Tentative Syllabus for CS 145: Introduction to Data Mining

Course Description

This course introduces basic concepts, algorithms, and techniques of data mining on different types of datasets, including (1) vector data, (2) set data, (3) sequence data, and (4) text data. The class project involves hands-on practice of mining useful knowledge from large data sets. The course is an undergraduate-level computer science course. Also, the course may attract students from other disciplines who need to understand, develop, and use data mining techniques to analyze large amounts of data.

Prerequisites

- You are expected to have background knowledge in data structures, algorithms, basic linear algebra, and basic statistics.
- You will also need to be familiar with at least one programming language, and have programming experiences.

Tentative Syllabus

- Week 1: Introduction and math review.
- Week 2: Vector data. Prediction: linear regression; Classification: logistic regression
- Week 3: Vector data. Classification: logistic regression, kNN.
- Week 4: Vector data. Classification: SVM and Neural Network.
- Week 5: Vector data. Clustering: k-means, hierarchical clustering, DBSCAN.
- Week 7: Set data. Frequent Pattern Mining and association rules.
- Week 8: Sequence data. Sequential pattern mining and similarity search.
- Week 9: Text data. Naïve Bayes and PLSA.
- Week 10: Review and Final.