Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SEVENTH SEMESTER B.TECH DEGREE EXAMINATION, DECEMBER 2018

Course Code: CS467

Course Name: MACHINE LEARNING

Max. Marks: 100 Duration: 3 Hours

PART A

Answer all questions, each carries 4 marks. Marks

1 Distinguish between classification and regression with an example. (4)

2 Define Hypothesis space and version space for a binary classification problem. Determine the hypothesis space H and version space with respect to the following data D. (4)

x	2	11	17	0	1	5	7	13	20
Class	0	1	1	0	0	0	0	1	1

3 State Occam's razor principle. Illustrate its necessity in learning hypothesis. (4)

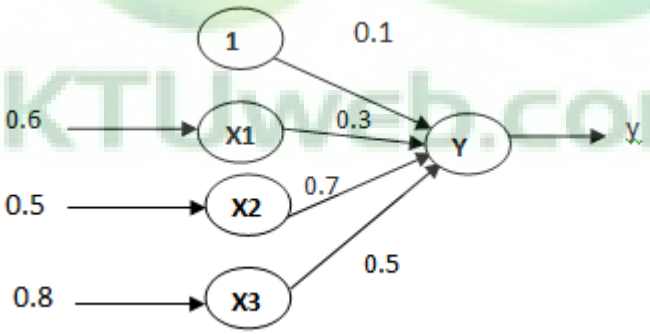
4 Define the following terms (a) sensitivity (b) Specificity (c) Precision (d) Accuracy for a classification problem. (4)

5 What is meant by k-fold cross validation. Given a data set with 1200 instances, how k- fold cross validation is done with k=1200. (4)

6 Calculate the output of the following neuron Y if the activation function as (4)

(a) Binary sigmoid

(b) Bipolar sigmoid



7 Distinguish between bagging and boosting. (4)

8 Show that the function $K(x, y) = (x.y)^3$ is a kernel function where x.y represents dot product of x and y. (4)

9 Briefly describe the concept of Expectation Maximization algorithm. (4)

10 Consider two data points in two dimensional A(5,8) and B(8,9).Calculate (4)

(a) City block distance (b) Chessboard distance between A & B.

PART B

Answer any two full questions, each carries 9 marks.

11 a) An open internal in R is defined as(a,b) = {x€R | a<x<b}. It has two parameters a and b. Show that the set of all open intervals has a VC dimension of 2 (4)

b) Compare Unsupervised Learning and Reinforcement Learning with examples. (5)

12 a) Discuss the necessity of dimensionality reduction in machine learning. (3)

b) Illustrate the idea of PCA for a two dimensional data using suitable diagrams. (6)

13 a) Let $X = R^2$ and C be the set of all possible rectangles in two dimensional plane which are axis aligned (not rotated). Show that this concept class is PAC learnable. (6)

b) Describe the applications of machine learning in any three different domains. (3)

PART C

Answer any two full questions, each carries 9 marks.

14 The following table consists of training data from an employee database. For a given row entry, count represents the number of data tuples having the values for department, status, age, and salary given in that row. Let status be the class label attribute. Given a data tuple having the values "systems", "31..35", and "46–50K" for the attributes department, age, and salary, respectively, what would a Naive Bayesian classification of the status for the tuple be? (9)

Department	Status	Age	Salary	Count
sales	senior	31..35	46K..50K	30
sales	junior	26..30	26K..30K	40
sales	junior	31..35	31K..35K	40
systems	junior	21..25	46K..50K	20
systems	senior	31..35	66K..70K	5
systems	junior	26..30	46K..50K	3
systems	senior	41..45	66K..70K	3
marketing	senior	36..40	46K..50K	10
marketing	junior	31..35	41K..45K	4
secretary	senior	46..50	36K..40K	4
marketing	junior	26..30	26K..30K	6

15 With the following data set, generate a decision tree and predict the class label for a data point with values <Female, 2, standard, high>. (9)

Gender	Car Ownership	Travel cost	Income level	Transport mode
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car
Female	1	Cheap	Medium	Train
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train

16 a) Point out the benefits of pruning in decision tree induction. Explain different (5)

approaches to tree pruning.?

b) Compute ML estimate for the parameter p in the binomial distribution whose probability function is (4)

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} \quad x = 0,1,2\ldots n.$$

PART D

Answer any two full questions, each carries 12 marks.

17 a) Explain the basic problems associated with hidden markov model. (6)

b) Describe the significance of soft margin hyperplane and optimal separating hyperplane and explain how they are computed. (6)

18 a) Suppose that the datamining task is to cluster the following seven points (with (x,y) representing location) into two clusters A1(1,1), A2(1.5,2), A3(3,4), A4(5,7), A5(3.5,5), A6(4.5,5), A7(3.5,4.5) The distance function is City block distance. Suppose initially we assign A1,A5 as the centre for each cluster respectively. Using the K-means algorithm to find the three clusters and their centres after two round of execution. (6)

b) Give the significance of kernel trick in the context of support vector machine. Describe different types standard kernel functions. (6)

19 a) Describe any one technique for Density based clustering with necessary diagrams. (6)

b) Given the following distance matrix, construct the dendogram using single linkage, complete linkage and average linkage clustering algorithm. (6)

Item	A	B	C	D	E
A	0	2	3	3	4
B	2	0	3	5	4
C	3	3	0	2	6
D	3	5	2	0	4
E	4	4	6	4	0

****

Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SEVENTH SEMESTER B.TECH DEGREE EXAMINATION(R&S), DECEMBER 2019

Course Code: CS467

Course Name: MACHINE LEARNING

Max. Marks: 100 Duration: 3 Hours

PART A

Answer all questions, each carries 4 marks. Marks

1	Identify the suitable learning method in each case and Explain it. (a) Grouping people in a social network (b)Training a robotic arm	(4)
2	Explain the concept of Overfitting and Underfitting model with suitable diagrams.	(4)
3	Define VC dimension. Show that VC dimension of a line hypothesis is three.	(4)
4	Compare Gain ratio with Information gain for attribute selection. Explain the advantage of using Gain ratio over Information gain for finding best split for constructing a decision tree.	(4)
5	Compute the Maximum Likelihood estimate for the parameter λ in the Poisson distribution whose probability function is $f(x) = \frac{e^{-\lambda}\,\lambda^{x}}{x!} \quad x = 0,1,2 \ldots n$	(4)
6	Why does a single perceptron cannot simulate simple XOR function ? Explain how this limitation is overcome?	(4)
7	Describe any two techniques used for Ensemble Learning.	(4)
8	Explain Kernel Trick in the context of support vector machine. List any two kernel function used in SVM.	(4)
9	Describe the basic concepts of Expectation Maximization Algorithm.	(4)
10	Calculate the dissimilarity between two data points x1(2,3,4) and x2(4,3,5) using (a) Euclidian distance (b) Manhattan Distance	(4)

PART B

Answer any two full questions, each carries 9 marks.

11	a)	Is regression a supervised learning technique? Justify your answer. Compare regression with classification with examples.	(5)

b) Explain (a) Hypothesis space (b) Version space (c) Most General hypothesis (d) Most specific hypothesis in the context of a classification problem. (4)

12 a) Explain the concept of PAC learning . Derive an expression for PAC learning in such a way that the selected function will have low generalized error. (5)

b) Briefly Explain the procedure for the computation of the principal components of a given data.. (4)

13 a) Describe the forward selection and backward selection algorithm for implementing the subset selection procedure for dimensionality reduction (6)

b) Explain the concept of association rule analysis with its application (3)

PART C

Answer any two full questions, each carries 9 marks.

14 a) The following table shows the midterm and final exam grades obtained for students in a database course. (6)

X Midterm exam	Y Final exam
72	84
50	63
81	77
74	78
94	90
86	75
59	49
83	79
65	77
33	52
88	74
81	90

(i) Use the method of least squares to find an equation for the prediction of a student's final exam grade based on the student's midterm grade in the course.

(ii) Predict the final exam grade of a student who received an 86 on the midterm exam.

b) Explain Bootstrapping method for evaluating accuracy of a classifier. (3)

15 Identify the first splitting attribute for decision tree by using ID3 algorithm with the following dataset. (9)

Age	Competition	Type	Class (profit)
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	Up

16 a) Explain back propagation algorithm for a multilayer Perceptron. (5)

b) Explain the concept of Reduced Error pruning (4)

PART D

Answer any two full questions, each carries 12 marks.

17 a) Explain Learning problem in Hidden Markov model and how it can be solved. (6)

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

18 a) Find the three clusters after one epoch for the following eight examples using the k-means algorithm and Euclidean distance (6)

A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9). Suppose that the initial seeds (centers of each cluster) are A1, A4 and A7.

b) Show the final result of hierarchical clustering with single 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

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

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

****

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

****

Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Seventh Semester B.Tech Degree Examination (Regular and Supplementary), December 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 supervised learning and unsupervised learning with example.	(4)
2		Differentiate between overfitting and underfitting with suitable diagrams.	(4)
3		Write down the forward selection algorithm for implementing the subset selection procedure under dimensionality reduction.	(4)
4		Express the confusion matrix table for two class datasets.	(4)
5		Calculate the Gini index of the given number of class label with bird=2, fish=2 and reptile=1.	(4)
6		Describe any four types of activation functions used in Artificial Neural Networks.	(4)
7		Explain the basic problems of Hidden Markov Models.	(4)
8		What is the purpose of Kernel functions used in Support Vector Machines?	(4)
9		Distinguish between K-means Clustering and Hierarchical Clustering technique.	(4)
10		Write any four commonly used equations for measures of distance between data points of numeric data in Hierarchical Clustering.	(4)

PART B

Answer any two full questions, each carries 9 marks.

11	a)	Distinguish between classification and regression with example.	(5)
	b)	List any four applications of machine learning.	(4)
12	a)	What is Probably Approximately Correct (PAC) Learning?	(4)
	b)	Illustrate the concept of Vapnik-Chervonenkis (VC) dimension.	(5)
13	a)	Explain the procedure for performing a Principal Component Analysis (PCA) on a given data set.	(5)

b) What is reinforcement learning in machine learning and list any two applications? (4)

PART C

Answer any two full questions, each carries 9 marks.

14 Show the first splitting attribute for decision tree by using ID3 algorithm with the following data set. (9)

Day_No.	View	Temperature	Moisture	Breeze	Play Cricket
1	Bright	Hot	High	Weak	NO
2	Bright	Hot	High	Strong	NO
3	Cloudy	Hot	High	Weak	YES
4	Rain	Mild	High	Weak	YES
5	Rain	Cool	Normal	Weak	YES
6	Rain	Cool	Normal	Strong	NO
7	Cloudy	Cool	Normal	Strong	YES
8	Bright	Mild	High	Weak	NO
9	Bright	Cool	Normal	Weak	YES
10	Rain	Mild	Normal	Weak	YES
11	Bright	Mild	Normal	Strong	YES
12	Cloudy	Mild	High	Strong	YES
13	Cloudy	Hot	Normal	Weak	YES
14	Rain	Mild	High	Strong	NO

15 a) Illustrate Naïve Bayes algorithm for the dataset having *n* features. (5)

b) Explain the Receiver Operating Characteristics (ROC) Space in machine learning. (4)

16 a) Describe various types of regression models based on type of functions. (4)

b) Explain the issues involved in decision tree learning. (5)

PART D

Answer any two full questions, each carries 12 marks.

17 a) Describe the features of soft margin hyperplane and explain how it is computed. (6)

b) Explain the bagging and boosting methods used in learning algorithms. (6)

18 a) Write the algorithm for DIANA (**DI**visive**ANA**lysis) of hierarchical clustering technique. (6)

b) Illustrate the algorithm for K-means clustering of unsupervised learning. Write the disadvantages and applications of K-means clustering. (6)

19 a) Demonstrate an algorithm to find the SVM classifier and describe the mathematical formulation of the SVM problem. (6)

b) Construct the dendrogram using Complete Linkage Method for the given data points. (6)

	a	b	c	d	e
a	0				
b	8	0			
c	2	6	0		
d	5	4	8	0	
e	10	9	1	7	0

****