

Registration No :

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

Total Number of Pages : 02

B.Tech.
PCS4D001

4th Semester Regular / Back Examination 2017-18

DATA ANALYTICS

BRANCH : CSE

Time : 3 Hours

Max Marks : 100

Q.CODE : C1147

Answer Part-A which is compulsory and any four from Part-B.

The figures in the right hand margin indicate marks.

Answer all parts of a question at a place.

Part – A (Answer all the questions)

Q1 Answer the following questions: multiple type or dash fill up type: (2 x 10)

- Data Analysis is a subcomponent of Data Analytics. Right/Wrong
- “A person interested to buy a laptop may also interested for Backpack” belongs to _____. Association rule learning/Clustering/supervised learning
- RSS stands for _____.
- Rosenblatt Perceptron is an algorithm for _____. Supervised Learning/Unsupervised learning/ Multiple Learning.
- Logistic regression is used for _____. (Classification problem/regression problem/ clustering problem/non of these)
- The day by day analysis of the effect of drug upon a mad person belongs to _____ learning.
- Decision trees where the target variable take discrete values are called regression trees. Right/Wrong
- Bootstrap method is generally used when data set is small. Right/Wrong
- Boosting is used for _____. Reducing bias and variance/Incrementing bias and variance/synchronization of network
- T-tests can be used to compare two means to assess whether they are from the same population. Right/Wrong

Q2 Answer the following questions: Short answer type: (2 x 10)

- What is Data analytics ? Explain.
- Define variance.
- Explain ridge regression ? Why it is essential ?
- How the weights of the inputs for an ANN can be determined ? Explain.
- What do you mean by multicollinearity ? explain.
- Explain few challenges of big data analytics.
- What is the role of kernel machine ? Explain.
- In k-means algorithm how the centers of the clusters initialized ?
- What is the necessity of back propagation in fitting a neural network.
- What do you mean by bagging ?

Part – B (Answer any four questions)

Q3 a) What is the difference between Supervised and Unsupervised Learning? Explain briefly. (10)

- What is clustering ? Give an example where clustering can be used and write the k-means clustering algorithm. (5)

- Q4 a)** Explain the role of SVM in machine learning. When it is essential ? Explain the linear classifier in relation to SVM. **(10)**
- b)** What is the objective of Bootstrap method ? How it is different from cross validation, Explain. **(5)**

- Q5 a)** What is Lift ? How it is related to Association rule ? Calculate the lift for **(10)**

the following data set.

(i) Lift ($A \rightarrow 0$)

(ii) Lift ($B \rightarrow 1$)

Antecedent	Consequent
A	0
A	0
A	1
A	0
B	1
B	0
B	1

- b)** "The Higher the Bias the Lower is the variance and vice versa" Explain the statement with a suitable diagram taking X as the independent variable and Y is the dependent variable. **(5)**

- Q6 a)** Define "ensemble learning". How it is related to Random forest ? Give an example with at least three learning parameters to explain an analysis of a medical patient. **(10)**

- b)** What is subset selection and why it is essential ? Write the best subset selection algorithm and state its limitation. **(5)**

- Q7 a)** Describe the McNemar's test and k-fold cross validated paired t-test to compare whether two given learning algorithms construct classifiers that have the same expected error rate. **(10)**

- b)** The following table provides a training data set with six observations, three predictors and one dependent variable **(5)**

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 $K=3$.

- Q8 a)** Consider two one-dimensional, Gaussian-distributed classes and that have a common variance equal to 1. Their mean values are **(10)**

$$\mu_1 = -10$$

$$\mu_2 = +10$$

These two classes are essentially linearly separable. Design a classifier that separates these two classes.

- b)** Explain the various components of a Neural Network. **(5)**

- Q9 a)** Write a short note on following data sets **(10)**
(i) California housing (ii) NewZealand fish

- b)** What is Bigdata ? Describe the main features of a data analytical system. **(5)**