B. Engg. CSE 8th Semester (23) 322 19 October 2014 (Morning)

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

SEMESTER FINAL EXAMINATION SUMMER SEMESTER, 2013-2014

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.

---

2) a) In the early 1900's, several investigators were interested in predicting behavioral and social outcomes among people based on physical characteristics. Macdonnell (1902) reported on the following seven physical variables measured on 3000 British criminals: head length, head breadth, face breadth, left finger length, left forearm length, left foot length and height. Assume all the variables have some values. 5+5

i. One of the goals of PCA is to reduce dimension of the original data. How would you choose the number of principal components to retain for subsequent analyses?

ii. Suppose that in addition to the seven variables described above, an eighth variable, computed as head length minus head breadth, had been included in the analysis to capture head shape. What would be the variance of the last principal component in such an analysis and why?

b) Suppose the covariance matrix **A** has trace value of 4, and two of its Eigen values are 1 and 0. Comment on the distribution pattern of the original samples in the 3D feature space from which that covariance matrix is computed. 5

c) In what sense is AdaBoost a greedy algorithm? Why is this necessary? 5

3) a) Given the following equations: 10

$$f(x,y) = x^3 + y^3$$

$$g_1(x,y) = x^2 - 1 \geq 0$$

$$g_2(x,y) = y^2 - 1 \leq 0$$

Find the extreme values.

b) Lagrange Multipliers are a mathematical method used to solve constrained optimization problems of differentiable functions. How are these multipliers used in a Support Vector Machine (SVM) for finding the support vectors? 10

c) Describe the 'Line of no Discrimination' in ROC curve. 5

4) a) Formulate the criterion function J of Linear Discriminant Analysis (LDA) for a two-class problem based on several factors which influence the design of that function. 10

b) What are the differences between PCA and LDA? 5

c) Suppose you are designing a 2D Human Activity Recognition (HAR) system capable of classifying four different activities, such as jogging, walking, running and skipping. By incorporating LDA in your feature representation module, how can you get the final feature vector for any activity? 10

4. a) Consider the following sixteen samples in a one-dimensional problem:

$\{x^1, x^2, \ldots, x^{16}\} = \{0, 1, 3, 4.5, 5.5, 6.0, 6.5, 7.0, 7.2, 7.5, 8.0, 8.8, 9.2, 9.3, 11, 13\}$

Give the values of the *k*-nearest neighbor estimate $p_n(x)$, for $n=16$ and $k_n = \sqrt{n}$, at $x=2$, $x=4$, $x=6$, $x=8.5$, and $x=10$.

b) Briefly discuss the differences between Maximum-Likelihood and Bayesian parameter estimation.

c) Why is *n*-fold cross validation used for evaluating a pattern recognition system? What does happen when *n* is equal to the total number of samples?

5. a) In a feed-forward neural network (NN), the weights w_{ij} of the edges to the hidden nodes are adjusted by the following term. Taking into consideration the usual meaning of the notations used, how did the back-propagation algorithm devise this adjustment factor?

$$\frac{\partial E}{\partial w_{ij}} = O_i \delta_j$$

$$\delta_j = O_j(1-O_j)\sum_{k \in K} \delta_j w_{jk}$$

b) The most common symbolic representation of numbers in the world today is done with 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. Suppose you are asked to recognize printed numbers from bank cheques from a designated area only. If you are constrained to use neural network as a classifier, how would you design the complete pattern recognition system?

6. a) In a two-class one-dimensional problem, the pdfs are the Gaussians $N(0, \sigma^2)$ and $N(1, \sigma^2)$ for the two classes, respectively. Show that the threshold x_0 minimizing the average risk is equal to

$$x_0 = \frac{1}{2} - \sigma^2 \ln \frac{\lambda_{21} P(\omega_2)}{\lambda_{12} P(\omega_1)}$$

where $\lambda_{11} = \lambda_{22} = 0$ has been assumed.

b) In a two-class, three-dimensional classification problem, the feature vectors in each class are normally distributed with the covariance matrix:

$$\Sigma = \begin{bmatrix} 0.3 & 0.1 & 0.1 \\ 0.1 & 0.3 & -0.1 \\ 0.1 & -0.1 & 0.3 \end{bmatrix}$$

The respective mean vectors are $[0, 0, 0]^T$ and $[0.5, 0.5, 0.5]^T$. Derive the corresponding linear discriminant functions and the equation describing the decision surface.

c) Briefly explain the properties that an efficient feature should hold for better recognition rate.

7. a) How does the Bayes Parameter method utilize the prior information of unknown parameters?

b) With the help of log-likelihood function, find the unknown parameter θ with the help of *MLE* method. Here $\theta=\{\mu,\sigma^2\}$.

c) Define the following terms with examples:
   i. False Positive Rate
   ii. True Position Rate
   iii. Precision
   iv. Accuracy

a) Draw a diagram illustrating the design cycle of a pattern recognition system and explain how feedback information from classifier evaluation may change each component of the design cycle. 10

b) Suppose you have two normal distributions with the same covariance but different means. In terms of their prior probabilities $P(\omega_1)$ and $P(\omega_2)$, state the condition that the Bayes decision boundary will not pass between the two means. 5

c) Consider the Batch Perceptron algorithm to find the weight coefficients of your discriminant function for a multi-class problem.Explain the differences in weight-updating as compared to the two-class problem. 10