

Registration No:

--	--	--	--	--	--	--	--	--	--

Total Number of Pages : 2

M.TECH

M.TECH 2<sup>ND</sup> SEMESTER (AR 17) SUPPLEMENTARY EXAMINATIONS, APRIL/MAY 2019  
DATA ANALYTICS

Branch: CSE, Subject Code:MCSPC2010

Time: 3 Hours

Max Marks : 70

**PART-A****(10 X 2=20 MARKS)****1. Answer the following questions.**

- Define SSE?
- Why subset selection is necessary, explain?
- What is the importance of training set in supervised learning?
- What is over fitting?
- Give an example of a regression problem specifying the predictors and the response?
- What is co-linearity?
- What is forward stepwise selection?
- In support vector machine method for classification, how are the support vectors defined?
- How are the weights in back-propagation algorithm determined?
- What is prescriptive analysis?

**PART-B****(5 X 10=50 MARKS)****Answer any five questions from the following.**

- Q.2. a) State the regression problem. Obtain the expression for the residual sum of squares (RSS) [ 5 ]  
in vector form for a simple linear regressor.
- b) What is subset selection problem? Why is subset selection necessary? Write the best [ 5 ]  
subset selection algorithm and state its limitations.
- Q.3. a) Write down the algorithm of linear discriminant analysis method for classification for a [ 5 ]  
single predictor case when the density function of each class is Gaussian.
- b) Explain Rosenblatt's perceptron with a figure and write the Perceptron learning algorithm. [ 5 ]
- Q.4. a) Analyze how logistic regression can be used for classification. [ 5 ]
- b) The table below provides a training data set containing six observations, three predictors, [ 5 ]  
and one qualitative response variable.
- | Obs. | X1 | X2 | X3 | Y     |
|------|----|----|----|-------|
| 1    | 0  | 2  | 0  | Red   |
| 2    | 2  | 0  | 0  | Red   |
| 3    | 0  | 1  | 3  | Red   |
| 4    | 0  | 1  | 2  | Green |
| 5    | -1 | 0  | 1  | Green |
| 6    | 1  | 1  | 1  | Red   |
- Predict the value of Y when  $X_1=0$ ,  $X_2=-1$ , and  $X_3=1$  using K-NN classifier with  $K=3$ .
- Q.5. a) Draw a diagram of a single hidden layer feed forward network. Outline the issues related [ 5 ]  
to training a multi layer perceptron neural network.
- b) Write the Principal component analysis algorithm. [ 5 ]

- Q.6. a) Giving an example, show how random forest algorithm is used to predict the response of a new pattern in a regression problem. [ 5 ]  
b) What is clustering? Explain, k-means clustering algorithm. [ 5 ]
- Q.7. a) A database has four transactions. Let minimum support  $s$  be 60% and the minimum confidence  $c$  be 80% T [ 5 ]
- | ID | item          |
|----|---------------|
| I1 | { K ,A,D,B }  |
| I2 | { D,A,C,E,B } |
| I3 | { C,A,B,E }   |
| I4 | { B,A,D }     |
- Find all frequent item sets using Apriori algorithm.
- b) Outline the issues and challenges of big data analytics. [ 5 ]
- Q.8. Write short notes on [ 5 ]
- a) Supervised Vs. Unsupervised learning [ 5 ]  
b) Market basket analysis

==0==