**STAT 665/365: Data Mining and Machine Learning**

Lisha Chen  
Spring 2008

**Practical Information**
- Lectures: Monday and Wednesday 11:35 – 12:50; Rm 107, 24 Hillhouse Ave  
- Office hours: TBA

**Objectives**
This course is intended for graduate students and advanced undergraduate students in all majors who are interested in statistical learning methods to deal with massive or complex data. Data mining and machine learning is a complex topic with links to many fields. A lot of ideas in data mining and machine learning have been developed by scientists from the fields outside of statistics, such as computer science and engineering. In this course, we will emphasize on the statistical viewpoint of these ideas.

**Prerequisites**
Some prior exposure to probability and linear algebra – you should have taken at least one course on the level of STAT 241/242. Some programming experience in R/Splus or Matlab.

**Grading and Homeworks**
Grades will be based on homeworks (60%) and a final project (40%). In the final project, you will work on a problem of your own interest.

**Required Textbook**
Hastie, Tibshirani and Friedman, “The Elements of Statistical Learning – Data Mining, Inference and Prediction”.

**Topics (tentative)**
- Linear Methods for Regression  
- Linear Methods for Classification  
- Model Assessment and Selection  
- Basis Expansion and Regularization  
- Kernel Methods  
- Additive Models and Trees  
- Boosting  
- Support Vector Machine