

Roll No

CS-7005 (2) (CBGS)

B.E. VII Semester

Examination, November 2019

Choice Based Grading System (CBGS)

Data Science and Big Data

Time : Three Hours

Maximum Marks : 70

- Note:* i) Attempt any five questions.
ii) All questions carry equal marks.

1. a) What is Data Wrangling? Define exploratory data analysis. Why EDA is required in Data Analysis? Discuss about some tools and techniques used for EDA.
b) Define and differentiate data cleaning and data transformation. What kind of issues affect the quality of data? How can I detect and fix these issues?
2. a) Discuss about data visualization. Discuss about the importance of data visualization with the help of an example.
b) What do you mean by feature extraction and feature selection? Explain some feature selection and feature extraction techniques available in R with the help of example.
3. a) What is reasoning under uncertainty? What is probability reasoning? How logical reasoning is differing with probability reasoning?
b) Explain and prove the Bayes Theorem. What is meant by conditional probability? Explain the method of performing exact inference in Bayesian Networks.
4. a) What is hypothesis testing? What are the steps of hypothesis testing? Explain in detail.
b) Discuss Maximum Likelihood Parameter Estimation (MLE) method with the help of an example.
5. a) What is Machine learning? What are the applications of machine learning? When and why we need machine learning? Explain with the help of an example.
b) Define and differentiate supervised and unsupervised learning. Explain the concept of learning using decision tree.
6. a) Explain about the over fitting and under fitting of the model. Explain the differences between "classification" and "clustering" and give an example of an application that would benefit from each techniques.
b) What is dimensionality reduction? Discuss about principal component analysis for dimensionality reduction.
7. a) Discuss about information retrieval. Discuss about various models used in information retrieval in detail.
b) Explain text mining. What kind of problem can be addressed using text mining? What kind of sentiments can be found in the text? Why is text mining useful in the age of social media?
8. a) What are the three types of web mining? What is click stream analysis? What are the two major ways by which a website can become popular?
b) Define Big data. What is Hadoop? How does Map-Reduce algorithm works? Discuss about the key issues in managing the big data.