

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Seventh Semester B.Tech Degree Regular and Supplementary Examination December 2021 (2018 Scheme)

**Course Code: CS467****Course Name: MACHINE LEARNING**

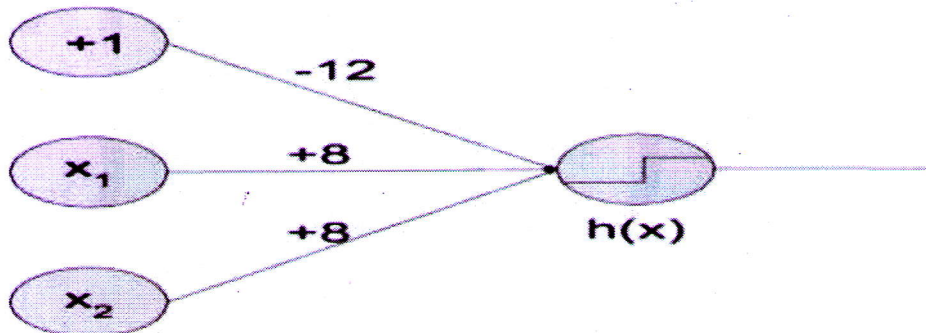
Max. Marks: 100

Duration: 3 Hours

PART A*Answer all questions, each carries 4 marks.*

Marks

- 1 Identify real life examples of classification problems. (4)
- 2 Describe the uses of dimensionality reduction in machine learning. (4)
- 3 Identify any four examples where machine learning is applicable. (4)
- 4 Discuss about ROC curve in machine learning and Relate ROC with Area Under Curve (AUC). (4)
- 5 Classify activation functions as linear and nonlinear. List any 2 examples for each. (4)
- 6 You are given the following neural networks which take two binary valued inputs $x_1, x_2 \in \{0, 1\}$ and the activation function is the threshold function ($h(x) = 1$ if $x > 0$; 0 otherwise). Find the logical functions does it compute? (4)



- 7 Identify the possible termination conditions in K-Means clustering. (4)
- 8 Suppose you are using RBF kernel in SVM with high Gamma value. Identify the significance. (4)
- 9 Analyze why clustering technique used in data compression. (4)
- 10 SVM is an example of a large margin classifier. Why? (4)

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

- 11 a) Explain the Hypothesis space, Version space, Most General hypothesis, Most specific hypothesis in the context of a classification problem. (4)
- b) Compare Supervised learning and Unsupervised learning with examples. (5)
- 12 a) Describe the backward selection algorithm for implementing the subset selection procedure for dimensionality reduction. (5)
- b) Illustrate with an example, overfitting and underfitting of data. (4)
- 13 a) Let X be the set of all real numbers. Describe a hypothesis for X for which the VC dimension is 2. Describe an example for which the VC dimension is 3. (4)
- b) Describe the steps involved in PCA algorithm. (5)

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 calculate the precision and recall scores for the search. (5)
- b) Explain the general MLE method for estimating parameters. (4)
- 15 a) Based on the following data to classify a particular species if its features are (Slow, Rarely, No). Use naive Bayes algorithm to classify the data. (9)

Sl. No.	Swim	Fly	Crawl	Class
1	Fast	No	No	Fish
2	Fast	No	Yes	Animal
3	Slow	No	No	Animal
4	Fast	No	No	Animal
5	No	Short	No	Bird
6	No	Short	No	Bird
7	No	Rarely	No	Animal
8	Slow	No	Yes	Animal
9	Slow	No	No	Fish
10	Slow	No	Yes	Fish
11	No	Long	No	Bird
12	Fast	No	No	Bird

- 16 a) Discuss Back propagation algorithm. (4)
- b) Discuss the issues in decision tree learning and Identify how to overcome it. (3)
- c) Describe about CART. (2)

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

- 17 a) Summarize the random forest algorithm. (5)
 b) Define a kernel function. Give an example. (3)
 c) Explain the kernel trick in context of support vector machines? How is it used to find a SVM classifier. (4)
- 18 a) Outline K-means algorithm. Apply k-means clustering algorithm to divide the following data into two clusters and also compute the representative data points for the clusters. Arbitrary cluster centers are (2,1) and (2,3). (12)

x_1	1	2	2	3	4	5
x_2	1	1	3	2	3	5

- 19 a) Discuss the pros and cons of HMM. (6)
 b) Summarize the DBSCAN algorithm for clustering. (6)
