Name:

Enrolment No:

UPES
UNIVERSITY WITH A PURPOSE

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2019

Course: B.Tech CSE+AI/ML **Semester:** III
Program: Machine Learning **Time** : 03 hrs.
Course Code: CSAI2001 **Max. Marks:** 100

Instructions:

SECTION A

S. No.		Marks	CO
Q 1	Define Machine Learning. Write down five application of it.	4	CO1
Q 2	List down four application of Linear Regression Model with their dependent and independent variable.	4	CO1
Q 3	Give an example of how specific clustering methods can be integrated, for example, where one clustering algorithm is used as a preprocessing step for another.	4	CO3
Q 4	<table><tr><td>Time Point</td><td>Infosys</td><td>TCS</td></tr><tr><td>Jan 2019</td><td>6</td><td>20</td></tr><tr><td>Feb 2019</td><td>5</td><td>10</td></tr><tr><td>March 2019</td><td>4</td><td>14</td></tr><tr><td>April 2019</td><td>3</td><td>5</td></tr><tr><td>May 2019</td><td>2</td><td>5</td></tr></table> It is given the average stock price of Infosys and TCS for five consecutive months. Find it either the stock price are independent to each other or not.	4	CO1
Q 5	Differentiate between Similarity Metrics and Term Weighting.	4	CO4

SECTION B

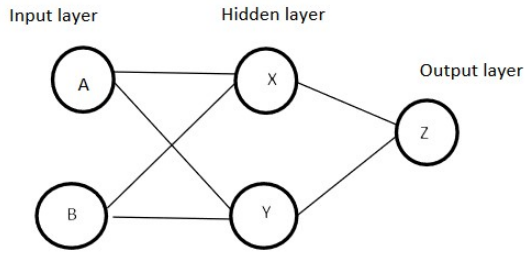
Q 6	Explain and discuss the architecture of information retrieval system of Google Search Engine.	10	CO4
Q 7	Discuss and derive the mathematical proof of linear regression model.	10	CO1
Q 8	Transactional data of AllElectronics <table><tr><th>TID</th><th>List of item_IDs</th></tr><tr><td>T100</td><td>I1, I2, I5</td></tr><tr><td>T200</td><td>I2, I4</td></tr><tr><td>T300</td><td>I2, I3</td></tr><tr><td>T400</td><td>I1, I2, I4</td></tr><tr><td>T500</td><td>I1, I3</td></tr><tr><td>T600</td><td>I2, I3</td></tr><tr><td>T700</td><td>I1, I3</td></tr><tr><td>T800</td><td>I1, I2, I3, I5</td></tr><tr><td>T900</td><td>I1, I2, I3</td></tr></table>	10	CO2

<table>
<tr><td></td><td>Evaluate the most frequent data item set of 3 items using FP-Growth algorithm for the above AllElectronics data set.</td><td></td><td></td></tr>
<tr><td>Q 9</td><td>Discuss confusion matrix. Explain the basis of Model Evaluation and selection. Suppose there are two models M1 and M2.
For M1: TP=6954, FN=46, FP=412 and TN=2588
For M2: TP=6800, FN=134, FP=566 and TN=2500
Calculate Accuracy, Recall, Specificity, Sensitivity and Z-Score. Among M1 and M2 which one is more preferable model?
OR
Explain KNN algorithm. Why it is also called Lazy Learner? What are the points to be subjected when choosing the value of k? For the below problem predict for the class of Davis using KNN and assume the value of k=3.
<table><tr><th>Customer</th><th>Age</th><th>Income (K)</th><th>No. of cards</th><th>Response</th></tr><tr><td>John</td><td>35</td><td>35</td><td>3</td><td>Yes</td></tr><tr><td>Rachel</td><td>22</td><td>50</td><td>2</td><td>No</td></tr><tr><td>Ruth</td><td>63</td><td>200</td><td>1</td><td>No</td></tr><tr><td>Tom</td><td>59</td><td>170</td><td>1</td><td>No</td></tr><tr><td>Neil</td><td>25</td><td>40</td><td>4</td><td>Yes</td></tr><tr><td>David</td><td>37</td><td>50</td><td>2</td><td>?</td></tr></table></td><td>10</td><td>CO2</td></tr>
<tr><td colspan="4">SECTION-C</td></tr>
<tr><td>Q 10</td><td>Suppose that the data mining task is to cluster the following eight points with (x, y) representing location into three clusters:
A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9):
The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively.
a) Write down k-means algorithm
b) Apply k-means algorithm for the three cluster centers after the first round execution
c) Find the final three clusters</td><td>8+6+6 =20</td><td>CO3</td></tr>
<tr><td>Q 11</td><td>“The support vector machine is highly accurate classification method”, justify the statement. SVM classifier suffers from slow processing when training with a large data set, why? How we can solve this problem and make the SVM scalable. Categorize the types of hyperplane, if any. Explain with the concept of projection</td><td>20</td><td>CO2</td></tr>
</table>

(orthonormal).

OR

.



Input		Output
A	B	Z
0	0	0
0	1	1
1	0	1
1	1	1

Learning rate=0.35

Biases are ϴx = ϴy = ϴz = 0. Neural Network of above diagram has two nodes (A,B) in the input layer, two nodes in the hidden layer (X,Y)and one node in the output layer (Z). The values given to weights are taken randomly and will be changed during back propagation iterations. Initial weights of the top input nodes taken at random are 0.4, 0.1 .Weights of bottom input node are 0.8 and 0.6. Weights of top hidden node is 0.3 and that of bottom hidden node is 0.9. Assume the number of iterations are two.