

**BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI  
(END SEMESTER EXAMINATION MO 2022)**

**CLASS: B.Tech.  
BRANCH: ECE/CHEM/CIVIL/MECH/PROD**

**SEMESTER: VII  
SESSION: MO2022**

**SUBJECT: IT340 MACHINE LEARNING**

**TIME: 3 Hours**

**FULL MARKS: 50**

**INSTRUCTIONS:**

1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Tables/Data handbook/Graph paper etc., if applicable, will be supplied to the candidates

- Q.1(a) Define Data matrix with proper representation. [CO1][BL2][2]  
 Q.1(b) Compare and contrast supervised and unsupervised learning with suitable example. [CO2][BL4][3]  
 Q.1(c) Illustrate the method to predict the outcome of a sample with a single attribute. [CO1, CO3][BL3][5]  
 Also discuss how cost function is related to solve such problem?
- Q.2(a) Explain pre-pruning and post-pruning process. [CO2][BL2][2]  
 Q.2(b) Illustrate the need of data transformation and what are the different ways to transform the values of a feature? [CO1][BL3][3]  
 Q.2(c) Consider the following training data set having three categorical attributes Homeowner, Family Status, Annual Income. Loan borrowers who defaulted on their payments are classified as 'Yes' while those who repaid their loans are classified as 'No'. Compute the value of Gain on Gini if the feature 'Family Status' is considered as a splitting attribute. [CO4][BL6][5]

TID	Homeowner	Family Status	Annual Income	Defaulted Borrower (Class)
1	Yes	Single	125K	No
2	No	Nuclear	100K	No
3	No	Single	70K	No
4	Yes	Nuclear	120K	No
5	No	Joint	95K	Yes
6	No	Nuclear	60K	No
7	Yes	Joint	220K	No
8	No	Single	85K	Yes
9	No	Nuclear	75K	No
10	No	Single	90K	Yes

- Q.3(a) Why is activation function used in artificial neurons network? [CO2][BL2][2]  
 Q.3(b) Design a neural network model with two input neurons with bias and two output neurons. The initial weight vectors are  $[v_{11}, v_{21}, v_{01}] = [0.2, 0.1, 0.3]$  and  $[v_{12}, v_{22}, v_{02}] = [-0.1, 0.4, 0.5]$ . Calculate the outputs of both the output neurons for input vector  $[1, 1]$ . Use Bipolar sigmoidal function as activation function. [CO4, CO5][BL4][3]  
 Q.3(c) Design the three phases of Back Propagation Neural Network model. Elucidate the error propagation phase in the backward direction. [CO3][BL4][5]
- Q.4(a) Compute the cosine similarity between the two document vectors X  $(3, 2, 0, 5, 0, 0, 0, 2, 0, 0)$  and Y  $(1, 0, 0, 0, 0, 0, 0, 1, 0, 2)$ . [CO1][BL3][2]  
 Q.4(b) Mention the limitations of k-means clustering algorithms. State at least two stopping conditions of partitional clustering method. [CO2, CO3][BL1, 2][3]  
 Q.4(c) Contrast and compare Divisive and Agglomerative hierarchical clustering algorithm. Also illustrate the methods using example and dendrogram. [CO3][BL3, 4][5]
- Q.5(a) Define classification accuracy. [CO2][BL1][2]  
 Q.5(b) How bias and variance are related to overfitting and underfitting, explain? [CO2][BL2][3]  
 Q.5(c) Demonstrate the Random Forest model with suitable diagram. [CO4][BL3][5]