

Indian Institute of Engineering Science and Technology, Shibpur

B.Tech (CST) 4 th Semester Mid Semester Examinations, April 2021

Subject: Introduction to Data Science

Code: CS2205

Full Marks: 30

Time: 45 minutes

Answer any three questions

Give your answer as much as possible in terms of logic, diagram and mathematical formulation

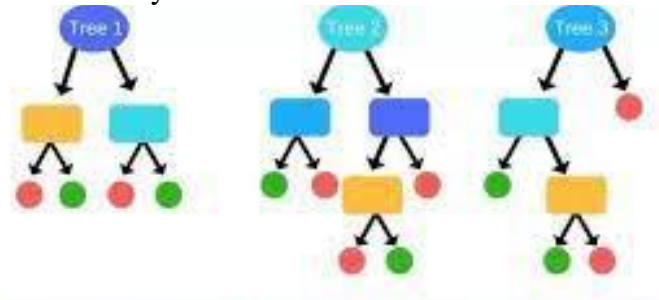
1. (a) Why do we need discretization of data? 2.5
(b) Describe Equal-Width Discretization and Equal-Frequency Discretization Methods. 2.5+2.5
(c) What is the main challenge of discretization? 2.5
2. (a) Write the significance of p-value in testing of hypothesis. 5
(b) Define Confidence Interval in parameter estimation of dataset? 3
(c) What do you mean by Sample and Population of a dataset? 2
3. (a) What do you mean by overfitting? 3
(b) Write equation of Loss function with L2 norm in regularization. 3
(c) Explain why Deep Learning based feature extraction approach is effective for predictive accuracy. 4
4. (a) construct a Decision tree using the following table. 5

| Home Owner | Marital Status | Annual Income | Defaulted Borrower |
|------------|----------------|---------------|--------------------|
| Y | Single | 125 | N |
| N | Married | 100 | N |
| N | Single | 70 | N |
| Y | Married | 120 | N |
| N | Divorced | 95 | Y |
| N | Married | 60 | N |
| Y | Divorced | 220 | N |
| N | Single | 85 | Y |
| N | Married | 75 | N |
| N | Single | 90 | Y |

- (b) How do you assign the class label of a new record with the following information: Home Owner = Y, Marital Status = Married, Annual Income = 150? 3
- (c) What do you mean by Pruning? 2

5. (a) What are the drawbacks of Decision tree based classifier? 4

- (b) Red and Green represent two different class labels of the decision trees. If you build a Random Forest Classifier using the decision trees, what will be the predicted class label and why? 3



(c) How do you choose the value of K in K-NN algorithm? 3

6. (a) How do you decide the degree of polynomial for fitting a good predictor model? 3
- (b) Describe the principle of gradient descent algorithm for minimizing the error? 4
- (c) Write Basic difference between collaborative and content based approaches to build a recommender system. 3

Jaya Sil