TYBSc CS - Sem VI

3/5/19   (2 ½ Hours)   [Total Marks: 75]

N.B. : 1) All questions are **compulsory**.
2) **Figures** to the right **indicate marks**.
3) **Illustrations**, in-depth **answers** and **diagrams** will be **appreciated**.
4) Mixing of **sub-questions** is **not allowed**.

Q.1   Attempt All(Each of 5marks)   (15M)

(a)   Choose the correct alternative.

i.   Ordinal data is a type of _____.
   a) Measurement   b) Catagorical
   c)Discrete   d)Continuous

ii.   _____ measures asymmetry about the mean of the probability distribution of a random variable.
   a)skewness   b)covariance   c)variance   d) Kurtosis

iii.   _____ shows all individual data points.
   a)Box-plot   b) scatter plot   c)line plot   d)pie chart

iv.   In _____, we start with all the features and removes the least significant feature at each iteration.
   a)Forward elimination   b)Backward elimination
   c) Recursive Feature elimination   d) None of the above

v.   XPath specification has _____ types of nodes
   a)Four   b)Five   c)Six   d)Seven

(b)   Fill in the blanks. Use following pool to answer question.

Pool (XPath, Mode,sample, Tidy Data, Principal Components Analysis)

i.   ___3___ is the subset of the population.

ii.   Five number summary does not include ___2___ value.

iii.   ___5___ refers to the process by which principal components are computed.

iv.   ___1___ is a query language that is used for traversing through an XML document.

v.   ___4___ data has each variable saved in its own column.

(c)   Answer in 1-2 sentences.

i.   Define Mode.

ii.   What is Data Aggregation?

iii.   What is Margin in SVM?

iv.   What is feature extraction?

v.   Define BIC.

Q.2 Attempt the following(Any Three)   (15M)

   a) Explain data along with its types.

   b) Describe the types of observational methods used in data collection.

   c) What is EDA? Explain any two types of visualization.

   d) Compare structured and unstructured data.

   e) What is kurtosis? Explain its types.

   f) What is data wrangling? Explain with any one package.

Q. 3 Attempt the following (Any THREE) (Each of 5Marks)   (15M)

   (a) What is NoSQL? Briefly explain its types.

   (b) What is JSON? How to read JSON file in R with an example?

   (c) Write a note on HBase.

   (d) What is collection in MongoDB? Give an example to create collection in MongoDB.

   (e) What are various ways to perform Web Crawling?

   (f) Write note on XPath.

Q. 4 Attempt the following (Any THREE) (Each of 5Marks)   (15)

   (a) Write short note on Time Series Analysis.

   (b) Explain Multiple Linear Regression.

   (c) What is Decision tree? What are its advantages?

   (d) Write note on Bias/Variance Tradeoff?

   (e) Explain hierarchical clustering.

   (f) Write a short note on Ensemble Methods.

Q. 5 Attempt the following (Any THREE) (Each of 5Marks)   (15)

   (a) Distinguish between Structured and Unstructured Data.

   (b) Explain Map Reduce Architecture.

   (c) Write note on Support vector Machines.

   (d) Explain Bayesian Information Criterion.

   (e) What is MongoDB? State its features.

****