ST.JOSEPHS COLLEGE (AUTONOMOUS), BANGALORE- 27

M.Sc.Big Data Analytics-II SEMESTER

SEMESTER EXAMINATION - April 2017

BDA 1216 : Multivariate Statistics

Time: 2 hrs   **Maximum marks: 70**

(This question paper has 2 printed pages and 1 part.)

1. (a) State any three axioms of probability [for example, P(sure event) $= 1$].   [6]

   (b) If $P(A) = 1/3$ and $P(B) = 2/3$, then what is $P(A \cup B)$ if:

   - A and B are independent   [2]
   - A and B are disjoint   [2]

2. Let X be your expected mark in this exam. Let Y be the number of hours that you studied for this exam. Create a dummy $X - Y$ data set for 10 students and then:

   - Compute the correlation coefficient between X and Y [4]
   - Draw the scatter plot of X vs Y [2]
   - Write down the regression equation of Y (dependent variable) on X [4]

3. Discuss how you can convert the bivariate problem of Question 2 into a multivariate problem. Specifically highlight the following points (dont write more than one page in all)

   - New independent variables you might add   [3]
   - Probable presence of collinearity   [3]
   - Using R squared, or adjusted R squared? Which one? Why?   [4]

4. (a) Describe (in no more than 5 sentences) the benefits of principal component analysis.   [5]

   (b) Sketch (as a flow chart) the different steps involved in PCA.   [5]

5. (a) Mention one application where PCA can make a big difference.   [2]

   (b) Write a short note on "Orthogonal factor model"   [8]

6. (a) Give two real-life examples (from sport of business) where cluster analysis helps.   [4]

   (b) Write a short note on "Solving classification problem by clustering"   [6]

7.   (a) What is the underlying principle of clustering?   [4]

     (b) Sketch (as a flow chart) the rationale of k-means clustering.   [6]

8. A bank has a tricky decision to make. Should it offer a credit card to a customer with a seemingly modest income?

- What sort of multivariate analysis would you recommend?   [1]
- List out 10-12 possibly predictive variables?   [4]
- Sketch (as a flow chart) your options of stepwise regression   [5]

9. Write short notes on the following: [5+5]

- Multivariate techniques in social media
- The promise of artificial intelligence