APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
Scheme for Valuation/Answer Key
Scheme of evaluation (marks in brackets) and answers of problems/key
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	Any 4 applications of machine learning (1*4=4 marks)	(4)
2	Model selection—Explanation on Choosing among possible hypothesis- 2 marks	(4)
	Generalisation – Explanation on how well a model performs on new data- 2 marks	
3	Definition of VC Dimension- 2 marks, Explanation of 4 point separation- 2 marks	(4)
4	Any 3 terms (precision, recall, accuracy, sensitivity, ROC curve etc)- 3 marks	(4)
	Corresponding equations- 1 mark	
5	Explanation of any two methods of cross validation- 2 marks each	(4)
6	Any 2 types(Binary, Bipolar, Sigmoid, ramp etc) of activation function with equation(2*2=4 marks)	(4)
7	Optimal separating hyper plane- 2 marks, Its significance in SVM- 2marks	(4)
8	Explaining the concept of bagging, boosting and voting- 3 marks	(4)
	Distinguishing between the three- 1 mark	
9	K means- Start by choosing k points arbitrarily as the centers of the clusters, one for each cluster and then associate each of the given data points with the nearest centre -2 marks Hierarchical- Building a hierarchy of clusters in a given dataset. Clusters at each level of the hierarchy are created by merging clusters at the next lower level- 2 marks	(4)
10	E step- 2marks, M step- 2 marks	(4)

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

11	a)	Regression explanation – 3 marks, Example- 1 mark	(4)
	b)	Supervised with example -2.5 unsupervised with example – 2.5 marks	(5)
12	a)	Explanation of Feature selection- 1.5 marks, Feature Extraction- 1.5 marks	(3)

b)   Steps of Forward Selection and Backward Selection carries 3 marks   (6)
     each(3*2=6marks)

13 a)   Definition of each term carries 2 mark each   (4)

    b)   Step by step procedure of Probably Approximately Correct learning with   (5)
         equations carries 5 marks

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

14 a)   Step by step procedure- 7 marks , Final prediction- 2 marks   (9)

         Play prediction for the day **<Sunny, Cool, High, Strong> is NO**

15 a)   IG(a1)= 0.082, IG(a2)= 0. Therefore a1 is selected as root attribute   (6)
         Step by step procedure- 4 marks, Final selection of attribute- 2 marks

    b)   Definition- 1 mark, Perceptron working +diagram- 2 marks   (3)

16 a)   Final answer carries 1 mark each(1*3=3 marks) and steps carries 1 mark   (4)

         1. P(cancer|+)=P(+|cancer) x P(cancer)/P(+)=0.0078

         2. P(~cancer|+)= P(+|~cancer) x P(~cancer)/P(+)=**0.0298**

         3. More likely to have no cancer

    b)   Explaining any 5 issues carries 1 mark each   (5)
         ▪ Overfitting
              ▪ Reduced error pruning
              ▪ Rule post-pruning
         ▪ Extensions
              ▪ Continuous valued attributes
              ▪ Alternative measures for selecting attributes
              ▪ Handling training examples with missing attribute values
              ▪ Handling attributes with different costs
              ▪ Improving computational efficiency

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

17 a)   SVM basic concept explanation, suitable diagrams, equations etc carries 6 marks   (6)

    b)   Concept of HMM- 2 marks, evaluation problem -2 marks   (6)
         Forward algorithm- (2 marks)

18 a)   Total marks for iteration and assigning points to cluster- 5 marks   (6)
         Final cluster formation carries 1 mark

C1={2,4,10,12,3,11} and C2={20,30,25}

   b) Concept of density based clustering- 2 marks   (6)

       DBSCAN algorithm – 4 marks

19 a) Explaining the basic concept on ensemble/random forest- 2 marks   (6)

       Random forest algorithm- (4 marks)

   b) Step by step procedure to find the cluster- 5 marks   (6)

       Figure of dendrogram- 1 mark

****