B.Sc. Engg. CSE 8th Semester | 29 October 2019

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)
ORGANISATION OF ISLAMIC COOPERATION (OIC)
Department of Computer Science and Engineering (CSE)

SEMESTER FINAL EXAMINATION | **SUMMER SEMESTER, 2018-2019**

DURATION: 3 Hours | **FULL MARKS: 150**

CSE 4835: Pattern Recognition

Programmable calculators are not allowed. Do not write anything on the question paper.

There are **8 (eight)** questions. Answer any **6 (six)** of them.

Figures in the right margin indicate marks.

---

1. a) Prove that the derivative of the Sigmoid transfer function is: 5

$$\sigma' = \sigma(1-\sigma)$$

b) Draw a simple Neural Network to represent a linear machine with generalized discriminant function of $g_i(x) = w^t x + w_0$. Which transfer function did you use and why? 3+2

c) Show that if the transfer function of the hidden units is linear, a three-layer network with one hidden layer is equivalent to a two-layer one. 7

d) In a feed-forward neural network (NN), the weights w_{jk} of the edges to the output nodes are adjusted by the following term. 8

$$\frac{\partial E}{\partial w_{jk}} = \mathrm{O}_j \delta_k$$

$$\delta_k = \mathrm{O}_k(1-\mathrm{O}_k)(\mathrm{O}_k - t_k)$$

Taking into consideration the usual meaning of the notations used, how did the back-propagation algorithm devise this adjustment factor?

2. a) "Difference of Gaussian (DoG) will give image information at different frequency band" – Do you agree or disagree with this statement. Justify you choice. 7

b) How does the Harris Corner detection algorithm find key-points with high corner response? 8

c) Scale Invariant Feature Transform (SIFT) is a powerful feature descriptor which can provide scale and rotation invariant properties. How are these invariance properties ensured by SIFT features? 10

3. a) "Irrespective of the dimensionality of the data space, a data set consisting of just two data points, one from each class, is sufficient to determine the location of the maximum-margin hyperplane" – Do you agree or disagree with this statement. Justify your choice. 7

b) If the training samples of a two-class problem cannot be linearly classified in the original feature space, how does Support Vector Machine (SVM) try to classify them? Give an example. 10

c) Given, 8

$$f(\mathrm{x}, \mathrm{y}) = 2 - x^2 - 2y^2$$
$$\text{subject to } \mathrm{x} + \mathrm{y} - 1 = 0$$

Find the extreme values using Lagrange multipliers.

4. a) What is Cluster Analysis? Briefly describe the requirements for cluster analysis. 1+9

   b) Consider the data set as given in Table 1 consisting of the scores of two variables on each of seven individuals. 10

Table 1

Sample	X_1	X_2
1	**1.0**	**1.0**
2	**1.5**	**2.0**
3	**3.0**	**4.0**
4	**5.0**	**7.0**
5	**3.5**	**5.0**
6	**4.5**	**5.0**

Apply the *k*-means clustering algorithm with the value $k = 2$. Show required calculations upto two cluster center updates.

   c) Which clustering algorithm is more robust in presence of outliers: *k*-means or *k*-medoids? Explain why. 5

5. a) What is the curse of dimensionality? How can you deal with this problem with a linear projection method? Briefly describe its working mechanism. 2+9

   b) Let an orthonormal transformation $y = \Phi^T x$, where the matrix Φ contains all eigenvectors. Show that for orthonormal transformations, Euclidean distances are preserved, i.e., $\|y\|^2 = \|x\|^2$. 7

   c) In which cases, the Mahalanobis distance gives a better sense of dissimilarity measure in compared to Euclidean distance? Explain with necessary figures. 7

6. a) What characteristics should an ideal feature extractor hold during extracting a feature vector representing an object? Briefly explain each of them with examples 7

   b) In a two-class problem, prove that the weight vector w representing a linear decision boundary is perpendicular to that boundary. Also show that the weight vector w can also be considered as a projection line on which samples are being projected and then classified by comparing against a threshold '$-w_0$' (Bias). 5+5

   c) Prove that the minimum distance classifier with respect to class member is nonlinear. Devise its discriminant function definition first. 8

7. a) In many pattern classification problems one has the option either to assign the pattern to one of c classes, or to *reject* it as being unrecognizable. If the cost for rejects is not too high, rejection may be a desirable action. 10

$$\text{Let } \lambda(\alpha_i \mid \omega_j) = \begin{cases} 0 & i = j \quad i, j = 1, 2, \ldots, c \\ \lambda_r & i = c + 1 \\ \lambda_s & \text{otherwise} \end{cases}$$

where, λ_r is the loss incurred for choosing the $(c+1)^{\text{th}}$ action that is rejection, and λ_s is the loss incurred for making a substitution error. Show that the minimum risk is obtained if we decide ω_i if $P(\omega_i|x) \geq P(\omega_j|x)$ for all j and if $P(\omega_i|x) \geq 1 - \lambda_r/\lambda_s$, and reject otherwise.

   b) Suppose, in a Bayes classifier the likelihood probability follows a normal distribution, each class has its own covariance matrix and a prior probability which is different from other classes. Devise the equation of the decision boundary for each pair classes. What conditions are required for this classifier to behave like a distance classifier? 8+2

c) Suppose you have a classifier which can classify samples into any of the two classes. Extend this classifier to make it suitable for a multi-class problems. 5

8. a) Suppose we have n samples $\{x^1, x^2, \ldots, x^n\}$ where each sample x be a d-dimensional binary (0 or 1) vector with a multivariate Bernoulli distribution: 10

$$P(x \mid \theta) = \prod_{i=1}^{d} \theta_i^{x_i} (1-\theta_i)^{1-x_i}$$

where $\theta = (\theta_1, \ldots, \theta_d)^T$ is an unknown parameter vector, θ_i being the probability that $x_i = 1$. Show that the maximum-likelihood estimate for θ is

$$\hat{\theta} = \frac{1}{n} \sum_{k=1}^{n} x^k$$

b) In Parzen window method, can we use a Gaussain window function instead of a rectangular window function $\varphi(u)$ for calculating k_n? If, yes, then define a Gaussian window function and explain the corresponding effects on the final density estimate of $p_n(x)$. 7

c) For the following sixteen samples in a one-dimensional problem: 8

$$D = \{0.5, 1.3, 2.3, 4.5, 5.5, 6.0, 6.5, 7.1, 7.2, 7.5, 8.3, 8.8, 9.2, 9.3, 12, 14\}$$

Give the values of the k-nearest neighbor estimate $p_n(x)$ at positions $x=1.0$, $x=6.5$, $x=8.0$, and $x=10.2$, for $n=16$ and $k_n = \sqrt{n}$,