}^1 = (0,4)$, ${R_2}^1 = (4,6)$ and ${R_1}^2 = (0,5)$, ${R_2}^2 = (5,6)$

4. ${R_1}^1 = (0,5)$, ${R_2}^1 = (5,6)$ and ${R_1}^2 = (0,5)$, ${R_2}^2 = (5,6)$

Exercise 13 (multiple choices question): Consider a two-class one-dimensional classification task, where the involved classes ω_1 and ω_2 are equiprobable (they have equal a priori probabilities) and they are modeled by the distributions $p(x|\omega_1)$ and $p(x|\omega_2)$, respectively. In the following figures, the regions assigned to each one of the classes by the classifier that minimizes the average risk, in each case, are shown. Which of the following statements is/are true?

(a) (b)

1. In case (a) the class ω_1 is more sensitive than class ω_2.

2. In case (b) the class ω_2 is less sensitive than class ω_1.

3. In case (a) both classes are equally sensitive.

4. In case (b) both classes are equally sensitive.

Exercise 14 (**mult. choice question**)**:** Which of the following statements regarding a classification task is/are true?

1. Each classifier associates a region of the feature space with a class.

2. If the feature space is the R^3, the class regions are delimited from each other by lines or curves.

3. A decision surface leaves all the points of the feature space on its positive side.

4. If $g: R^l \to R$ is the function associated with a (hyper)surface S, then, if for a specific $\boldsymbol{x} \in R^l$ it is $g(\boldsymbol{x}) = 0$, then $\boldsymbol{x} \in S$.

Exercise 15 (**mult. choice question**)**:** Consider a two-class l-dimensional classification task, where classes ω_1 and ω_2 are involved and let S be the decision hypersurface defined by a specific classifier, which is modeled by the equation $g(\boldsymbol{x}) = 0$. Let also R_1 and R_2 be the class regions associated with ω_1 and ω_2, respectively. Which of the following statements is/are true?

1. For all points in R_1 (R_2) it is $g(\boldsymbol{x}) > (<) 0$ or for all points in R_2 (R_1) it is $g(\boldsymbol{x}) > (<) 0$.

2. In the case of the Bayesian classifier, it is $g(\boldsymbol{x}) = P(\omega_1) - P(\omega_2) = 0$.

3. There are cases where some points of R_1 may lie on the positive side of S and some others on the negative side of S.

4. In the case of the Bayesian classifier, it is $g(\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\omega_1)P(\omega_1)}{p(\boldsymbol{x})} - \frac{p(\boldsymbol{x}|\omega_2)P(\omega_2)}{p(\boldsymbol{x})} = 0$.

5. For all points in ω_1 (ω_2) it is $g(\boldsymbol{x}) > (<) 0$ or for all points in ω_2 (ω_1) it is $g(\boldsymbol{x}) > (<) 0$.

Exercise 16 (**mult. choice question**)**:** Consider the l-dimensional Gaussian distribution with mean $\boldsymbol{\mu} = [\mu_1, \dots, \mu_l]^T$ and covariance matrix $\Sigma = \begin{bmatrix} {\sigma_1}^2 & \sigma_{12} & \cdots & \sigma_{1l} \\ \sigma_{12} & {\sigma_2}^2 & \cdots & \sigma_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1l} & \sigma_{2l} & \cdots & {\sigma_l}^2 \end{bmatrix}$, described by the equation

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right)$$

where $\boldsymbol{x} = [x_1, \dots, x_l]^T$. Which of the following statements is/are true?

1. The diagonal elements of Σ are the variances along each dimension, x_i, of $\boldsymbol{x}$.

2. The non-diagonal elements σ_{ij} of Σ are the covariances between x_i and x_j.

3. If Σ is diagonal, then the random variables x_i's are dependent on each other.

4. In the case where Σ is non-diagonal, $p(\boldsymbol{x})$ is centered at $\Sigma^{-1}\boldsymbol{\mu}$.

Exercise 17 (**mult. choice question**)**:** Consider a two-class l-dimensional classification task, where the probability density functions of the involved classes ω_1 and ω_2, $p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_2)$, are normal distributions with mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and covariance matrices Σ_1 and Σ_2, respectively. Then, the decision hypersurfaces defined by the Bayes classifier are described, in general, by

1. Second degree equations

2. First degree equations

3. Third degree equations

4. A linear combination of exponential functions.

Exercise 18 (**mult. choice question**)**:** Consider a two-class l-dimensional classification task, where the probability density functions of the involved classes ω_1 and ω_2, $p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_2)$, are normal distributions with mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and covariance matrices Σ_1 and Σ_2, respectively. Also, let $P(\omega_1)$ and $P(\omega_2)$ be the a priori class probabilities for ω_1 and ω_2, respectively. Consider the decision hypersurface associated with the Bayes classifier for the problem under study

$$g(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{x}^T \Sigma_2^{-1} \boldsymbol{x} - \boldsymbol{x}^T \Sigma_1^{-1} \boldsymbol{x}\right) + \boldsymbol{\mu}_1^T \Sigma_1^{-1} \boldsymbol{x} - \boldsymbol{\mu}_2^T \Sigma_2^{-1} \boldsymbol{x} - \frac{1}{2} \boldsymbol{\mu}_1^T \Sigma_1^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} \boldsymbol{\mu}_2^T \Sigma_2^{-1} \boldsymbol{\mu}_2$$
$$+ \ln \frac{P(\omega_1)}{P(\omega_2)} + \frac{1}{2} \ln \frac{|\Sigma_2|}{|\Sigma_1|} = 0$$

Which of the following statements is/are true?

1. If $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, then the equation $g(\boldsymbol{x}) = 0$ is of linear nature.

2. The equation $g(\boldsymbol{x}) = 0$ is of quadratic nature for $\Sigma_1 \neq \Sigma_2$.

3. If $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and $\Sigma_1 = \Sigma_2$, then the expression $g(\boldsymbol{x})$ has a constant value.

4. If $P(\omega_1) = P(\omega_2)$, then the equation $g(\boldsymbol{x}) = 0$ is necessarily of linear nature.

Exercise 19 (**mult. choice question**)**:** Consider a two-class one-dimensional classification task, where the probability density functions of the involved classes ω_1 and ω_2, $p(x|\omega_1)$ and $p(x|\omega_2)$, are normal distributions with mean values μ_1 and μ_2 and variances $\sigma_1{}^2$ and $\sigma_2{}^2$, respectively. Also, let $P(\omega_1)$ and $P(\omega_2)$ be the a priori class probabilities for ω_1 and ω_2, respectively. In this case, the decision hypersurface associated with_the Bayes classifier breaks down to a set of single points, which are the roots of the following equation,

$$g(x) = \frac{1}{2}\left(\frac{1}{\sigma_2{}^2} - \frac{1}{\sigma_1{}^2}\right) x^2 + \left(\frac{\mu_1}{\sigma_1{}^2} - \frac{\mu_2}{\sigma_2{}^2}\right) x - \frac{1}{2}\frac{\mu_1{}^2}{\sigma_1{}^2} + \frac{1}{2}\frac{\mu_2{}^2}{\sigma_2{}^2} + \ln \frac{P(\omega_1)}{P(\omega_2)} + \frac{1}{2} \ln \frac{\sigma_2{}^2}{\sigma_1{}^2} = 0$$

Which of the following statements is/are true?

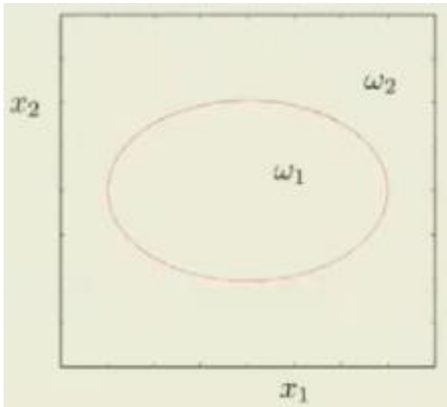
1. For $\sigma_1{}^2 \neq \sigma_2{}^2$ and $P(\omega_1) = P(\omega_2)$, the equation $g(x) = 0$ has a single solution.

2. For $\mu_1 = \mu_2$, the equation $g(x) = 0$ has always a single solution.

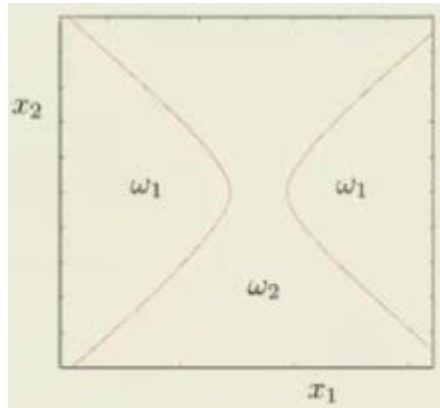
3. For $\sigma_1{}^2 = \sigma_2{}^2$ and $\mu_1 \neq \mu_2$, the decision regions that are associated with the two classes are two disjoint half-lines, whose union is the whole feature space.

4. For $\sigma_1{}^2 = \sigma_2{}^2$ and $P(\omega_1) = P(\omega_2)$, the solution of the equation $g(x) = 0$ is $x = \frac{\mu_1 + \mu_2}{2}$.

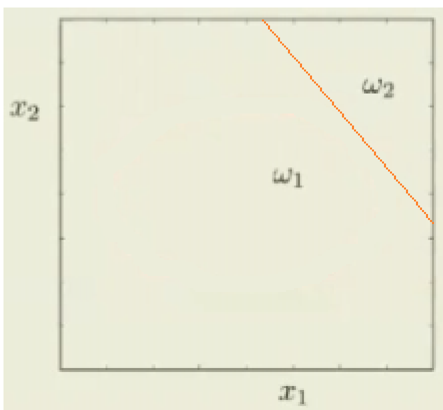
Exercise 20 (**mult. choice question**)**:** Consider a two-class two-dimensional classification task, where the probability density functions of the involved classes ω_1 and ω_2, are denoted as $p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_2)$, respectively. Also, let $P(\omega_1)$ and $P(\omega_2)$ be the a priori class probabilities for ω_1 and ω_2, respectively. In which of the following cases, the probability density functions that model the classes are definitely **not** normal distributions?



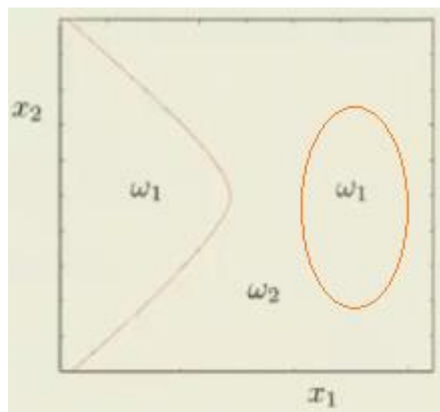
(1)



(2)

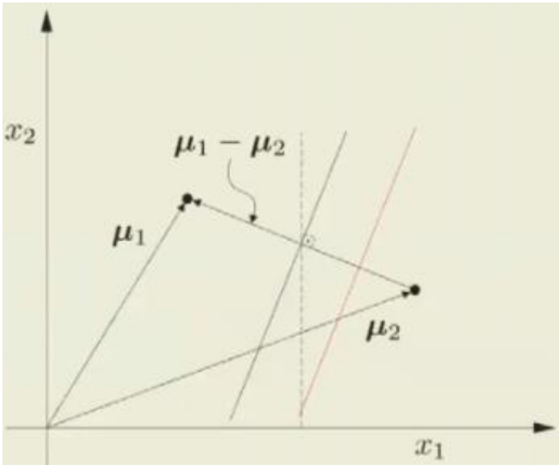


(3)



(4)

Exercise 21 (mult. choice question): Consider a two-class two-dimensional classification task, where the probability density functions of the involved classes ω_1 and ω_2, $p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_2)$, are normal distributions with mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and common diagonal covariance matrix Σ. Also, let $P(\omega_1)$ and $P(\omega_2)$ be the a priori class probabilities for ω_1 and ω_2, respectively. In the figure below, the vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ as well as the vector of their difference $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ are shown. The full red line, full gray line and dashed line in the figure that intersect the line segment $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, are different borders between the two classes, that correspond to different scenarios concerning $P(\omega_1)$, $P(\omega_2)$ and Σ. Which of the following pairs "line $\longleftrightarrow$ scenario" are compatible to each other (note that the full gray and the dotted gray lines pass through the middle point of the line segment $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, while the full gray and the full red lines are perpendicular to that)?



1. full red line $\leftrightarrow P(\omega_1) = P(\omega_2), \Sigma \neq \sigma^2 I$

2. full red line $\leftrightarrow P(\omega_1) \neq P(\omega_2), \Sigma = \sigma^2 I$

3. full gray line $\leftrightarrow P(\omega_1) = P(\omega_2), \Sigma = \sigma^2 I$

4. full gray line $\leftrightarrow P(\omega_1) \neq P(\omega_2), \Sigma = \sigma^2 I$

5. dashed line $\leftrightarrow P(\omega_1) = P(\omega_2), \Sigma = \sigma^2 I$

6. dashed line $\leftrightarrow P(\omega_1) = P(\omega_2), \Sigma \neq \sigma^2 I$

Exercise 22 (**mult. choice question**)**:** Consider the minimum Euclidean distance (MED) and the minimum Mahalanobis distance (MMD) classifiers. Consider also an M-class classification problem, where the a priori class probabilities $P(\omega_j), j = 1, \dots, M$, are equal to each other and let $p(\boldsymbol{x}|\omega_j)$ denotes the probability density functions (pdf) associated with the j–th class, $j = 1, \dots, M$. Which of the following statements is/are true?

1. The MED classifier is equivalent to the Bayes classifier, if $p(\boldsymbol{x}|\omega_j) = N(\boldsymbol{\mu}_j, \Sigma_j), j = 1, \dots, M$.

2. The MED classifier is equivalent to the Bayes classifier, if $p(\boldsymbol{x}|\omega_j) = N(\boldsymbol{\mu}_j, \Sigma)\ , j = 1, \dots, M$.

3. The MED classifier is equivalent to the Bayes classifier, if $p(\boldsymbol{x}|\omega_j) = N(\boldsymbol{\mu}_j, \sigma_j^2 I)\ , j = 1, \dots, M$.

4. The MED classifier is equivalent to the Bayes classifier, if $p(\boldsymbol{x}|\omega_j) = N(\boldsymbol{\mu}_j, \sigma^2 I)\ , j = 1, \dots, M$.

5. The MMD classifier is equivalent to the Bayes classifier, if $p(\boldsymbol{x}|\omega_j) = N(\boldsymbol{\mu}_j, \sigma_j^2 I)\ , j = 1, \dots, M$.

6. The MMD classifier is equivalent to the Bayes classifier, if $p(\boldsymbol{x}|\omega_j) = N(\boldsymbol{\mu}_j, \Sigma)\ , j = 1, \dots, M$.

7. The MMD classifier is equivalent to the Bayes classifier, if $p(\boldsymbol{x}|\omega_j) = N(\boldsymbol{\mu}_j, \sigma^2 I)\ , j = 1, \dots, M$.

Hint: $N(\boldsymbol{\mu}, \Sigma)$ denotes the normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. I is the identity matrix and $\sigma^2 I$ is a diagonal matrix with all its diagonal entries equal to σ^2.

Exercise 23 (**mult. choice question):** Consider a two-class classification task where the two classes are denoted by ω_1 and ω_2. Which of the following statements is/are true?

1. If the data points of ω_1 and ω_2 stem from different normal distributions, then the minimum Euclidean distance classifier for this problem minimizes the probability of classification error.

2. If the a priori probabilities of ω_1 and ω_2 are not equal, then the Euclidean minimum distance classifier for this problem is definitely not equivalent to the Bayes classifier.

3. Assuming that ω_1 and ω_2 are equiprobable and they are modeled by normal distributions whose covariance matrices are different from each other, the Minimum Mahalanobis distance classifier is equivalent to the Bayes classifier.

4. Assuming that ω_1 and ω_2 are equiprobable and they are modeled by normal distributions with common covariance matrix, the Minimum Euclidean distance classifier is equivalent to the Bayes classifier.

Exercise 24 (**mult. choice question):** Consider a two-class two-dimensional classification task where the involved classes, ω_1 and ω_2, are equiprobable and they are modeled by normal probability density functions, with mean vectors $\boldsymbol{\mu}_1 = [0, 0]^T$ and $\boldsymbol{\mu}_2 = [3, 3]^T$ and common covariance matrix $\Sigma = \begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}$. The minimum Euclidean and minimum Mahalanobis distance classifiers are utilized for solving this classification task and let $\boldsymbol{x}_1 = [1.4, 1.7]^T$, $\boldsymbol{x}_2 = [1, 2.2]^T$, $\boldsymbol{x}_3 = [1, 1]^T$ be the data points that are to be classified. Which of the following statements is/are true?

1. The minimum distance Euclidean classifier assigns $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ to the classes $\omega_1, \omega_2, \omega_1$, respectively.

2. The minimum distance Mahalanobis classifier assigns $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ to the classes $\omega_2, \omega_1, \omega_1$, respectively (the assignment of $\boldsymbol{x}_1$ to ω_2 is arbitrary).

3. The two classifiers do not assign all three points to the same class.

4. None of the data points under consideration remains unclassified from any of the two classifiers.

Exercise 25 (**mult. choice question**)**:** Consider a two-class two-dimensional classification task where the involved classes, ω_1 and ω_2, are equiprobable and they are modeled by normal probability density functions, with mean vectors $\boldsymbol{\mu}_1 = [0, 0]^T$ and $\boldsymbol{\mu}_2 = [3, 3]^T$ and common covariance matrix $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. The minimum Euclidean and minimum Mahalanobis distance classifiers are utilized for solving this classification task and let $\boldsymbol{x}_1 = [1.4, 1.7]^T$, $\boldsymbol{x}_2 = [1, 2.2]^T$, $\boldsymbol{x}_3 = [1, 1]^T$ be the data points that are to be classified. Which of the following statements is/are true?

1. The minimum distance Euclidean classifier assigns $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$ to the classes $\omega_2, \omega_2, \omega_1$, respectively.

2. The minimum distance Mahalanobis classifier assigns $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$ to the classes $\omega_2, \omega_1, \omega_1$, respectively.

3. The two classifiers do not assign all three points to the same class.

4. None of the data points under consideration remains unclassified from any of the two classifiers.

Exercise 26 (**mult. choice question**)**:** Consider a two-class two-dimensional classification task where the involved classes, ω_1 and ω_2, are equiprobable and they are modeled by normal probability density functions, with mean vectors $\boldsymbol{\mu}_1 = [0, 0]^T$ and $\boldsymbol{\mu}_2 = [3, 3]^T$ and common covariance matrix $\Sigma = \begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}$. The minimum Euclidean and minimum Mahalanobis distance classifiers are utilized for solving this classification task and let $\boldsymbol{x} = [1, 2.2]^T$ be a data point that needs to be classified. It turns out that the minimum Mahalanobis distance classifier assigns $\boldsymbol{x}$ to class ω_1, while the minimum Euclidean distance classifier assigns $\boldsymbol{x}$ to class ω_2. Which of the following statements is/are true?

1. Since, under certain conditions, both classifiers are optimal w.r.t the probability of error, we can decide arbitrarily for $\boldsymbol{x}$.

2. The classification of the minimum Mahalanobis distance classifier is more preferable, since for this case this classifier is optimum w.r.t the probability of classification error.

3. The two classifiers perform exactly the same partitioning of the feature space.

4. For any data point $\boldsymbol{x}' \in R^2$, the two classifiers give different from each other classification results.

Exercise 27 (**mult. choice question**)**:** Which of the following statements regarding the (optimal) Bayes classifier and other suboptimal classifiers is/are true?

1. In the unrealistic case where the number of the available data points approaches infinity, we should definitely resort to suboptimal classifiers.

2. If the probability density functions (pdfs) of all classes in a specific classification task are known, then the Bayes classifier is preferable.

3. If the size of the available data set is limited and fixed, then adoption of classifiers that involve only a few parameters to be learned is preferred.

4. In practice, where the number of the available data points is fixed and finite, one should carefully consider classifiers that are described by as few parameters as possible, in order to avoid overfitting.

5. In the case where the classes that are involved in the classification task under study are modeled by normal pdfs of fixed covariance matrices, the Bayes and the naïve Bayes classifiers are equivalent.

Exercise 28 (**mult. choice question):** Consider a three-class four-dimensional classification task. Assuming that the classes are modelled by normal probability density functions, the total number of parameters that are involved in the Bayes and the naïve Bayes classifiers, respectively, are

1. $24, 42$

2. $42, 24$

3. $12, 12$

4. $14, 8$

Exercise 29 (**mult. choice question):** Which of the following statements regarding the Bayes and the naïve Bayes classifiers is/are true?

1. There are certain cases where the naïve Bayes classifier achieves better classification error than the theoretically optimal Bayes classifier.

2. If the features that are involved in the representation of the patterns in a classification task are statistically independent, the naïve Bayes classifier is equivalent to the optimal Bayes classifier.

3. In the naïve Bayes classifier the (multidimensional) probability density function of each class is written as a product of one-dimensional probability density functions.

4. In one-dimensional classification tasks, the naïve Bayes and the Bayes classifiers are not equivalent.

5. For small size data sets, it is more preferable to have suboptimal classifiers (e.g., naïve Bayes classifier), whose parameters are more reliably estimated, than the optimal Bayes classifier whose parameters are poorly estimated.

Exercise 30 (**mult. choice question**)**:** Consider a two-class two-dimensional classification task, where the involved classes, ω_1 and ω_2, are modeled by normal distributions, with mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and covariance matrices Σ_1 and Σ_2, respectively, i.e. $p(\boldsymbol{x}|\omega_1) = N(\boldsymbol{\mu}_1, \Sigma_1)$ and $p(\boldsymbol{x}|\omega_2) = N(\boldsymbol{\mu}_2, \Sigma_2)$. Assume that we have at our disposal a data set of ten data points, namely, $\boldsymbol{x}_1 = [0.1,\ 2.5]^T, \boldsymbol{x}_2 = [0.2,\ 2]^T, \boldsymbol{x}_3 = [0.3,\ 1.5]^T, \boldsymbol{x}_4 = [0.4,\ 1]^T, \boldsymbol{x}_5 = [0.5,\ 0.5]^T, \boldsymbol{x}_6 = [2.1,\ 5]^T, \boldsymbol{x}_7 = [2.2,\ 4.5]^T, \boldsymbol{x}_8 = [2.3,\ 4]^T, \boldsymbol{x}_9 = [2.4,\ 3.5]^T, \boldsymbol{x}_{10} = [2.5,\ 3]^T$. The first five of these points stem from class ω_1, while the remaining five stem from class ω_2. Based on the above data set, estimates of the class a priori probabilities $P(\omega_1)$ and $P(\omega_2)$, as well as the maximum likelihood estimates of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and Σ_1, Σ_2, associated with the **naïve** Bayes classifier are

1. $P(\omega_1) = 0.5$, $\boldsymbol{\mu}_1 = [1.5,\ 0.3]^T, \Sigma_1 = \frac{1}{5}\begin{bmatrix} 2.5 & 0 \\ 0 & 0.1 \end{bmatrix}$ and $P(\omega_2) = 0.5$, $\boldsymbol{\mu}_2 = [2.3,\ 4]^T, \Sigma_2 = \frac{1}{5}\begin{bmatrix} 0.1 & 0 \\ 0 & 2.5 \end{bmatrix}$

2. $P(\omega_1) = 0.5$, $\boldsymbol{\mu}_1 = [0.3,\ 1.5]^T, \Sigma_1 = \begin{bmatrix} 0.1 & 2.5 \\ 2.5 & 0.1 \end{bmatrix}$ and $P(\omega_2) = 0.5$, $\boldsymbol{\mu}_2 = [2.3,\ 4]^T, \Sigma_2 = \begin{bmatrix} 0.1 & 2.5 \\ 2.5 & 0.1 \end{bmatrix}$

3. $P(\omega_1) = 0.5$, $\boldsymbol{\mu}_1 = [0.3,\ 1.5]^T, \Sigma_1 = \frac{1}{5}\begin{bmatrix} 0.1 & 0 \\ 0 & 2.5 \end{bmatrix}$ and $P(\omega_2) = 0.5$, $\boldsymbol{\mu}_2 = [2.3,\ 4]^T, \Sigma_2 = \frac{1}{5}\begin{bmatrix} 0.1 & 0 \\ 0 & 2.5 \end{bmatrix}$

4. $P(\omega_1) = 0.5$, $\boldsymbol{\mu}_1 = [1.5, 0.3]^T, \Sigma_1 = \begin{bmatrix} 2.5 & 0 \\ 0 & 0.1 \end{bmatrix}$ and $P(\omega_2) = 0.5$, $\boldsymbol{\mu}_2 = [4, 2.3]^T, \Sigma_2 = \begin{bmatrix} 2.5 & 0 \\ 0 & 0.1 \end{bmatrix}$

Hint: The maximum likelihood estimate of the mean of a set of N real numbers is $\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$, while the maximum likelihood estimate of the variance is $\sigma_{ML}{}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2$.

Exercise 31: Consider a two-class 2-dim. classification problem of two equiprobable classes ω_1 and ω_2 that are modeled by the normal distributions $N(\boldsymbol{\mu}_1, \Sigma)$ and $N(\boldsymbol{\mu}_2, \Sigma)$, where $\Sigma = \sigma^2 I$.

(a) Show that the Bayesian classifier borders the decision regions R_1 and R_2 (corresponding to ω_1 and ω_2, respectively) by the perpendicular bisector of the line segment whose endpoints are $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$.
(b) What would be the border in the case where $\Sigma \neq \sigma^2 I$? (give intuitive arguments).

Hint: The equation that describes the perpendicular bisector of a line segment whose endpoints are $\boldsymbol{\mu}_1 = [\mu_{11}, \mu_{12}]^T$ and $\boldsymbol{\mu}_2 = [\mu_{21}, \mu_{22}]^T$, is $||\boldsymbol{x} - \boldsymbol{\mu}_2||^2 = ||\boldsymbol{x} - \boldsymbol{\mu}_1||^2$ or $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{x} - \frac{1}{2}||\boldsymbol{\mu}_1||^2 + \frac{1}{2}||\boldsymbol{\mu}_2||^2 = 0$, where $\boldsymbol{x} = [x_1, x_2]^T$.

Exercise 32 (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.

Recall that $\boldsymbol{x} = [x_1, x_2]^T$

(a) Adopt the **naïve Bayes classifier**.
  i. Use the training set to estimate $P(\omega_1), P(\omega_2), p(x_1|\omega_1), p(x_2|\omega_1), p(x_1|\omega_2), p(x_2|\omega_2)$ (Each $p(\boldsymbol{x}|\omega_j)$ is written as $p(\boldsymbol{x}|\omega_j) = p(x_1|\omega_j) \cdot p(x_2|\omega_j)$. Use the ***ML* estimates** of the mean and variance for each one of the 1-dim. pdfs).
  ii. Classify the points $\boldsymbol{x}_i = [x_{i1}, x_{i2}]^T$ of the test set, using the naïve Bayes classifier (Estimate $p(\boldsymbol{x}|\omega_j)$ with $p(x_{i1}|\omega_j) \cdot p(x_{i2}|\omega_j)$ and then apply the Bayes rule. Keep the class labels, to an a N_{test}–dim. column **vector** , called ***NBtest_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 ***NBtest_y*** , (2) **count** the positions where both of them have the same class label and (3) **divide** with the total number of test vectors).

(b) Depict graphically the training set, using different colors for points from different classes.

(c) Report the classification results obtained by (a) the Bayes classifier (see exercise 32 of Homework 7) and (b) the naïve Bayes classifier and comment on them. Under what conditions, the Bayes and the naïve Bayes classifiers would exhibit the same performance?

Hint: Use the attached Python code in file *HW8.ipynb* (also given in Homework 7).