ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

SCHOOL OF INFORMATION SCIENCES AND TECHNOLOGY

DEPARTMENT OF INFORMATICS

MASTER OF SCIENCE IN DATA SCIENCE

SPRING QUARTER 2020 ***31/3/2020***

Examination of the course

"Machine Learning and Computational Statistics"

1. **(a) (***grade* **1.2**/10**)** Consider an one-dimensional three-class classification problem where the three involved classes ω_1, ω_2 and ω_3 are equiprobable and are modeled by the following pdfs

$$p(x|\omega_1) = \begin{cases} 1/3, & x \in (3,4) \cup (6,8) \\ 0, & \text{otherwise} \end{cases}, \; p(x|\omega_2) = \begin{cases} 1/2, & x \in (5,7) \\ 0, & \text{otherwise} \end{cases}, \; p(x|\omega_3) = \begin{cases} 1/8, & x \in (0,8) \\ 0, & \text{otherwise} \end{cases}$$

   (i) Give the common plot of $P(\omega_i)p(x|\omega_i)$ versus x, for both classes and determine the decision regions for each class, as they result from the application of the Bayes rule (give a short explanation).
   (ii) Compute the probability of error classification for the Bayes classifier.
   (iii) Classify the points $x_1 = 2$, $x_2 = 4.5$, $x_3 = 5.5$ to one of the two classes.

   Hint: The probability of error for *M*-class classification problems is

$$P_{error} = \sum_{j=1}^{M} \int_{R_j} \left(\sum_{k=1, k \neq j}^{M} p(\mathbf{x} \mid \omega_k) P(\omega_k) \right) dx$$

   (b) (*grade* **1.8**/10**)** (i) The Rayleigh distribution is defined as follows

$$p(x) = 2\theta x e^{-\theta x^2} u(x)$$

   where $u(x) = 1(0)$, for $x \geq (<) 0$. Given a set of independent observations $X=\{x_1,...,x_N\}$ that stem from a one-dimensional Rayleigh distribution, prove that the maximum likelihood estimation of the parameter θ, with respect to X, is given by the relation

$$\theta_{ML} = \frac{N}{\sum_{i=1}^{N} {x_i}^2}$$

   (ii) Consider a two-class one-dimensional classification problem where the two classes ω_1 and ω_2 are modeled by Rayleigh distributions. Assuming that the sets of observations $X_1 = \{0, 0.5, 1, 1.5, 2\}$, and $X_2 = \{3, 3.5, 4, 4.5, 5\}$ stem from the classes ω_1 and ω_2, respectively:
   I. Determine the ML estimates of the parameter θ for the two classes and write down explicitly the pdfs for the two classes.
   II. Estimate the class a-priori probabilities and then determine the decision regions for each class, according to the Bayes classifier.

   Hint: (i) Consider the log-likelihood function $L(\theta) = \sum_{i=1}^{N} \ln p(x_i; \theta)$. (ii) It is

x	0	0.5	1	1.5	2	3	3.5	4	4.5	5
x^2	0	0.25	1	2.25	4	9	12.25	16	20.25	25

   (ii)-II Consider the points where the quantities $P(\omega_1)p(x|\omega_1)$ and $P(\omega_2)p(x|\omega_2)$ are equal.

2. **(a)** Consider the two-dim. two-class problem where the vectors $[1,0]^T$, $[0,-3]^T$ belong to class 1, while the vectors $[-1,0]^T$ and $[0,3]^T$, belong to class -1.
   (i) **(***grade* **1.5**/10**)** Draw the above vectors on your paper (design the figure carefully). Determine and plot the linear classifier that separates the (points of the) two classes and results from the minimization of the sum of error squares criterion.
   (ii) (grade **0.35**/10) Classify the vectors $[5,1]^T$ and $[1,5]^T$ with the above classifier.
   (iii) **(***grade* **0.65**/10**)** In your opinion, which is the linear classifier that would result from the SVM methodology? Explain very briefly.

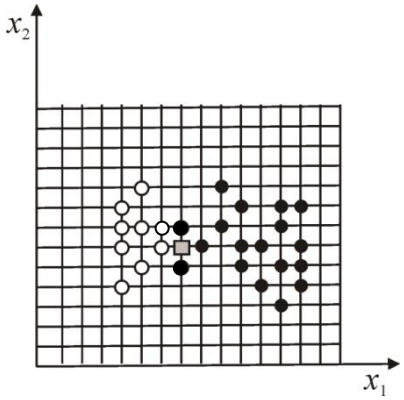
   (b) (*grade* **0.5**/10**)** Consider the following non-linearly separable one-dimensional classification problem where the points -3 and 2 belong to class 1, while -1 and 0 belong to class -1. Draw the points on the paper and propose a simple (one-dimensional) transformation $x \dashrightarrow \varphi(x)$ that makes the problem linearly separable.

Hints: **(a)** *(i)* If $A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_k \end{bmatrix}$ then $A^{-1} = \begin{bmatrix} 1/a_1 & 0 & \cdots & 0 \\ 0 & 1/a_2 & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/a_k \end{bmatrix}$

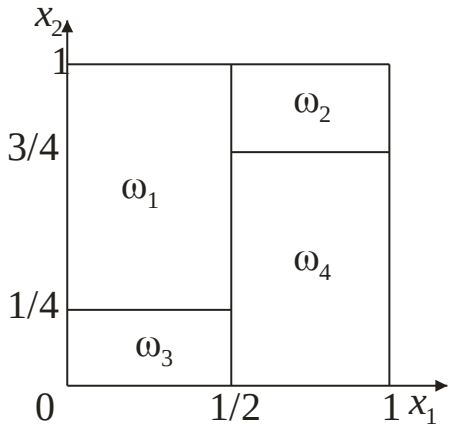
(b) In the 1-dim. case, linear separability means that there exists a point on the (feature) axis, that leaves all data points of the one class on its one side and all the others from its other side.

3. **(a) (***grade* **1.25**/10**)** Let (e1) $x_1 + 2x_2 = 0$ and (e2) $x_1 - 2x_2 = 0$ be two lines with the point (2,0) lying on their positive side. Consider the two-dimensional two-class classification problem where all points that lie on the positive or the negative side of both lines belong to class 0 while all the rest points of the plane belong to class 1. Design a neural network that implements the above classification, where each node is modeled by the function $y = f(\boldsymbol{w}^T\boldsymbol{x} + w_0)$, where $f(z) = 1$, for $z > 0$ and $f(z) = 0$, otherwise (describe briefly the steps followed for the network design).

(b) (*grade* **1.25**/10**)** Consider a two-class two-dimensional classification problem of two equiprobable classes ω_1 and ω_2, where the data of ω_1 are denoted by «•», while those of ω_2 are denoted by «°» (see figure 1). Classify the gray colored point, denoted by «□» using ***(i)*** the **5-nearest neighbor rule** and ***(ii)*** **Bayes classifier**, estimating the pdfs using the **3**-nearest neighbor density estimation method. The size of the edge of the grid is equal to 1.



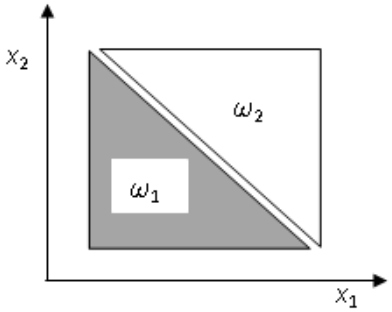
4. **(a) (***grade* **0.8**/10**)** Consider the partition of the unit square shown in the following figure. Determine a decision tree that implements this partition.



(b) (*grade* **0.2**/10**)** Consider a (scalar) parameter θ whose true value is θ_0. Let θ be an unbiased estimator of θ. Consider the estimator $\theta' = (a^2 - 3) \cdot \theta$. Under what conditions θ' is an undiased estimator of θ?
(*Hint:* For an unbiased estimator: $E[\theta] = \theta_0$).

(c) (*grade* **0.2**/10**)** Consider the quadratic regressor $\boldsymbol{x} = [x_1, x_2]^T \in R^2 \mapsto y = f(\boldsymbol{x}) \in R$, with $f(\boldsymbol{x}) = {x_1}^2 + 4x_2 + 6$. Estimate the values y_1, y_2 associated with $\boldsymbol{x}_1 = (2,3)$, $\boldsymbol{x}_2 = (-1,1)$.

(d) (*grade* **0.2**/10**)** Consider a two-class two-dimensional classification problem, as it is depicted in the following figure. Which of the following combinations of features x_1 and x_2 can discriminate well the classes: (α) only x_1, (β) only x_2 and (γ) both x_1 and x_2? Explain briefly.



(e) (*grade* **0.2**/10) Explain very briefly, why the dimensionality reduction through the Principal Component Analysis (PCA) method, does not necessarily retain class separability.

(f) (*grade* **0.9**/10) Consider the data points $\boldsymbol{x}_1 = [1,0]^T, \boldsymbol{x}_2 = [3,0]^T, \boldsymbol{x}_3 = [1,10]^T, \boldsymbol{x}_4 = [3,10]^T$. Run the k-means clustering algorithm for two representatives $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, which are initialized at $\boldsymbol{\theta}_1(0) = [4,10]^T$ and $\boldsymbol{\theta}_2(0) = [4,0]^T$, respectively.