UNIVERSITY OF KELANIYA – SRI LANKA

FACULTY OF COMPUTING AND TECHNOLOGY

Bachelor of Science Honours in Computer Science Degree Examination – January 2022

Academic Year 2019/2020 – Semester I

Computer Science

BSCS 31032 – Machine Learning and Pattern Recognition

No. of Questions: **Five (05)** No. of Pages: **Three (03)** Time: **Two (02) hours.**

Answer any ***four*** questions.

---

1. (a) What is the key difference between a parametric and non-parametric model? **(4 marks)**

 (b) Write down two examples for each parametric and non-parametric machine learning algorithms. **(8 marks)**

 (c) List three limitations of Logistic Regression algorithm. **(6 marks)**

 (d) Suppose that the collected dataset $X = \{x_1, x_1, \dots, x_N\}$ follows the Poisson distribution given below:

$$p(x;\lambda) = \frac{\lambda^x exp\{-\lambda\}}{x!}.$$

 Write down the likelihood function and estimate the maximum likelihood parameter λ_{ML} using maximum likelihood parameter estimation method. **(7 marks)**

2. (a) Why do we need K-nearest neighbours (KNN) algorithm? **(5 marks)**

 (b) Explain the main steps of the KNN algorithm in point form. **(6 marks)**

 (c) Let A = $\{x_1, x_1, \dots, x_n\}$ be a set of n independent labeled samples and let $A_k(x) = \{x_1', x_2', \dots, x_k'\}$ be the k-nearest neighbours of x.

 Consider a two-category problem with $P(\omega_1) = (\omega_2) = 1/2$. Assume further that the conditional densities $p(x/\omega_i)$ are uniform within the unit hyperspheres a distance of ten units apart.

 (i) Show that if k is odd, the average probability of error is given by,

$$P_n(e) = \frac{1}{2^n}\sum_{j=0}^{(k-1)/2}\binom{n}{j}.$$

(6 marks)

(ii) Show that for this case the single-nearest neighbour rule has a lower error rate than the k-nearest neighbour error rate for $k > 1$. **(8 marks)**

3. Consider two univariate Gaussian distributions with means μ_1 and μ_2, respectively, and an arbitrary, but known and equivalent variance σ^2. Assume that you know the priors $P(\omega_1)$ and $P(\omega_2)$. The class-conditional densities are given by,

$$P\left(\frac{x}{\mu_i}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\frac{(x-\mu_i)^2}{\sigma^2}\right], i = 1,2.$$

Consider only the case of the zero-one loss function.

(a) Write down the Bayes discriminants, $g_1(x)$ and $g_2(x)$. **(5 marks)**

(b) What role do the priors play in this discriminant-based classification? **(8 marks)**

(c) The dichotomizer (a classifier that classifies into two categories) is a single discriminant function $g(x)$. For two-class situations $g(x) = g_1(x) - g_2(x)$. We decide ω_1 if $g(x) > 0$. Derive the log-likelihood ratio form of the dichotomizer (the rule for selecting class ω_1). **(12 marks)**

4. (a) State What do you mean by Principal Component Analysis (PCA)? **(5 marks)**

(b) What quantity is PCA maximizing during dimension reduction? **(4 marks)**

(c) The problem of principal component analysis ultimately reduces to the eigenproblem:

$$Se = \lambda e.$$

(i) Describe the variables S, e and λ in the above equation. **(10 marks)**

(ii) Given this problem setup for PCA, and assumed you are given a very high-dimensional dataset to work with (on the order of 100,000 or even 1,000,000 features). Would it be a good idea to start your analysis by reducing the data's dimensionality with this algorithm? Justify your answer. **(6 marks)**

5. Let Parzen window be defined as $\varphi(x) = e^{-x}$ for $x > 0$ and 0 for $x \leq 0$ and let $p(x) \sim U(0, a)$ be uniform from 0 to a.

(a) Show that the mean of such a Parzen-window estimate is given by,

$$\bar{p}_n(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{a}\left(1 - e^{-\frac{x}{h_n}}\right) & 0 \leq x \leq a \\ \frac{1}{a}\left(e^{\frac{a}{h_n}} - 1\right)e^{-\frac{x}{h_n}} & a \leq x. \end{cases}$$

(15 marks)

(b) Plot $\bar{p}_n(x)$ versus x for a=1 and h_n=1, 1/4 and 1/16. **(10 marks)**

---------------------------------- XXX ----------------------------------