CSE 6339: SPECIAL TOPICS IN ADVANCED DATABASE SYSTEMS

Advanced Data Mining using Matrices and Graphs

Instructor: Prof. Chris Ding

Spring 2010. Time: TuTh 5:30-6:50pm, Location: TBD

This course covers practical and useful areas of data mining: (1) Techniques based matrices and scientific computing, (2) techniques based on graph and network theories, and (3) probabilistic and statistical approaches. Intended for students with basic knowledge of data mining, this course will improve the understanding of basic algorithms and methodologies, enrich the knowledge of data mining, and enhance the ability of designing new algorithms. Overall, students will gain a good comprehension of data mining field, understand why certain techniques are out of favor (such as neural networks and decision trees), and certain old methods remain popular (such as KNN and Linear Discriminant Analysis), and learn state-of-art and up-and-coming techniques. Topics covered:

1. **Classification methods**
   kNN, LDA, Naïve Bayes, Decision trees, logistic regression, SVM and statistical learning theory
2. **Clustering methods**
   K-means clustering, mixture models, EM algorithm, Hierarchical clustering, partitional clustering
3. **Feature Selection**
   Filter-type methods, F-test, mutual information
   max-relevance min-redundancy algorithm, feature stability algorithms
   Wrapper methods, search methods, floating search methods
4. **Graph Laplacian and related clustering and semi-supervised learning**
5. **PCA and its equivalence to K-means and kernel K-means clustering**
6. **Laplacian embedding and other embedding methods for clustering**
7. **Semi-supervised learning and positive-examples-only learning**
8. **Nonnegative Matrix Factorizations and its related clustering**
9. **Graphical Models, Random Markov Fields, Bayesian Networks**
10. **Data Ranking, PageRank, Web related graph algorithms**

The class emphasizes algorithm design and theory. This course is based on 6 tutorials I gave in data mining conferences with additional new materials.