

Nirma University

Institute of Technology

Semester End Examination (IR), May - 2019

B. Tech. in Computer Engineering / Information Technology, Semester-VI
CE623 Machine Learning

Roll/
Exam No

Supervisor's initial
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

Q.1 Answer the following questions.

(A) Generate and draw C4.5 decision tree for the following training data.

CO3,
BL3

[16]

[16]

Size	Color	Shape	Class
Medium	Blue	Brick	Yes
Small	Red	Sphere	Yes
Large	Green	Pillar	Yes
Large	Green	Sphere	Yes
Small	Red	Wedge	No
Large	Red	Wedge	No
Large	Red	Pillar	No

OR

(A) Answer the following questions.

(i) Compare ID3, C4.5 and CART.

CO1,
BL2

[4]

(ii) Clark Coffee operates a chain of five luxury coffee shops in Cheshire. It is looking at two options to increase revenues across the chain. The estimated impact of the two options on sales (and their probabilities) are shown below as are the associated costs of each option.

[12]

	Launch Loyalty Card	Cut Prices
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.

Q.2 Answer the following:

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

	Actual Win	Actual Loss
Predicted Win	85	4
Predicted Loss	2	9

- (B) Compute model accuracy, precision, recall and f1-measure. [08]
What is feature normalization? Why is it required? Explain any two feature normalization methods with a suitable example.

- Q-3. Assume text collection shown in the following table. [18]
CO3, BL3

Document ID	Text of Document	Class
1	great game	Sports
2	election over	Not Sports
3	clean match	Sports
4	clean but forgettable game	Sports
5	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.

Section - II**Q-4. Answer the following**

- (A) Suppose you have given the following data where x and y are the 2 input variables and Class is the dependent variable. Suppose, you 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? [20]
[08]
CO3, BL3

x	y	Class
-1	1	-
0	1	+
0	2	-
1	-1	-
1	0	+
1	2	+
2	2	-
2	3	+

- (B) Maximize the fitness function $f(x) = x^2$ where x varies between 1 and 31. [12]
CO3, BL3 Assuming that decision variable of the problem is coded as some finite length string, explain how genetic algorithm works to show improvement

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? [12]
CO3, Discuss suitable encoding scheme, selection method, crossover and
BL3 mutation operators to solve this problem using genetic algorithm

Q-5. Answer the following

- (A) Critically Compare Supervised, Unsupervised, Partially Supervised and [16]
CO2, Reinforcement Learning with suitable example and application for each. [8]
BL4

- (B) Consider the data given in the following table. [8]
CO3,

BL3

Age in Years	Weight in KGs
12	40
19	60
40	80
55	85
23	70
13	70
15	75
14	66
58	79

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.

Q.6

CO3,
BL3

Consider the data given in the following table.

[14]

Number of Shopping Centers (x)	Retail Sales (in billions of \$) (y)
630	15.5
370	7.5
616	13.9
700	18.7
430	8.2
568	13.2
1200	23.0
2976	87.3

Fit a first order regression model to the data ($y = \beta_0 + \beta_1 x$). Estimate parameters of the model (β_0 and β_1) through one iteration of gradient descent assuming learning rate = 0.1. Assume initial value of β_0 and β_1 to be -5 and 0.025 respectively. How can you approach this problem using normal equation method? Just mention steps for normal equation method.