

ME (Comp) Choice Based

Sem II Data Science

19/5/2012
Q.P. Code:16984

Marks: 80

Time: 3 Hours

Note: 1. Question 1 is compulsory

2. Answer any three out of remaining questions.
3. Assume suitable data wherever required and justify the same.

- Q1 a) Explain Linear discriminant analysis. [5]
b) What are the differences between supervised learning and unsupervised learning. [5] elaborate with an example.
c) What is the background of the formation of data journalism? [5]
d) Explain cross-validation for accuracy estimation. [5]
- Q2 a) Use PCA to transform 2D data space to 1D data space for the given matrix A. [10]
$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$$

b) Explain Maximum Likelihood estimation using expectation-maximization (EM). [10]
- Q3 a) Describe null hypothesis and alternative hypothesis with appropriate example [10]
b) Explain how Gaussian approximation works to a posterior distribution. [10]
- Q4 a) What type of problem were you looking to solve with text mining? How did you know how to text mine? What could be the challenges when text mining? [10]
b) Explain the process of collaborative based Recommendation System with suitable example. [10]
- Q5 a) Draw and describe the information visualization process. [10]
b) What infrastructure is most appropriate for Hadoop? Draw and describe Hadoop Ecosystem Architecture. [10]
- Q6 a) List and explain basic relational operators used in pig. [10]
b) What is the big deal with Big Data? If your company is just starting to consider using Big Data in you marketing research, what would be most useful to include? Explain in Data Science perspective. [10]
-