F 10000CS467122105

Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Seventh Semester B.Tech Degree (S,FE) Examination January 2023 (2015 Scheme)

Course Code: CS467

Course Name: MACHINE LEARNING

Max. Marks: 100 Duration: 3 Hours

PART A

Answer all questions, each carries 4 marks. Marks

1 Explain Outlier detection. 4

2 List out any four applications of Pattern recognition. 4

3 Compare forward and backwards subset selection methods. 4

4 State No Free Lunch theorem. 4

5 Three factories A, B, C of an electric bulb manufacturing company produces 45%, 35% and 20% of the total output. Approximately 1.5%, 1% and 2% of the bulbs produced by these factories are known to be defective. If a randomly selected bulb manufactured by the company was found to be defective, what is the probability that the bulb was manufactured in factory B? 4

6 Explain Decision trees using an example. 4

7 When do we say data is linearly separable? 4

8 Explain Soft Margins SVM Hyperplanes. 4

9 Illustrate clustering with an example. 4

10 By using K-means clustering algorithm divide the following data into two clusters and also compute the representative data points for the clusters: 4

x1	1	2	2	3	4	5
x2	1	1	3	2	3	5

PART B

Answer any two full questions, each carries 9 marks.

11 a) Explain Reinforcement learning with an example, how is it different from Supervised and Unsupervised learning? 5

b) Compare Supervised Learning with Unsupervised Learning. 4

12 a) Explain PAC learning with example. 5

b) Explain the following: 4

a) Training b)Validation c)Cross validation d)Test set

13 a) What is Triple Trade off? 3

b) Differentiate between Feature Selection and Feature Extraction. 2

c) Explain Unsupervised Learning with example. 4

PART C

Answer any two full questions, each carries 9 marks.

14 a) A database contains 80 records on a particular topic of which 55 are relevant to a certain investigation. A search was conducted on that topic and 50 records were retrieved. Of the 50 records retrieved, 40 were relevant. Construct the confusion matrix for the search and compute the precision and recall scores for the search. 5

b) Explain Cross Validation and resampling method. 4

15 a) Using the data given in the following table, construct a tree to predict the value of y: 5

x_1	1	3	4	6	10	15	2	7	16	0
x_2	12	23	21	10	27	23	35	12	27	17
y	10.1	15.3	11.5	13.9	17.8	23.1	12.7	43.0	17.6	14.9

b) How can we calculate the entropy of a data set? What does the entropy value signify? 4

16 a) Discuss some issues in Decision Tree Learning. 4

b) What are the problems encountered in cross validation method? 3

c) What do you mean by stratification? 2

PART D

Answer any two full questions, each carries 12 marks.

17 a) Explain the ADABOOST algorithm. 6

b) Using SVM algorithm, find the SVM classifier for the following data: 6

Example No	x1	x2	Class
1	2	1	+1
2	4	3	-1

18 a) Explain types of hierarchical clustering. 6

b) Given the dataset {a,b,c,d,e} and the distance matrix in the following table, construct a dendrogram by Single-Linkage Hierarchical clustering using the Agglomerative method: 6

	a	b	c	d	e
a	0	9	3	6	11
b	9	0	7	5	10
c	3	7	0	9	2
d	6	5	9	0	8
e	11	10	2	8	0

19 a) What are the different ways to achieve diversity in learning? 5

b) Explain algorithm for agglomerative hierarchical clustering. 4

c) Illustrate Kernel Trick Method. 3

****