

Statistics 315a

Modern Applied Statistics: Elements of Statistical Learning (Winter 2006)

Table of Contents

- ▶ [Course Description](#)
- ◆ [Course Overview](#)
- ▶ [Contact Information](#)
- ◆ [Supplements](#)
- ▶ [Assignments](#)

Course Description

The meteoric rise in computing power has been accompanied by a rapid growth in the areas of statistical modeling and data analysis. New techniques have emerged for both predictive and descriptive learning that were not possible ten years ago, using ideas that bridge the gap between statistics, computer science and artificial intelligence. In this two-part series we cover many of these new methods, with emphasis on the statistical aspects of their application and their integration with more standard statistical methodology.

Predictive learning refers to estimating models from data with the specific goal of predicting future outcomes, in particular regression and classification models. Regression topics include linear regression with recent advances to deal with large numbers of variables, smoothing techniques, additive models, and local regression. Classification topics include discriminant analysis, logistic regression, support vector machines, generalized additive models, naive Bayes, mixture models and nearest neighbor methods.

Objectives

The objectives of this course are:

- To learn some of the traditional as well as the more recent tools for classification and regression
 - To understand these concepts from within a statistical decision theoretic framework
 - To learn some of the statistical inference tools for model selection and inference.
 - To get hands-on experience in using some of these techniques, through the homework assignments
-

Lectures

Monday, Wednesday 2:15-3:30

First Lecture: Wednesday January 11, 2006

Last Lecture: Wednesday March 15

Location: Skilling 193

This course will be broadcast [online](#) via SCPD (Stanford Center for Professional Development)

Instructor

[Trevor Hastie](#), Sequoia Hall room 104, hastie@stanford.edu, 5-2231. Office hours: Wednesday after class.
Secretary: Helen Tombropoulos 3-2620 Sequoia Hall first floor

Teaching Assistants

Mee Young Park (Head TA) (mypark@stanford.edu), Andreas Eckner (andreas@eckner.com), Yaqian Guo (yaqiang@stat.stanford.edu) and Ping Li (pingli98@stanford.edu).

Scheduled Office Hours

TA	Day	Start	End	Room
Park	?	?	?	Sequoia 237
Li	?	?	?	?
Guo	?	?	?	Sequoia 216
Eckner	Thurs	2	4	Sequoia 223A

E-Mail List

There are four mailing lists for this class. Students who have enrolled in the class in AXESS are automatically added in the mailing list. If you are just auditing the class and would like your name to be added to the mailing list, please send an email to stats315a-win0506-staff@lists.stanford.edu. You will not be able to subscribe to the mailing list unless you get approval from one of the staff.

- stats315a-win0506-staff@lists.stanford.edu: student can send professor and TAs email here
- stats315a-win0506-students@lists.stanford.edu: professor can reach students, but students can not post to this list
- stats315a-win0506-guests@lists.stanford.edu: auditors, or other guests can subscribe to this list
- stats315a-win0506-all@lists.stanford.edu: send email to this one address, and it will send to the other three lists. The other lists are subscribed to the "all" list

Prerequisites

Prerequisites: Some knowledge of probability and statistical inference, regression, linear algebra.

Computing

The course encourages students to have access to Splus or R. They both implement the S language, which is a programming language very useful for statistical applications.

[R](#) is available for free, and runs on all platforms (including Macs). [Rintro.pdf](#) is an introduction to the R language, providing a description of basic functionality and commands. Elizabeth Purdom has an excellent introduction to R and a list of basic functions on her [webpage](#) along with more specific computer labs on various subjects.

Splus exists on the Leland computers, on the Statistics computing environment, and is available on many other machines on campus. If you need help getting started with Splus, have a look at [S-guide.ps](#) Splus is also available for PCs, but not for Macs.

Students are encouraged to use the Emacs editor, since there exists a wonderful [ESS](#) mode for Emacs, which understands Splus and R syntax and more. [Online documentation for ESS](#) is available.

Homework

There will be regular biweekly homework assignments. Homework will include analysis of datasets, theoretical problems, and programming assignments.

Homework and other handouts will be distributed at lectures, and are available online (see the [Assignments](#) button in the left panel.) All of the homework assignments will be graded, and solutions will be made available.

LATE HOMEWORK will be penalized at 5% of the maximum score per day. Homework turned in more than 7 days late will not be graded. The final homework has a sharp deadline of the due date, which will be Friday March 17 at 5pm.

Exam

There will be no final exam for this course. The final grade will be determined entirely from the performance on assignments.

Texts

The course will follow closely [Elements of Statistical Learning](#) by Hastie, Tibshirani and Friedman (2001), Springer, NY. The book is available from the bookstore (class section).

Other References

Pattern Recognition and Neural Networks by B. Ripley, Oxford University Press, 1995.

Modern Applied Statistics with S (fourth edition) by W. Venables, and Ripley, B. Springer 2002.

Machine Learning, Neural and Statistical Classification edited by Michie, Spiegelhalter and Taylor, 1994, Ellis Norwood.

Pattern Classification (2nd Edition) by R. Duda, P. Hart and D. Stork, Wiley, 2000. *Neural Networks for Pattern Recognition* by C. Bishop, Clarendon Press, Oxford, 1995.

Short Course

An intensive two-day short course will be given on April 3-4, 2006 in Palo Alto:

[Statistical Learning and Data Mining](#)

by Professors Trevor Hastie and Robert Tibshirani. This course is also based on the text

[Elements of Statistical Learning](#), and will also cover topics in bioinformatics.

Please consult the course web page regularly. The URL is <http://www-stat-stanford.edu/~hastie/stats315a/>
