Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Seventh semester B.Tech examinations (S), September 2020

Course Code: CS467

Course Name: MACHINE LEARNING

Max. Marks: 100 Duration: 3 Hours

PART A

Answer all questions, each carries 4 marks. Marks

1		Define VC dimension. How VC dimension is related with no of training examples used for learning.	(4)
2		Compare Classification with regression with an example.	(4)
3		Distinguish between overfitting and underfitting. How it can affect model generalization?	(4)
4		Explain the general MLE method for estimating the parameters of a probability distribution.	(4)
5		Compare Cross validation with Bootstrapping Techniques.	(4)
6		Calculate the output y of a three input neuron with bias. The input feature vector is (x1, x2, x3) = (0.8,0.6,0.4) and weight values are [w1,w2,w3, b] =[0.2, 0.1, -0.3, 0.35]. Use binary Sigmoid function as activation function.	(4)
7		Describe the significance of Kernal functions in SVM. List any two kernel functions.	(4)
8		Explain the basic elements of a Hidden Markov Model (HMM). List any two applications of HMM.	(4)
9		Explain any two model combination scheme to improve the accuracy of a classifier.	(4)
10		Compare K means clustering with Hierarchical Clustering Techniques.	(4)

PART B

Answer any two full questions, each carries 9 marks.

11	a)	Distinguish between supervised learning and Reinforcement learning. Illustrate with an example.	(5)
	b)	Discuss any four examples of machine learning applications.	(4)
12	a)	Define Probably Approximately Learning.	(3)
	b)	Explain the procedure for the computation of the principal components of the data.	(6)
13	a)	Compare Feature Extraction and Feature Selection techniques. Explain how dimensionality can be reduced using subset selection procedure.	(5)
	b)	Explain the methods used to learn multiple classes for a K class Classification Problem.	(4)

PART C

Answer any two full questions, each carries 9 marks.

14 a) Identify the first splitting attribute for decision tree by using ID3 algorithm with the following dataset. (6)

Major	Experience	Tie	Hired?
CS	programming	pretty	NO
CS	programming	pretty	NO
CS	management	pretty	YES
CS	management	ugly	YES
business	programming	pretty	YES
business	programming	ugly	YES
business	management	pretty	NO
business	management	pretty	NO

b) Explain perceptron learning algorithm. (3)

15 a) Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the precision and recall for the data. (4)

b) Consider the training data in the following table where Play is a class attribute. In the table, the Humidity attribute has values "L" (for low) or "H" (for high), Sunny has values "Y" (for yes) or "N" (for no), Wind has values "S" (for strong) or "W" (for weak), and Play has values "Yes" or "No". (5)

Humidity	Sunny	Wind	Play
L	N	S	No
H	N	W	Yes
H	Y	S	Yes
H	N	W	Yes
L	Y	S	No

What is class label for the following day (Humidity=L, Sunny=N, Wind=W), according to naïve Bayesian classification?

16 a) What are the benefits of pruning in decision tree induction? Explain different approaches to tree pruning? (5)

b) Given the set of values X = (3, 9, 11, 5, 2)^T and Y = (1, 8, 11, 4, 3)^T . Evaluate the regression coefficients. (4)

PART D

Answer any two full questions, each carries 12 marks.

17 a) Explain DBSCAN algorithm for density based clustering. List out its advantages compared to K-means. (6)

b) Describe How Evaluation problem can be solved by using Hidden Markov Model. (6)

18 a) Describe the significance of soft margin hyperplane and explain how they are computed. (6)

b) Illustrate K means clustering algorithm with an example. (6)

19 a) State the mathematical formulation of the SVM problem. Give an outline of the method for solving the problem. (6)

b) Show the final result of hierarchical clustering with complete link by drawing a dendrogram. (6)

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.70	0.45	0	
F	0.34	0.61	0.93	0.20	0.67	0

****