Dimension Reduction
and
Metric Learning
STAT 946

Instructor	Ali Ghodsi
Lecture Time/Room	1:00-2:20 TTh/MC 6007
Office Hour/Room	2:300-3:30 T or by appointment /MC 6081G

Description

The difficulty of extracting information from high-dimensional data is the main motivation for renewed interest in the problem of Distance Metric Learning and Dimensionality Reduction. High-dimensional data take many different forms: from Digital image libraries to gene expression microarrays and financial time series. Researchers in fields as diverse as finance, physics, medicine, and bioinformatics have to deal with such large data sets. By formulating this problem in a general setting, however, many different types of data can be analyzed in the same underlying mathematical framework. This course will explore this general framework. In addition several methods that are central to distance metric learning including positive semi-definite programming, kernel learning, large margin classification, and graph-based approaches will be discussed.

Textbook

There is no required textbook for the class. Some classic papers will be assigned as readings. A recommended book that cover the similar material is:
Hastie, Tibshirani, Friedman *Elements of Statistical Learning* (2nd Edition).

We will also be covering material similar to a variety of chapters from a few other books which I will point out in class.

Tentative Marking Scheme

Project 50%
Two paper critiques 30%
Paper presentation 20%

Prerequisite:

Some knowledge of calculus, linear algebra, and statistics

Tentative Topics:

- **Unsupervised Algorithms**
  - Linear Methods
    - Principal Components Analysis (PCA), Dual PCA
    - Metric Multidimensional Scaling (MDS)
    - Landmark MDS (Nystrom Approximation)
    - Non-negative Matrix Factorization
  - Nonlinear Methods
    - Locally Linear Embedding, ISOMAP, Local MDS, Laplacian Eigenmaps
    - Connection with Spectral Clustering
    - Stochastic Neighbor Embedding (SNE, t-SNE)
    - Deep Belief Networks
  - Unified Framework for Dimensionality reduction Algorithms
- **Supervised Algorithms**
  - Neighborhood Components Analysis
  - Relevant Component Analysis (RCA)
  - Large Margin Nearest Neighbor
- **Kernel Methods**
  - Kernel Alignment and Learning a Kernel with Semidefinite Programming
  - Maximum Variance Unfolding (MVU)
  - Colored MVU (Hilbert-Schmidt independence criterion (HSIC))
  - Distance Metric Learning based on Support Vector Machine (SVM)
- **Applications**
  - Graph Realization and Sensor Networks
  - Graph Partitioning (Clustering, Classification)
  - Microarray Clustering, Text mining (Biclustering, Co-clustering)
  - Search, Ranking (The Google PageRank Algorithm)
  - Image Processing and Image Segmentation