Stat 675b: Unsupervised Learning: Dimension Reduction and Clustering Analysis Lisha Chen

Class meeting time and classroom: 10:30 – 12:30 Thursday, Room 107, 24 Hillhouse

Unsupervised learning, distinguished from supervised learning, is concerned with exploring data structure and extracting meaningful information from data without the guidance of a particular variable of interest. This course will be focused on two subfields of unsupervised learning, dimension reduction and clustering analysis. We cover both classical and recently developed methods concerning these two topics. Some applications of these methods will be discussed.

This course is intended for graduate students in statistics and related fields who are interested in obtaining a relatively comprehensive view and in-depth understanding in dimension reduction and clustering analysis. One purpose of this course is to complement the topics covered in Stat 665b Statistical Machine Learning (also known as Data Mining and Machine Learning) which focuses on statistical methods for supervised learning, though you are not assumed to take Stat 665b before this class. The grades will be based on class participation and presentations.

The requisites include matrix algebra, probability (at least on the level of Stat 241/541), and statistical inference (at least on the level of Stat 242/542).

Topics:

- 1. Principal component analysis (PCA) and some related methods (PCA, Principal Curve and Surface, Exploratory Projection Pursuit, Kernel PCA)
- 2. Multidimensional Scaling (Metric Scaling and Classical Scaling)
- 3. Recent development in nonlinear dimension reduction/manifold learning(Isomap, LLE, Laplacian Eigenmaps, Hessian Eigenmaps, Diffusion maps)
- 4. Clustering methods and algorithms (Probabilistic Formulation, Hierarchical Clustering, Partition Clustering)

References:

- 1. Hastie, Tibshirani and Friedman, The Element of Statistical Learning
- 2. Borg and Groenen, Modern Multidimensional Scaling
- 3. Mardia, Kent and Bibby, *Multivariate Analysis*
- 4. Jain and Dubes, Algorithms for Clustering Data
- 5. Research papers on recent advances in the fields