# STAT 365/665 Data Mining and Machine Learning Spring 2009

## Instructors:

Harrison H. Zhou, Email: huibin.zhou@yale.edu TA: Peisi Yan, Email: peisi.yan@yale.edu

#### Textbook:

The Element of Statistical Learning by Hastie, Tibshirani and Friedman Download the book PDF: http://www-stat.stanford.edu/~tibs/ElemStatLearn/

### Class Time and Place:

Time: Mon, Wed 11:35 - 12:50; Place: 24 Hillhouse

## Tentative schedule:

Week 1: Overview and Applications to bioinformatics (No class on Friday)
Week 2: Linear Methods for Regression (Chapter 3)
Week 3: Linear Methods for Classification (Chapter 4)
Week 4: Basis Expansion and Regularization (Chapter 5)
Weeks 5 & 6: Additive Models, Trees and Related Methods (Chapter 9)
Weeks 7 & 8: Boosting and Additive Trees (Chapter 10)
Spring Break
Weeks 9 &10: Support Vector Machines and Flexible Discriminants (Chapter 12)
Weeks 11, 12 & 13: Some selected topics (TBA): Kernel Smoothing Methods, Neural

Networks, Bagging, Nearest-Neighbors, Cluster Analysis, Association Rules, Independent Component Analysis, Random Forests etc.

### Intended Audience:

It's necessary to have some background in linear algebra, probability and statistics (one the level of Math222, Stat 241 and Stat 242). Some programming experience is also a must.

#### Grade:

Homework: 65% Final project: 25% Participation: 10%