Math 569 – Statistical Learning

Course Description from Bulletin: The wealth of observational and experimental data available provides great opportunities for us to learn more about our world. This course teaches modern statistical methods for learning from data, such as, regression, classification, kernel methods, and support vector machines.

Enrollment: Elective for AM MS, PhD plus MMF plus other majors.

Textbook(s): T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer (2001), ISBN 0-387-95284-5.

Other required material: Software packages such as R, JMP, and MATLAB

Prerequisites: MATH 350 Numerical Methods and MATH 474 or 475, or consent of the instructor

Objectives:

1. Students will learn modern statistical techniques for modeling and drawing inferences from large data sets,
2. Students will learn to use visual and numerical diagnostics to assess the soundness of their models,
3. Students will become familiar with the computational requirements and compromises to be made in analyzing large data sets, and
4. Students will gain experience in analyzing real data sets and communicating their results.
5. Students will learn how to implement and use these numerical methods in Matlab (or another similar software package),
6. Students will improve their problem solving skills in computational mathematics,
7. Students will improve their presentation and writing skills.

Lecture schedule: 3 50 minutes (or 2 75 minutes) lectures per week

Course Outline:	Hours
1. Overview of Statistical Learning	6
a. Types of data, terminology	
b. Regression versus nearest neighbor models	
c. Variety of approaches and challenges	

2. Linear Methods for Regression and Classification 9
   a. Linear Regression and Least Squares
   b. Subset Selection and Coefficient Shrinkage for Linear Regression
   c. Linear Discriminant Analysis
   d. Logistic Regression
3. Basis Expansions and Kernel Methods 9
   a. Piecewise polynomials
   b. Smoothing Splines
   c. Regularization via Reproducing Kernel Hilbert Spaces
   d. Kernel Smoothers
   e. Local Regression
4. Model Assessment, Selection, and Inference 9
   a. Bias, Variance and Model Complexity
   b. Bayesian Approach and BIC
   c. Vapnik-Chernovenkis Dimension
   d. Cross-Validation
   e. Bootstrap and Maximum Likelihood Estimation
5. Advanced Topics 9
   a. Additive Models and Trees
   b. Neural Networks
   c. Support Vector Machines
   d. Nearest Neighbor Methods
   e. Unsupervised Learning

Assessment:	Homework	10-30%
	Computer Programs/Project	10-30%
	Tests	20-40%
	Final Exam	30-50%

Syllabus prepared by: Snejana Abarji and Fred J. Hickernell

Date: February 6, 2008