B.Sc. Engg. CSE 8th Semester (15)

335

19 October 2015

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

SEMESTER FINAL EXAMINATION **SUMMER SEMESTER, 2014-2015**

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) For the following dataset in Table 1, plot the Receiver Operating Curve (ROC) for different score value as a threshold. The true class label is given in the active/decoy column. 15

Table 1

Id	Score	Active/Decoy	Id	Score	Active/Decoy
1	.03	a	10	.48	a
2	.08	a	11	.56	d
3	.10	d	12	.65	d
4	.11	a	13	.71	d
5	.22	d	14	.72	d
6	.32	a	15	.73	a
7	.35	a	16	.80	d
8	.42	d	17	.82	d
9	.44	d	18	.99	d

b) Write short notes on the followings: 5×2

i. False Positive Rate
ii. Accuracy
iii. Recall
iv. Specificity
v. Line of No Discrimination

2. a) For the given two-class problem in Figure 1, which Support Vector Machine (SVM) classifier will you use: Linear or Non-linear? How will you find the decision boundary? Explain your answer. 10

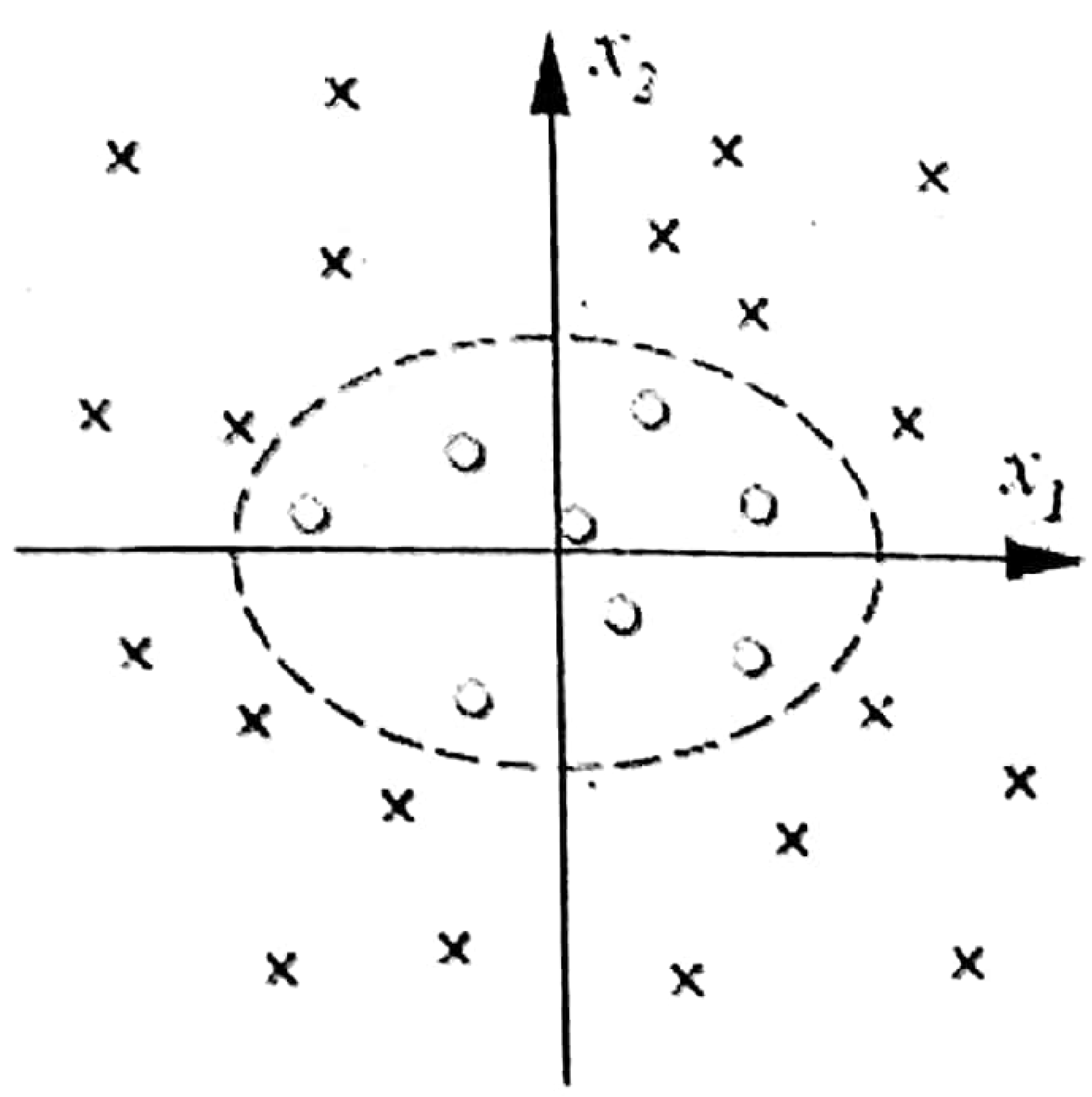

Figure 1

b) List the different kernels used in SVM, along with their kernel function definitions. 5

336

c) Given the following equations:

$$f(x,y) = x^2 + y^2$$
$$g_1(x,y) = x + 1 \le 0$$
$$g_2(x,y) = y + 1 \le 0$$

Find the extreme values using Lagrange Multipliers.

3. a) In an expression classification system, how can eigen-vector and eigen-value concepts be useful for training your recognizer for six different expressions (happy, sad, angry, surprise, fear, neural) and make it computationally efficient?

b) From the newly transformed feature vector **y**, how can you retain your original vector **x**, when the eigen-vectors are stored in matrix V as columns $v^1, v^2, v^3 \ldots v^n$?

c) In Figure 2, an object has been rotated from its vertical position to an unknown angle. How can you recognize this same object after rotation and determine its orientation?

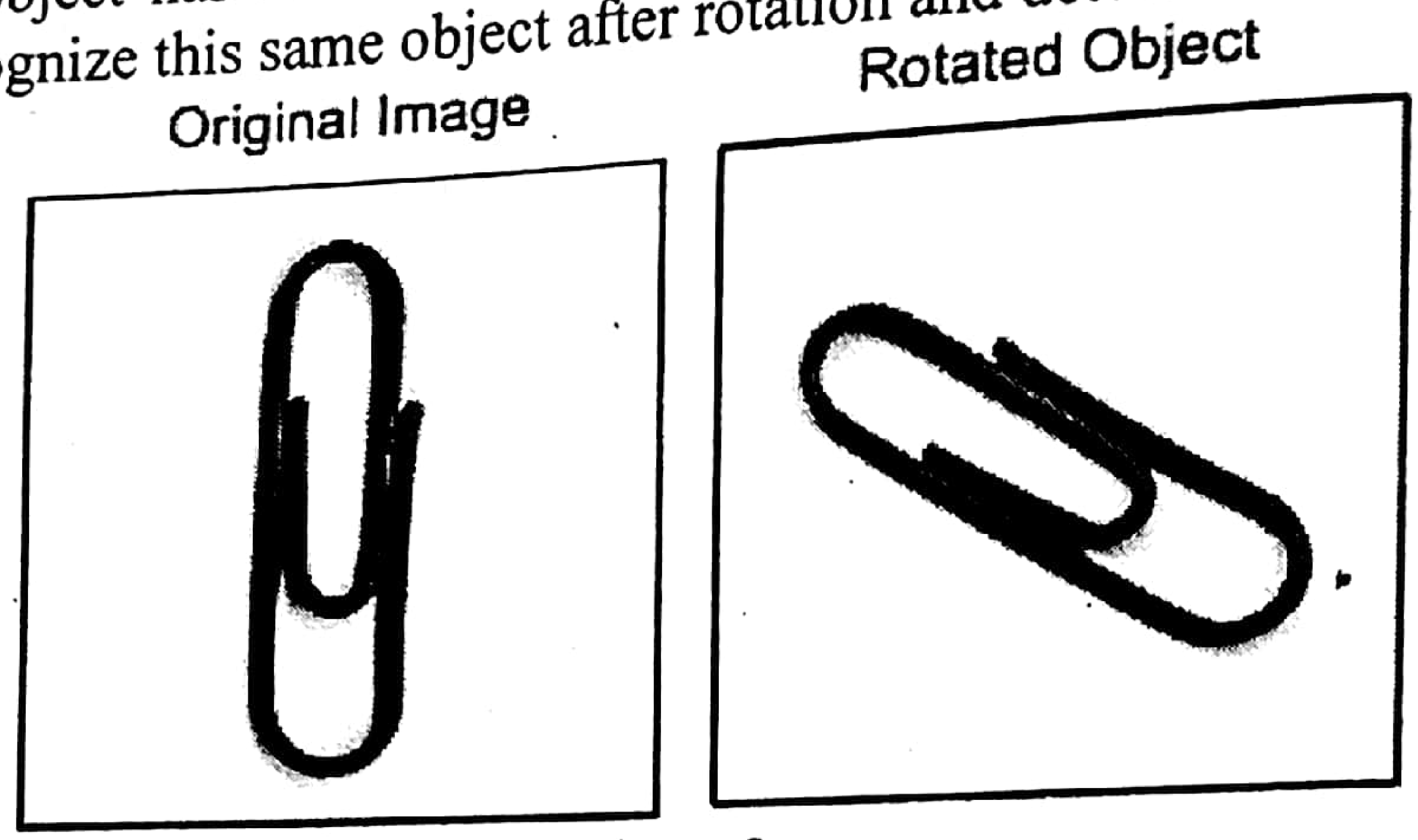

Figure 2

4. a) Explain why and how distance normalization is incorporated in Mahalanobis distance measure. Illustrate with figures.

b) Suppose two discriminant functions g_1 and g_2 are defined based on the Bayes formula. Provide the equation of their decision boundary. What happens if the prior probabilities change their values, e.g., one class has higher prior probability than the other class.

c) Suppose the discriminant functions in Question 4.(b) are defined on a multivariate normal density function in d dimensions. If feature elements in all classes are statistically independent and has the same variance, then derive the general form of the discriminant function g. Which conditions will convert this classifier to the 'minimum distance to class mean' classifier?

5. a) Which factors influence the design of the criterion function J in Linear Discriminant Analysis (LDA)? Discuss the effects of each of those terms and write the final form of J.

b) For feature dimensionality reduction, a research student has used PCA first and then on the new feature set he has applied Linear Discriminant Analysis (LDA). An anonymous reviewer of a reputed journal doubted the use of LDA after PCA, and suggested that using LDA is sufficient. Would you accept the opinion of the reviewer or defend your idea? If you are that research student, what will be your reply to the reviewer?

c) What are the differences between PCA and LDA?

6. 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.

b) Suppose x have a uniform density 9

$$p(x|\theta) \doteq U(0,\theta) = \begin{cases} 1/\theta & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}.$$

If n samples D={$x_1, x_2, \ldots, x_n$} are drawn independently according to p(x| θ), then show that the maximum likelihood estimate for θ is max[D], i.e., the value of the maximum element in D.

c) Compare between Supervised, Unsupervised and Reinforcement Learning with examples. 6

7. a) For the following sixteen samples in a one-dimensional problem: 10

{$x^1, x^2, \ldots, x^{16}$} = {0, 1.2, 2.5, 4.5, 4.7, 6.0, 6.5, 7.0, 7.2, 7.2, 8.0, 8.8, 9.2, 9.3, 11, 13}

Find the values of the k-nearest neighbor estimate $p_n(x)$, for n=16 and $k_n = \sqrt{n}$, at $x=0$, $x=3$, $x=6$, $x=7.2$, and $x=9$.

b) What conditions should be satisfied for the nonparametric density estimate $p_n(x)$ to converge to true density $p(x)$? 6

c) In generalized non-parametric techniques, how is the density function estimated to $p(x) \simeq \frac{k/n}{V}$ where symbols have their conventional meaning. 6

d) Among the most popular two non-parametric techniques, which one is more preferable for density estimation? Justify your answer. 3

8. a) What are the differences between Maximum Likelihood Estimation (MLE) and Bayes Parameter estimation methods? 6

b) In Bayes parameter estimation for $\theta = \{\mu\}$, the estimates are found by 8

$$\mu_n = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} * \hat{x}_n + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} * \mu_0$$

where prior probability is defined by $p(\mu) \doteq N(\mu_o, \sigma_0)$. Explain how the two extreme values of the estimated μ are either average of samples $\hat{x}_n$ or initial guess μ_0.

c) You are given the following sample points in a 3-class problem: 11

$$\omega_1 = (0,2),\ (1,3),\ (1,1)$$
$$\omega_2 = (1.5,0),\ (0,-1),\ (2,-1)$$
$$\omega_3 = (4,0),\ (4,3),\ (5,7)$$

i. Plot the above samples and determine whether the classes are linearly separable in the original space.

ii. Design a ϕ-function to generate new higher dimensional samples y.

iii. Use Perceptron criteria function (one-at-a-time) to find the linear weight-coefficients for each class. Initial weight vector can be chosen as you wish and update them using the following formula:

"For each sample y, if y belongs to ω_i but mistakenly assigned to ω_j then:

$$w^i(t+1) = w^i(t) + \alpha y$$
$$w^j(t+1) = w^j(t) - \alpha y$$
$$w^k(t+1) = w^k(t) \text{ for all } k \neq i,j$$

where α is a learning step."

[Note: During weight update, show calculations up to 2 complete iterations.]