Reg No.:_____   Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
SEVENTH SEMESTER B.TECH DEGREE EXAMINATION(S), MAY 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	List out any four applications of machine learning.	(4)
2	Briefly describe the concept on model selection and generalisation.	(4)
3	Define VC dimension. Show that an axis aligned rectangle can shatter 4 points in 2 dimension.	(4)
4	What are the different methods for measuring classifier performance?	(4)
5	Explain the various methods to perform cross validation.	(4)
6	With suitable equations, explain any two types of activation functions used in neural networks.	(4)
7	What is the significance of optimal separating hyperplane in SVM?	(4)
8	Differentiate between bagging, boosting and voting.	(4)
9	Write down the major differences between K-means clustering and hierarchical clustering.	(4)
10	Explain the steps involved in expectation maximization algorithm.	(4)

PART B
Answer any two full questions, each carries 9 marks.

11	a)	Explain regression with an example.	(4)
	b)	Differentiate between supervised and unsupervised training. Explain with suitable examples.	(5)
12	a)	Explain feature selection and feature extraction method for dimensionality reduction.	(3)
	b)	Illustrate the two approaches used in subset selection.	(6)
13	a)	Define the terms Hypothesis space and Version space. Illustrate with an example.	(4)
	b)	Explain the concept of Probably Approximately Correct learning.	(5)

PART C
Answer any two full questions, each carries 9 marks.

14   The following data set contains factors that determine whether tennis is played   (9)

or not. Using Naive Bayes classifier, find the play prediction for the day

<Sunny, Cool, High, Strong>

DAY	OUTLOOK	TEMP	HUMIDITY	WIND	PLAY
Day 1	Sunny	Hot	High	Weak	NO
Day 2	Sunny	Hot	High	Strong	NO
Day 3	Overcast	Hot	High	Weak	YES
Day 4	Rain	Mild	High	Weak	YES
Day 5	Rain	Cool	Normal	Weak	YES
Day 6	Rain	Cool	Normal	Strong	NO
Day 7	Overcast	Cool	Normal	Strong	YES
Day 8	Sunny	Mild	High	Weak	NO
Day 9	Sunny	Cool	Normal	Weak	YES
Day 10	Rain	Mild	Normal	Weak	YES
Day 11	Sunny	Mild	Normal	Strong	YES
Day 12	Overcast	Mild	High	Strong	YES
Day 13	Overcast	Hot	Normal	Weak	YES
Day 14	Rain	Mild	High	Strong	NO

15   a)   For the following set of training samples, find which attribute can be chosen as   (6)

the root for decision tree classification

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

b)   What is a Perceptron? Explain the working of a perceptron with a neat diagram.   (3)

16   a)   A patient takes a lab test and the result comes back positive. It is known that the   (4)

test returns a correct positive result in only 98% of the cases and a correct negative result in only 97% of the cases. Furthermore, only 0.008 of the entire population has this disease.

1. What is the probability that this patient has cancer?

2. What is the probability that he does not have cancer?

3. What is the diagnosis?

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

PART D
Answer any two full questions, each carries 12 marks.

17 a) Explain how Support Vector Machine can be used for classification of linearly separable data.   (6)

b) Define Hidden Markov Model. What is meant by evaluation problem and how is this solved?   (6)

18 a) Use K Means clustering to cluster the following data into two groups. Assume cluster centroid are m1=2 and m2=4. The distance function used is Euclidean distance. { 2, 4, 10, 12, 3, 20, 30, 11, 25 }   (6)

b) Describe the concept on density based clustering and write the steps involved in DBSCAN algorithm.   (6)

19 a) Describe the random forest algorithm to improve classifier accuracy.   (6)

b) For the given data points, construct the dendrogram using Complete Linkage method.   (6)

	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

****