Question Paper

Exam Date & Time: 17-Apr-2018 (10:00 AM - 01:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

SCHOOL OF INFORMATION SCIENCES (SOIS)
SECOND SEMESTER ME -(BIG DATA AND DATA ANALYTICS)
DEGREE EXAMINATION- APRIL/MAY 2018
Tuesday, April 17, 2018
Time : 10.00 am to 1.00 pm
Applied Multivariate Analysis [BDA 616.2]

Marks: 100 **Duration: 180 mins.**

Answer all the questions.

1) a) Discuss the importance of multivariate analysis in big data analytics. Define a sample covariance matrix. (3+2 = 5 marks) (5)

b) Let $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$ be a multivariate normal random vector with mean vector $\mu = \begin{bmatrix} 20 \\ 13 \\ 23 \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} 67 & 29 & 76 \\ 29 & 85 & 90 \\ 76 & 90 & 50 \end{bmatrix}$. Find mean and variance of $Z = 10Y_1 + 2Y_2 - 4Y_3$. (5)

2) Discuss the data structure, model, hypothesis and between and within variance structure in multivariate one-way analysis of variance (MANOVA). Also mention the names of different test statistics used in MANOVA. (2+2+1+3+2=10 Marks) (10)

3) Describe principal component analysis (PCA) for 'p' number of variables? Mention the different methods used for estimating factor loadings and communalities in factor analysis. (6+4 = 10 marks) (10)

4) Write the similarities and dissimilarities between principal component analysis and exploratory factor analysis. Write short notes on methods on deciding how many principal components to be retained. (6+4=10 Marks) (10)

5) What do you mean by rotation in factor analysis and discuss its uses? Explain different type of rotation used in (10)

factor analysis. (4+6 = 10 marks)

6) Describe the discriminant analysis for two group cases (10)

7) Describe the classification analysis for two group cases (10)

8) Explain the divisive and agglomerative method of identifying clusters in cluster analysis. (5+5=10 Marks) (10)

9) What do you mean by hierarchical clustering? Explain the average linkage and centroid approaches to measure the distance between clusters. (2+4+4=10 Marks) (10)

10) Differentiate between cluster analysis and discriminant analysis (10)

-----End-----