Syllabus for STAT 293 : Multivariate Staistical theory and methods

Instructor: Jayanta Kumar Pal.

Email: jkp3@duke.edu.

Class: W F 8:30 AM - 9:45 AM

Homepage: www.samsi.info/jpal.

Pre-requisite: Linear Algebra, univariate statistical methods and inference. STA 213, 214, 215 are required. STA 244 or consent of instructor.

Office hours: W F 10:00 AM - 11:00 AM

Software: R, Matlab, GGOBI.

Text: Class notes.
Applied Multivariate Statistical Analysis : Richard Johnson, Dean Wichern.
An introduction to multivariate Statistics : T.W. Anderson
Multivariate Analysis : K.V. Mardia, J.T. Kent, J.M. Bibby.
The Elements of Statistical Learning -data mining, inference and prediction : T. Hastie, R. Tibshirani, J. Friedman.

Grading: Homework - 5 sets - 75%. The homework will be a combination of theoretical problems and data analysis. Data sets will be either provided or referred in the sets. You may partner with someone else but submit your own solutions. Type-written solutions are preferable, but not absolutely mandatory. However, legibility should not be a problem.
Project - 25%. The final project will be based on a data analysis of a real problem and you are encouraged to find a data set or two of your own. The best scenario will involve using a data set from your own research. Start thinking about this now, and discuss with me about the methods you may need to use. Creativity is also encouraged. The final project will be graded subjectively, and excellence will be rewarded.

TOPICS

1. Introduction to multivariate analysis : Multivariate distribution theory, special emphasis on multivariate normal distribution - Modeling and inference - Estimation and testing of hypothesis - Multivariate regression - MANOVA - Correlation techniques - Multivariate medians. 8 lectures.

2. Dimension reduction techniques : Principal component analysis and SVD techniques - Factor Analysis - Canonical correlation - Multidimensional scaling - Correspondence analysis. 6 lectures.

3. Supervised and unsupervised learning : Discriminant analysis - Cluster analysis - Classification and regression trees - Support vector machines - bagging - Boosting. 8 lectures.

4. Theory of random matrices : Distribution of the eigenvalues and spacings - Tracy-Widom law - Application to high-dimension data - Regularization of covariance matrices. 4 lectures.

5. High dimensional problems and curse of dimensionality : large p small n scenario. recent techniques including data mining. As time permits.