STAT 7620 Elements of Statistical Learning
Spring 2016

Lecture: MWF 11:30AM – 12:25PM in Dreese Lab 264

Instructor: Yoonkyung Lee
Office: 440B Cockins Hall
Office Hours: MW 4:10 – 5:10PM or by appointment
Email: yklee@stat.osu.edu

Grader: Yanan Jia
Office Hours: by appointment only
Email: jia.66@osu.edu

Text: *The Elements of Statistical Learning – Data Mining, Inference and Prediction*, Hastie, T., Tibshirani, R., and Friedman, J. (2009), 2nd edition, Springer.
The second edition (2009) is available as an ebook via the OhioLINK Electronic Book Center (see the course webpage on Carmen for access).

Course web page:

This course has a web page on Carmen. You will find class notes, homework assignments, solutions, references and other course announcements on the web page. Please check it on a regular basis.

Prerequisites: Probability/Math/Statistics - STAT 6301/6302 (610/623) or higher, or ECE 6001/7001, or equivalent, or permission of instructor. Familiarity with matrix algebra and linear regression analysis, and working knowledge of standard software packages such as R or Matlab.

Course Description: Statistical learning or machine learning methodology explores various ways of estimating functional dependencies between a response variable and possibly a large set of explanatory variables (features), when one is trying to find and understand an unknown, regular component within the realm of noisy, complex data. Modern regression and pattern recognition analyses fall in this framework. This course will provide an overview of supervised learning and discussions of statistical learning algorithms such as Discriminant Analysis, Classification Tree, Support Vector Machines, and Boosting, and illustrate practical uses of the algorithms. In addition, this course will cover cluster analysis and dimension reduction for unsupervised learning.

Tentative Course Schedule:

Week	Topics
1	Overview of Statistical Learning/Machine Learning (Chapter 1)
	Linear Methods for Regression (Chapter 3)
2	Penalized Regression
3-4	Linear Methods for Classification (Chapter 4)
5	Support Vector Machines (Chapter 12)
6	Basis Expansion and Regularization (Chapter 5)
7-8	Kernel Smoothing Methods (Chapter 6)
	Tree-Based Methods (Chapter 9)
9	Bagging and Boosting (Chapter 10)
10-11	Model Assessment and Selection (Chapter 7)
12	Cluster Analysis (Chapter 14)
13-14	Dimension Reduction (Chapter 14)
15	Project Presentations

Grading: Grades will be assigned on the basis of performance on homework assignments (40%) and a team project (60%).

Homework assignments: Homework will involve reading, analytical exercises, computational work, and data analyses. Homework assignments and solutions will be posted on the course web page.

Project: Class presentation and a project report are required for each team, which will be a key component of the course. Your project may involve new applications of statistical learning methods to real data, comparisons of various learning algorithms, or numerical or analytical studies of statistical issues of interest pertaining to the main theme of this course. Topics are not limited to those listed, and some specific topics may also be suggested by the instructor.

Tentative Timeline for Project:

Week 10	Select a topic and form a team.
Week 11	Project proposal (1 page) due March 28
Week 13	Preliminary report (5-6 pages) due April 11
Week 15	Class presentations*
Final Week	Final report (10-12 pages) due May 2

* The class may meet during the regularly scheduled final hours, 12:00-1:45PM on *Friday, April 29* for class presentations.