Nirma University

Institute of Technology

Semester End Examination (IR), May - 2019 B. Tech. in Computer Engineering */* Information Technology, Semester-VI

CE623 Machine Learning Roll*/*

Supervisor's initial Exam No

with date

Time: 3 Hours

Max Marks: 100

Instructions:

1. Attempt all questions. 2. Figures to the right indicate full marks. *3*. Assume suitable data wherever necessary and specify them. 4. Draw neat sketches wherever necessary. 5. CO\_, BO \_ has been mentioned against each question to map it as per

Course Learning Objective and Bloom's Taxonomy.

**Section - I**

[16] [16]

Yes

**Q.1 Answer the following questions.** (A) Generate and draw C4.5 decision tree for the following training data. CO3,

**Size**

**Color**

**Shape**

**Class** BL3

Medium

Blue

Brick

Yes Small

Red

Sphere Large

Green

Pillar

Yes Large

Green

Sphere

Yes Small

Red

Wedge

No Large

Red

Wedge Large

Red

Pillar

No

No

**OR**

**(A) Answer the following questions.** (i) Compare ID3, C4.5 and CART. C01,

[4]

BL2

[12]

(ii) Clark Coffee operates a chain of five luxury coffee shops in Cheshire. It is CO3, looking at two options to increase revenues across the chain. The BL3

estimated impact of the two options on sales (and their probabilities) are shown below as are the associated costs of each option.

**Launch Loyalty Cut Prices**

**Card** Cost of Option

Rs. 500,000 Rs. 300,000 Probability of High Sales

0.6

0.8 Probability of Low Sales

0.4

0.2 Result of High Sales

Rs. 1,000,000 Rs. 800,000

Result of Low Sales

Rs. 750,000

Rs. 500,000

Generate and draw decision tree and help the owner of the coffee chain to take the best decision.

Page 1 of 3

CE623: Machine Learning

2.2 **Answer the following:**

**[16]** (A) Consider the confusion matrix of win/loss prediction of a cricket match [08] CO2, problem given below. BL5

Actual Win

Actual Loss Predicted Win

85 Predicted Loss

9 Compute model accuracy, precision, recall and fl-measure. **(B)** What is feature normalization? Why is it required? Explain any two [08] CO2- feature normalization methods with a suitable example. BL4

D.

[18]

Q-3. Assume text collection shown in the following table. CO3,

**Document Text of Document** BL3

**Class**

**ID**

great game

Sports election over

Not Sports clean match

Sports clean but forgettable game Sports

close election

Not Sports Last column shows the class of the text document. Build naïve Bayes text classifier considering the text collection shown in the table as the training set. Construct bag-of-words index (In bag-of-words index, rows correspond to text documents while columns relate to words in the text document. Cell value indicates number of times the corresponding word appears in the corresponding document) and fit Multinomial distribution to the data. Classify the document "close forgettable match” using the built classifier. Build the classifier considering Laplacian correction.

**[20]** [08]

**Section - II Q-4. Answer the following *(*A)** Suppose you have given the following data where x and y are the 2 CO3, input variables and Class is the dependent variable. Suppose, you BL3

want to predict the class of new data point x=1 and y=1 using Euclidean distance using 3-NN and 5-NN. In which class this data point belongs to?

**Class** -1

+

+ 11 + +

+

[12]

(B) Maximize the fitness function f(x) = x2 where x varies between 1 and 31. CO3, Assuming that decision variable of the problem is coded as some finite BL3 length string, explain how genetic algorithm works to show improvement

Page 2 of 3

CE623: Machine Learning

in the next generation.

String No 1 = [01101], String No 2 = [11000), String No 3 = (01000), String No 4 = [10011]

**OR** (B) How can Travelling Salesman Problem be solved using genetic algorithm? CO3, Discuss suitable encoding scheme, selection method, crossover and

mutation operators to solve this problem using genetic algorithm

[12]

[16] [8]

Q-5**. Answer the following** (A) Critically Compare Supervised, Unsupervised, Partially Supervised and CO2, Reinforcement Learning with suitable example and application for each. BL4 **(B)** Consider the data given in the following table. CO3,

Age in Years Weight in KGs BL3

40

*181*

12 19

60

55

[14]

Q.6 CO3, BL3

40

80

85 23

70 13

7*0* 15

75 14

66 58 T

7*9* Divide these data points in three clusters using k-means algorithm. Assume (12, 40), (19, 60) and (40, 80) as initial centroid of three clusters. Use Euclidean Distance as the distance measure. Perform 1 iteration. Consider the data given in the following table.

**Number of Shopping Retail Sales (in billions of**

**Centers (x)**

$) (y) 630

15.5 370

7.5 616

13.9 700

18.7 430

8.2 568

13.2 1200

23.0 297*6*

*8*7.3 Fit a first order regression model to the data (y = *Bo + B*1x). Estimate parameters of the model (*Bo and B*.) through one iteration of gradient descent assuming learning rate = 0.1. Assume initial value of *B*o and *B*i to be -5 and 0.025 respectively. How can you approach this problem using normal equation method? Just mention steps for normal equation method.

Page 3 of 3