Term Information

Effective Term	Autumn 2014

General Information

Course Bulletin Listing/Subject Area	Statistics
Fiscal Unit/Academic Org	Statistics - D0694
College/Academic Group	Arts and Sciences
Level/Career	Undergraduate
Course Number/Catalog	4620
Course Title	Introduction to Statistical Learning
Transcript Abbreviation	Intr. Stat. Learn.
Course Description	The course provides an introduction to the principles of statistical learning and standard learning techniques for regression, classification, clustering, dimensionality reduction, and feature extraction.
Semester Credit Hours/Units	Fixed: 2

Offering Information

Length Of Course	14 Week
Flexibly Scheduled Course	Never
Does any section of this course have a distance education component?	No
Grading Basis	Letter Grade
Repeatable	No
Course Components	Lecture
Grade Roster Component	Lecture
Credit Available by Exam	No
Admission Condition Course	No
Off Campus	Never
Campus of Offering	Columbus

Prerequisites and Exclusions

Prerequisites/Corequisites	Stat 3302
Exclusions	

Cross-Listings

Cross-Listings	

Subject/CIP Code

Subject/CIP Code	27.0501
Subsidy Level	Baccalaureate Course
Intended Rank	Junior, Senior

Requirement/Elective Designation

Required for this unit's degrees, majors, and/or minors

Course Details

Course goals or learning objectives/outcomes

- Recognize the types of learning problems and understand their statistical formulations.
- Understand the foundational principles of statistical learning including statistical modeling, computation and evaluation.
- Comprehend the rationale and algorithms behind statistical learning techniques and know their relative merits and limitations.
- Evaluate and compare different learning techniques numerically in terms of generalization error.
- Use statistical learning methods for data analysis and interpret the results in the context of the data problem.

Content Topic List

- Predictive modeling and model evaluation
- penalized and nonparametric regression
- nearest neighbor methods
- classification and regression trees
- model selection and validation
- high dimensional data and variable selection

Attachments

- 4620_Syllabus.pdf

(Syllabus. Owner: Hans,Christopher M)

Comments

- This is a required course for the proposed major in Data Analytics. *(by Craigmile,Peter F on 10/11/2013 03:17 PM)*

Workflow Information

Status	User(s)	Date/Time	Step
Submitted	Hans,Christopher M	10/09/2013 02:52 PM	Submitted for Approval
Revision Requested	Craigmile,Peter F	10/11/2013 07:12 PM	Unit Approval
Submitted	Hans,Christopher M	10/12/2013 12:24 PM	Submitted for Approval
Approved	Craigmile,Peter F	10/13/2013 06:08 PM	Unit Approval
Approved	Hadad,Christopher Martin	10/14/2013 06:48 AM	College Approval
Pending Approval	Vankeerbergen,Bernadette Chantal Nolen,Dawn Jenkins,Mary Ellen Bigler Hogle,Danielle Nicole Hanlin,Deborah Kay	10/14/2013 06:48 AM	ASCCAO Approval

Statistics 4620

Introduction to Statistical Learning

2-semester-hour course

Prerequisite: Stat 3302 (Statistical Modeling for Discovery II)

Exclusions:

Class distribution: Two 55-minute lectures per week

Course Description and Learning Outcomes

The course provides an introduction to the principles of statistical learning and standard learning techniques for regression, classification, clustering, dimensionality reduction, and feature extraction.

Upon successful completion of the course, students will be able to

1. Recognize the types of learning problems and understand their statistical formulations.
2. Understand the foundational principles of statistical learning including statistical modeling, computation and evaluation.
3. Comprehend the rationale and algorithms behind statistical learning techniques and know their relative merits and limitations.
4. Evaluate and compare different learning techniques numerically in terms of generalization error.
5. Use statistical learning methods for data analysis and interpret the results in the context of the data problem.

Required Text and Other Course Materials

The required textbook for the course is (books currently under review). The book is available for purchase at the official University bookstore (ohiostate.bkstore.com) and elsewhere online. The book is available on reserve in the 18th Avenue Library.

Students will be required to use the R[1] software environment for statistical computing and graphics. R can be downloaded for free at http://www.r-project.org. Instructions for using the software will be given in class. Many students prefer to use RStudio, an IDE designed for use with R. RStudio is available for free at http://www.rstudio.com.

---

[1] For information on the use of R in data analytics, see:

- http://www.revolutionanalytics.com/why-revolution-r/whitepapers/r-is-hot.php
- http://techcrunch.com/2012/10/27/big-data-right-now-five-trendy-open-source-technologies/
- http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html
- http://bits.blogs.nytimes.com/2009/01/08/r-you-ready-for-r/

Assignments

Homework will be assigned (approximately) bi-weekly, will be due on the dates announced in class and will be graded. Assignments will consist of a mix of technical questions to assess students' understanding of the statistical models, and questions asking students to perform analyses of data sets. The grade for the analysis portion of each assignment will be based on both the accurateness and appropriateness of the analysis, as well as the clarity of the description of the analysis and results.

Project: Each student will be responsible for completing an individual project. Proposals for project ideas will be due mid-way through the semester, and the project will be due near the end of the semester. The project will consist of finding a data set, formulating questions that can be answered with the data, and performing an appropriate analysis to answer the questions.

Exams

There will be two in-class midterms that cover material from lecture, the assigned readings and homework.

A cumulative final examination will be given during the university's examination period.

Grading Information

The final course grade will be based on homework assignments, two projects, two midterms and a comprehensive final examination. The weights for each component of the grade are:

Homework	Midterm 1	Midterm 2	Project	Final Exam
15%	20%	20%	15%	30%

Outline of topics

1. Overview of predictive modeling and model evaluation
2. Penalized regression and nonparametric regression
3. Nearest neighbor methods
4. Classification and regression trees
5. Model selection and validation
6. High dimensional data and variable selection

Statement on Academic Misconduct

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term “academic misconduct” includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct http://studentlife.osu.edu/csc/.

Special Accommodations

Students with disabilities that have been certified by the Office for Disability Services will be appropriately accommodated and should inform the instructor as soon as possible of their needs. The Office for Disability Services is located in 150 Pomerene Hall, 1760 Neil Avenue; telephone 292-3307, TDD 292-0901; http://www.ods.ohio-state.edu/.