



Register Number:

DATE:

ST. JOSEPH'S COLLEGE (AUTONOMOUS), BENGALURU-27
M.Sc. BIG DATA ANALYTICS – II SEMESTER
SEMESTER EXAMINATION: APRIL 2019
BDA 2318 - MACHINE LEARNING

TIME 2.5 HOURS

MAXIMUM MARKS 70

This Question Paper Contains TWO Printed Paper And ONE Part

ANSWER ANY SEVEN QUESTIONS

7 X10 = 70

1. Classify the following algorithms in terms of their uses as supervised or unsupervised learning. State one end application of each of these algorithms, also.
 - a. Linear Regression
 - b. Logistics Regression
 - c. KNN
 - d. ANN
 - e. SVM

[5×2=10]

2. a. Define and sketch a logistic function.[properly indicating its range]
b. Describe the procedure of estimating the logistic regression coefficients.

[3+7=10]

3. Suppose we collect data for a group of students in a statistics class with variables X_1 = hours studied, X_2 = undergrad GPA, and Y = receive an A. We fit a logistic regression and produce estimated coefficient, $\hat{\beta}_0 = 6$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 1$.
 - a. Estimate the probability that a student who studies for 40 h and has an undergrad GPA of 3.5 gets an A in the class.
 - b. How many hours would the student in part (a) need to study to have a 50% chance of getting an A in the class?

[5+5=10]

4. a. Explain the concept of Single Neuron.
b. Discuss the usage of Naive Bayes classifier in email spam filtering.

[5+5=10]

5. a. Explain the phenomenon of bias-variance trade-off in case of modeling.
b. Given a training data-set consisting of n observations, briefly describe one method of cross-validation to ensure robust predictions on the test data-set.

[5+5=10]

6. Calculate the cost function $\hat{\beta} = [\beta_0 \ \beta_1]'$ for a linear regression model $Y_j = \beta_0 + \beta_1 z$ fit to data (5 pairs of observations)

| | | | | | |
|---|---|---|---|---|---|
| z | 0 | 1 | 2 | 3 | 4 |
| Y | 1 | 4 | 3 | 8 | 9 |

[10]

7. a. How is a multi-layer neural network able to form nonlinear decision boundaries?
 b. What is the difference between forward and backward propagation in a neural network?

[5+5=10]

8. We are given $n = 7$ observations in $p = 2$ dimensions. For each observation, there is an associated class label.

| Observation | X ₁ | X ₂ | Y |
|-------------|----------------|----------------|------|
| 1 | 3 | 4 | Red |
| 2 | 2 | 2 | Red |
| 3 | 4 | 4 | Red |
| 4 | 1 | 4 | Red |
| 5 | 2 | 1 | Blue |
| 6 | 4 | 3 | Blue |
| 7 | 4 | 1 | Blue |

Sketch the observations and indicate the support vectors for the maximal margin classifier.

[10]

9. Write short notes on the following
 a. Batch gradient descent algorithm.
 b. Support vector machine.

[5+5=10]