M18002005

Registration No: | | | | | | | | | | |

Total Number of Pages : 02 **M.TECH**

M.TECH 2ND SEMESTER REGULAR EXAMINATIONS, MAY 2018
DATA ANALYTICS
Branch: CS, Subject Code:MCSPC2010
Time: 3 Hours
Max Marks : 70

PART-A (10 X 2=20 MARKS)

1. Answer the following questions.

a) Define SSE? CO-1
b) Why subset selection is necessary, explain? CO-2
c) What is the importance of training set in supervised learning? CO-1
d) What is over fitting? CO-1
e) Give an example of a regression problem specifying the predictors and the response? CO-2
f) What is co-linearity? CO-1
g) What is forward stepwise selection? CO-1
h) In support vector machine method for classification, how are the support vectors defined? CO-2
i) How are the weights in back-propagation algorithm determined? CO-2
j) What is prescriptive analysis? CO-1

PART-B (5 X 10=50 MARKS)

Answer any five questions from the following.

2. a) State the regression problem. Obtain the expression for the residual sum of squares (RSS) in vector form for a simple linear regressor. [5] CO-1

b) What is subset selection problem? Why is subset selection necessary? Write the best subset selection algorithm and state its limitations. [5] CO-1

3. a) Write down the algorithm of linear discriminant analysis method for classification for a single predictor case when the density function of each class is Gaussian. [5] CO-1

b) Explain Rosenblatt's perceptron with a figure and write the Perceptron learning algorithm. [5] CO-1

4. a) Analyze how logistic regression can be used for classification. [5] CO-3

b) The table below provides a training data set containing six observations, three predictors, and one qualitative response variable.

Obs.	X_1	X_2	X_3	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. [5] CO-2

5. a) Draw a diagram of a single hidden layer feed forward network. Outline the issues related to training a multi layer perceptron neural network. [5] CO-2

   b) Write the Principal component analysis algorithm. [5] CO-1

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] CO-3

   b) What is clustering? Explain, k-means clustering algorithm. [5] CO-1

7. a) A database has four transactions. Let minimum support s be 60% and the minimum confidence c be 80% T

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. [5] CO-2

   b) Outline the issues and challenges of big data analytics. [5] CO-3

8. Write short notes on

a) Supervised Vs. Unsupervised learning [5]CO-1
b) Market basket analysis [5]CO-1

==0==