Roll No.: _

Amrita VishwaVidyapeetham

Amrita School of Engineering, Coimbatore

B.Tech. Degree Examinations – October 2019

Seventh Semester

Computer Science and Engineering

15CSE361 Pattern Recognition

Time: Three hours Maximum: 100 Marks

Course Outcomes (COs):

CO	Course Outcomes
CO01	Understand the working principles of pattern recognition system and algorithms
CO02	Apply statistical methods for decision making
CO03	Understand non parametric decision making system
CO04	Apply and evaluate Non-metric approach for real world problems
CO05	Apply and analyze unsupervised learningmethods for real world problems.

Answer all questions

1. In a set of 400 samples, x varies from 600 to 800. At what values of x would the boundaries of the histogram intervals be located using the square root rule? [3] [CO 03]

2. Suppose the average number of lions seen on a 1-day safari is 5. What is the probability that tourists will see fewer than four lions on the next 1-day safari? [5] [CO 01]

3. a. There are two classes and the penalty for misclassifying a sample that belongs to class A is 1. The reward for correctly classifying a sample from class A is 3. For a sample from B, the penalty for misclassifying it is 4 and the reward for correctly classifying it is 6. If a sample has P(A/x)=2/3. What is the expected loss (risk) of choosing class A? of choosing class B? What class should be chosen? [6] [CO 02]

b. Class A has a symmetric triangular density ranging from 0 to 4, and class B has a uniform density ranging from 2 to 6. The prior probabilities and costs are the same for both classes. What are the probabilities of error for class A and for class B if the decision region x=3 has been used? [6] [CO 03]

4. In a two-class problem, a classifier called 45 of the 60 As as As, and all 60 of the Bs as Bs. Form a confusion matrix and fractional confusion matrix for these results. [6] [CO 04]

5. a. How the new sample is being classified using bayesian classifier? Explain how decision boundary is obtained and how decision making can be done effectively using decision boundary. [5] [CO 02]

Comment [k1]: Font: Times New Roman –size

Comment [k2]: Font: Times New Roman –size

Comment [k3]: Font: Times New Roman –size

Comment [k4]: Font: Times New Roman –size

Comment [k5]: Font: Times New Roman –size

Comment [k6]: If the course is common to ma branches please indicate here as (common to.....

Comment [k7]: Font: Times New Roman –size

Comment [GK8]: Font: Times New Roman (BOLD)-size-12

Comment [k9]: Font: Times New Roman –size

b. Two binary tests, x and y, are relevant in determining three disease states A, B, and, C which are mutually exclusive. The probabilities that x is positive, given states A, B, and C, are 0.3, 0.5 and 0.7, respectively. For test y, these numbers are 0.8, 0.2, and 0.6. The prior probabilities of A, B, and C are 1/6, 1/3, and 1/2. What are the probabilities of A, B, and C if x is positive and y is negative? What assumptions did you have to make to obtain the result? [5] [CO 02]

6. a. At some point in training an adaptive decision boundary, D= 1-2x+3xy-4y. D>0 for members of class A. The next sample is with x=1, y=-2, and was a member of class B. If c=1 and k=2, what will the decision boundary become after adapting it for this sample?

[5] [CO 03]

b. Use a symmetric triangular window with a base of 4 to estimate the density at x=4, given samples at 2,3,4, and 5. [5] [CO 03]

7. There are 7 medicines in the training data points object and each medicine has 2 attributes. Each attribute represents coordinate of the object. Determine which medicines belong to cluster 1, cluster 2 and which medicines belong to the other cluster after the first iteration. Randomly choose the following three centroids for three clusters. m1=(1.0,1.0), m2=(5.0,7.0) and m3=(3.0,4.0) [10] [CO 04]

Object	Attribute1 (X): weight index	Attribute 2 (Y): pH
Medicine A	1	1
Medicine B	1.5	2
Medicine C	3	4
Medicine D	5	7
Medicine E	3.5	5
Medicine F	4.5	5
Medicine G	3.5	4.5

8.Assume that the database D is given by the table below. Follow single linkage agglomerative technique to find clusters in D. Use Euclidean distance measure.

[10] [CO 04]

Database D

	x	y
p1	0.40	0.53
p2	0.22	0.38
p3	0.35	0.32
p4	0.26	0.19
p5	0.08	0.41
p6	0.45	0.30

9. Construct a decision tree for the following training set using ID3 algorithm. Classify the following record using the decision tree.Name =Human, Body Temperature =warm-blooded, Gives Birth=yes, Four legged=no, Hibernates=no, Class label = ? [10] [CO 04]

Name	Body Temperature	Gives Birth	Four legged	Hibernates	Class label (Mammals)
salamander	Cold-blooded	no	yes	yes	no
Komodo dragon	Cold-blooded	no	yes	no	no
python	Cold-blooded	no	no	yes	no
salmon	Cold-blooded	no	no	no	no
porcupine	Warm-blooded	Yes	Yes	yes	yes
cat	Warm-blooded	Yes	Yes	no	yes
bat	Warm-blooded	yes	no	yes	no
whale	Warm-blooded	yes	no	no	no
eagle	Warm-blooded	no	no	no	no
guppy	cold-blooded	yes	no	no	no

10. The true classes are known for a testset of six samples and the estimated probabilities of class membership produced by a classifier are given in the following table. [10] [CO 02]

Sample	True class	$\hat{P}(A\|x)$	$\hat{P}(B\|x)$	$\hat{P}(C\|x)$
1	A	0.3	0.4	0.3
2	A	0.5	0.1	0.4
3	B	0.5	0.3	0.2
4	B	0.1	0.8	0.1
5	C	0.2	0.6	0.2
6	C	0.1	0.1	0.8

a. Find the estimated probability of error using simple counting. (3M)
b. Find the estimated probability of error using fractional counting. (3M)
c. What is the integer confusion matrix for the above test set? (2M)
d. What is the fractional confusion matrix for the above test set?(2M)

11. Classify the new sample with using three nearest neighbour for the feature values: X1=3,x2=7 ,Y=?. Use Euclidean distance measure. [6][CO 03]

X1=Acid Durability	X2= strength	Y= classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

b. Use the same samples as training set and testing set and estimate the error rate and accuracy of the K NN classifier for k=3. [8][CO 03]

****