香 港 大 學

THE UNIVERSITY OF HONG KONG

Bachelor of Engineering

Department of Electrical & Electronic Engineering

ELEC3249 Pattern Recognition and Machine Intelligence

2019-2020 Semester 2
Online Examination

Date: __20 May 2020______ Time: ___7:00pm-9:00pm_________

Answer **ALL** questions.
Use a different page for each question.
*Open book exam. You can refer to printed/written/electronic materials, but you cannot search information from the Internet.

Use of Electronic Calculators:

"Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script."

1.

a) Let the likelihood of the two classes ω_1 and ω_2 with respect to x be

$$p(x \mid \omega_1) = \frac{1}{4\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-3}{4}\right)^2} \text{ and } p(x \mid \omega_2) = \frac{1}{4\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x+2}{4}\right)^2}.$$

The *a priori* probabilities for the two classes are given by $P(\omega_1) = 0.7$ and $P(\omega_2) = 0.3$, respectively.

(Sub-total: 16)

i) Explain the difference between one vs all and one vs one classifiers for multi-categorial classification.

(2 marks)

ii) Mention **ONE** difference between each of the following classifiers: 1) Maximum Likelihood, 2) Maximum A Posteriori and 3) Bayes Minimum Risks Classifiers.

(3 marks)

iii) Find the Maximum Likelihood Classifier.

(4 marks)

iv) Using the Bayes rule $P(\omega_i \mid x) = \frac{p(x \mid \omega_i)P(\omega_i)}{p(x)}$, find the Maximum A Posteriori Classifier.

(3 marks)

Client X wants to apply the above classifier for bio-medical applications and has suggested the following loss function for Bayes classification:

is actually ω_1	is actually ω_2	
0	5	choosing ω_1
1	0	choosing ω_2

v) Write down the 4 different values of the loss function $\lambda(\alpha_1 \mid \omega_1)$, $\lambda(\alpha_1 \mid \omega_2)$, $\lambda(\alpha_2 \mid \omega_1)$ and $\lambda(\alpha_2 \mid \omega_2)$, where α_1 and α_2 are the actions of choosing classes ω_1 and ω_2 respectively.

(1 mark)

vi) Find the Bayes Minimum Risk Classifier using the loss function in v).

(3 marks)

b) Consider the following criterion function for finding a hyperplane to separate the two classes of samples, which are $\boldsymbol{x}_1 = [3,1]^T$, $\boldsymbol{x}_2 = [1,4]^T$ (Class 1) and $\boldsymbol{x}_3 = [13,9]^T$, $\boldsymbol{x}_4 = [7,7]^T$ (Class 2), respectively,

$$J_q(\boldsymbol{a}) = \sum_{y_i \in Y_C} -\boldsymbol{a}^T \widetilde{\boldsymbol{y}}_i \,,$$

where Y_C is the set of correctly identified samples. $\boldsymbol{a}$ are the coefficients of the Perceptron.

(Sub-total: 17)

i) What does the vector $\boldsymbol{a}$ geometrically represent?

(2 marks)

ii) $\widetilde{\boldsymbol{y}}$ is referred to as a normalized sample. Describe how a normalized sample can be obtained for the following two cases:

1.) $\boldsymbol{y}$ belongs to class 1 and 2.) $\boldsymbol{y}$ belongs to class 2.

(1 mark)

iii) The Gradient Descent can be used to solve $J_q(\boldsymbol{a})$. Write down the expression in terms of $\rho^{(k)}$, $\nabla_{\boldsymbol{a}} J_q(\boldsymbol{a})$, $\boldsymbol{a}^{(k+1)}$ and $\boldsymbol{a}^{(k)}$ that solves $\boldsymbol{a}$ iteratively.

(1 mark)

iv) Suppose the augmented feature vector is defined as $\boldsymbol{y} = [1, x_1, x_2]^T$. Using ii) and iii), find $\boldsymbol{a}^{(2)}$ and $\boldsymbol{a}^{(3)}$ with an initialization $\boldsymbol{a}^{(1)} = [0,0,0]^T$ and a step size $\rho^{(k)} = 1$.

(6 marks)

v) Write down the objective function of the soft-margin SVM. **Indicate** which part of the objective function **1) maximizes the margin** and **2) penalizes the mis-classified samples**, and **explain** how these functionalities can be achieved.

(7 marks)

2.

a) Let the likelihood of a parameter θ of the density function be

$$p(x \mid \theta) = \frac{1}{4\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\theta}{4}\right)^2}.$$

(Sub-total: 14)

i) Given a set of independent feature samples $\{x_1, x_2\} = \{-1, 7\}$, determine the Maximum Likelihood Estimate (MLE) of θ.

(3 marks)

Assume that the parameter θ has an a priori probability

$$p(\theta) = 0.5[\delta(\theta-2) + \delta(\theta-5)],$$

where $\delta(.)$ is the ideal unit impulse function informally defined as:

$$\delta(y) = \begin{cases} \infty & y = 0 \\ 0 & otherwise \end{cases} \text{ and } \int_{-\infty}^{\infty} \delta(y) = 1.$$

ii) Determine the posterior probability $p(\theta \mid x_1, x_2)$.

(5 marks)

iii) Find the Maximum A Posteriori (MAP) Estimate of θ.

(2 marks)

iv) State **TWO** differences between parametric and non-parametric estimation methods.

(4 marks)

b) Consider the following independently drawn samples

$$X = \{1,1,4,8,8,9\}, \quad N = 6.$$

(Sub-total: 19)

i) Find $p(x)$ for $x = 2$ using the Parzen window (rectangular) with a bandwidth $h_d = 3$.

(6 marks)

ii) Find $p(x)$ for $x = 5$ using the k-Nearest Neighbour (kNN) with $k_n = 3$.

(6 marks)

iii) Suggest a method to choose the bandwidth for the Parzen window for the uniform kernel.

(1 mark)

iv) Suppose $X_1 = \{1,3,3,4,5\}$ and $X_2 = \{7.5, 8,9,11,12,13\}$ belong to class 1 and class 2, respectively. The kNN method is employed with $k_n = 3$. Compute the class label for $x = 7$.

(2 mark)

v) Explain why an even k_n should not be used in a two-class classification problem.

(2 mark)

vi) For the Pazen window and kNN, suggest **ONE** potential issue for each of the approaches.

(2 marks)

3.

a) Consider the following 10 independent samples $x_1, x_2, \ldots x_{10}$.

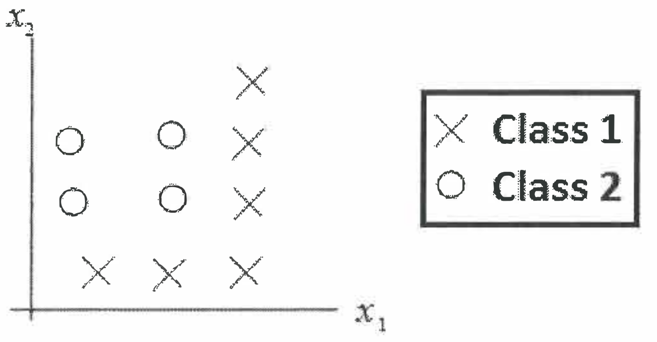

Figure 1. A plot of the 10 samples.

(Sub-total: 17)

i) Draw a neural network with fewest units that can separate the samples, in which the net activations contain only linear combinations of the inputs or the output of the previous layer.

(2 marks)

ii) Copy Figure 1 onto your answer booklet and indicate the separation planes created by the hidden layers.

(2 marks)

iii) <u>**List**</u> the total number of unknown parameters to be solved using back propagation for the neural network architecture defined in i) and <u>**explain**</u> how this number is obtained.

(3 marks)

iv) What is the advantage of deep learning over traditional machine learning approaches? Name **TWO** examples of deep learning architectures.

(4 marks)

v) Suppose $x_1, x_2, \ldots x_6$ belong to class 1 and $x_7, x_8, \ldots x_{10}$ belong to class 2. Student A plans to use a quadratic programming solver to find a primal hard-margin SVM classifier for the above samples. <u>**Write down the formulation**</u> of a quadratic programmer solver and <u>**explain**</u> how it can be used to solve the hard-margin SVM problem.

(6 marks)

b) Suppose there is a hidden class among the 10 samples in Figure 1.

(Sub-total: 17)

i) Describe how the *k*-means procedure can be used to perform clustering on the 10 samples in Figure 1 in words.

(4 marks)

ii) State the criterion employed in the *k*-means algorithm and suggest a short-coming of that criterion.

(2 marks)

iii) Define hard and soft memberships.

(1 mark)

iv) Find the 1st and 2nd Principal Components for the matrix $X = [[1\ 2]^T, [3\ 1]^T, [5,0]^T, [7,8]^T]^T$.

(6 marks)

v) Describe the working principle of Fourier descriptors.

(4 marks)

*** END OF PAPER ***