Reg. No									

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

2nd SEMESTER M.TECH. (COMPUTER SCIENCE & ENGINEERING)
END SEMESTER EXAMINATIONS, April 2019
SUBJECT: MACHINE LEARNING (PROGRAME ELECTIVE-I) [CSE 5255]
REVISED CREDIT SYSTEM
(29/04/2019)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer **ALL** the questions.
- Missing data may be suitably assumed.

1A. Define the terms 'Machine Learning', and 'Concept Learning'. **2M**

1B. Explain Modified k-Nearest Neighbor (MkNN) Machine Learning Algorithm.: **4M**

1C. Consider the following training set to be applied on modified K nearest neighboring method of Classification/Learning. **4M**

X1 = (0.8, 0.8, 1), X2 = (1.0, 1.0, 1), X3 = (1.2, 0.8, 1)
X4 = (0.8, 1.2, 1), X5 = (1.2, 1.2, 1), X6 = (4.0, 3.0, 2)
X7 = (3.8, 2.8, 2), X8 = (4.2, 2.8, 2), X9 = (3.8, 3.2, 2)
X10 = (4.2, 3.2, 2), X11 = (4.4, 2.8, 2), X12 = (4.4, 3.2, 2)
X13 = (3.2, 0.4, 3), X14 = (3.2, 0.7, 3), X15 = (3.8, 0.5, 3)
X16 = (3.5, 1.0, 3), X17 = (4.0, 1.0, 3), X18 = (4.0, 0.7, 3)

Let the test sample P = (3.0, 2.0). For the five nearest points, the distances from P are to be determined, that is, d(P,X16) = ? ; d(P,X7) = ?; d(P,X14) = ?; d(P,X6) = ?; d(P,X17) = ?; For the given three class labels, find to which class the point P belongs to by explaining the steps of the modified K nearest neighbor method.

2A. Model the two components of the probability of error in two class classification process. Show that how diagrammatically the p (error) is least when the decision boundary at a point g1(x) = g_2(x). Where g1(x) and g2(x) are two discriminant functions. **4M**

2B. Derive the Naïve Bayes classification model using the Bayes classification model. **4M**

2C. Why we need parameter estimation in statistical classification systems? Mention any two methods for parameter estimation. **2M**

3A. What is clustering? Why is Clustering Important? **1M**

3B. Consider the two-dimensional data set of 16 labelled patterns given below. Let the two classes be labelled as ''X'' and ''O''. **5M**

X : (1, 1), (1, 2), (2,1), (2, 2), (1, 5), (1, 6), (2, 5), (2, 6)
O : (6, 1), (6, 2), (7, 1), (7, 2), (6, 6), (6, 7), (7, 6), (7, 7)
Let (2, 3) be a test sample which needs to be classified using the NN on the above 16 labelled patterns after forming clusters.

(i)Find its nearest neighbor to find the right class using the NN method.
(ii) Also, find the centroids using a clustering approach and show that clustering procedure can reduce the number of distance values to be computed from the test sample p to all the samples by using only centroids or cluster representatives. Also, comment on the extent of space optimization using only cluster representatives or centroids.

3C. Consider the data given in the following Table. Find the entropy and information gain using the following methods: **4M**
(i). Using entropy impurity
(ii) Using variance impurity.

X = *a* left branch	X = *b* right branch	Total	Class
40	0	40	1
10	20	30	2
10	20	30	3

4A. Write the Kohonen self-organizing algorithm. Write the meaning of competitive learning. **3M**

4B. Distinguish between MLP and RBF. **3M**

4C. Write and discuss the use of Backpropogation algorithm in multilayer neural networks. Why it is called Backpropogation? **4M**

5A. Briefly explain single layer perceptron algorithm and what are its limitations with regard to learning of Boolean functions? **5M**

5B. Discuss the working of a radial basis function network and compare radial basis function with Kohonen self-organizing algorithm **3M**

5C. What is bootstrap aggregation? What is its role in ensemble classifiers? **2M**