Term Information

Effective Term	Spring 2019

General Information

Course Bulletin Listing/Subject Area	Statistics
Fiscal Unit/Academic Org	Statistics - D0694
College/Academic Group	Arts and Sciences
Level/Career	Graduate
Course Number/Catalog	6500
Course Title	Statistical Machine Learning
Transcript Abbreviation	Stat Machine Learn
Course Description	Statistical models and algorithms for supervised and unsupervised learning; linear and logistic regression; classification and LDA; cross-validation and bootstrap; variable selection; ridge and LASSO penalization; smoothing splines and GAMs; SVM and kernels; CART and random forests; bagging; boosting; feed-forward and convolutional neural networks; k-means clustering and Gaussian mixtures; PCA.
Semester Credit Hours/Units	Fixed: 3

Offering Information

Length Of Course	14 Week, 7 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	6450 or permission of instructor
Exclusions	Not open to students with credit for 7620
Electronically Enforced	No

Cross-Listings

Cross-Listings

Subject/CIP Code

Subject/CIP Code	27.0501
Subsidy Level	Masters Course
Intended Rank	Masters, Doctoral

Requirement/Elective Designation

The course is an elective (for this or other units) or is a service course for other units

Course Details

Course goals or learning objectives/outcomes

- Students will understand the statistical learning framework, including core concepts such as loss, learning, and generalization; they will be able to judge when the framework is applicable and be able to formulate problems within this framework.
- Students will recognize the role of statistical models that are appropriate for a variety of statistical learning problems; they will understand the assumptions, formulation, and evaluation of these models.
- Students will understand the rationale and algorithms behind statistical learning methods, and they will know the merits and limitations of these methods.
- Students will be able to quantitatively evaluate and compare different statistical learning methods.
- Students will be able to apply statistical learning methods for data analysis and be able to interpret the results in the context of the application.

Content Topic List

- Logistic regression and Gaussian discriminant analysis
- Cross-validation and bootstrap
- Model selection and regularization
- Basis expansion and additive models
- Support vector machines
- Tree-based models
- Bagging and boosting
- Neural networks
- Principal components analysis
- Clustering
- Deep learning

Sought Concurrence

Yes

Attachments

- FW: Concurrence for STAT 6500.pdf: ECE Concurrence

  (Concurrence. Owner: Lee,Yoonkyung)
- STAT6500_Rationale.pdf: Rationale for the course proposal

  (Other Supporting Documentation. Owner: Lee,Yoonkyung)
- STAT6500-Syllabus.pdf: Syllabus

  (Syllabus. Owner: Lee,Yoonkyung)
- Concurrence for STAT 6500.pdf: CSE Concurrence

  (Concurrence. Owner: Lee,Yoonkyung)

Comments

Workflow Information

Status	User(s)	Date/Time	Step
Submitted	Lee,Yoonkyung	05/07/2018 09:27 AM	Submitted for Approval
Approved	Lee,Yoonkyung	05/07/2018 09:28 AM	Unit Approval
Approved	Haddad,Deborah Moore	05/07/2018 05:35 PM	College Approval
Pending Approval	Nolen,Dawn Vankeerbergen,Bernadette Chantal Oldroyd,Shelby Quinn Hanlin,Deborah Kay Jenkins,Mary Ellen Bigler	05/07/2018 05:35 PM	ASCCAO Approval

STAT 6500 – Statistical Machine Learning
3 CREDIT HOURS

Term: Spring, 2019
Instructor:
Office:
Email:
Phone:
Office Hours:
Location:
Format: Lecture
Course Website:

Course Description:
Statistical Machine Learning explores the methodology and algorithms behind modern supervised and unsupervised learning techniques to explore relationships between variables in large, complex datasets. Topics include linear and logistic regression, classification, clustering, resampling methods, model selection and regularization, and non-linear regression. Students will also gain exposure to popular statistical machine learning algorithms implemented in R. A focus will be on understanding the formulation of statistical models and their implementation, and the practical application of learning methods to real-world datasets.

Expected Learning Outcomes:

- Students will understand the statistical learning framework, including core concepts such as loss, learning, and generalization; they will be able to judge when the framework is applicable and be able to formulate problems within this framework.
- Students will recognize the role of statistical models that are appropriate for a variety of statistical learning problems; they will understand the assumptions, formulation, and evaluation of these models.
- Students will understand the rationale and algorithms behind statistical learning methods, and they will know the merits and limitations of these methods.
- Students will be able to quantitatively evaluate and compare different statistical learning methods.
- Students will be able to apply statistical learning methods for data analysis and be able to interpret the results in the context of the application.

Course Prerequisites:
STAT 6450 (Applied Regression Analysis) or permission of instructor. Familiarity with calculus, linear algebra and linear regression analysis will be assumed. Basic proficiency in a programming language, such as `R` or `Python`, is required.

Computing and Software:
One of the goals of the course is to train students in the computing aspects of statistical machine learning and develop the skills to implement machine learning algorithms. Many homework assignments will have a computing

and programming component (knowledge of software packages such as `Stata`, `SAS` or `JMP` will *not* be sufficient). There will be example codes provided, primarily written in the language `R`. Students are welcome to use `Python` and other languages, but with the understanding that support may be limited.

Textbooks:
The following textbooks are required for the course and are available to purchase at the university bookstore.

- James, Witten, Hastie, Tibshirani: *An Introduction to Statistical Learning with Applications in R.* (Freely downloadable PDF available at `http://www-bcf.usc.edu/~gareth/ISL`)
- Murphy: *Machine Learning: A Probabilistic Perspective* (An electronic version is available for online reading through the OSU library website)

Coursework:
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 datasets. 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. A tentative schedule for assignments is shown below. The assignments and assignment solutions will be posted on the course website. Late submissions will NOT be accepted.

A take-home midterm exam will be given. The midterm will be completed by each student individually.

Students will also complete projects in groups consisting of 3 to 4 members (depending on the enrollment size). The project will consist of selecting a data set (by week 2), performing an exploratory data analysis (EDA, by week 4), making a 5 page proposal (by week 10), presenting a poster (in week 15), and submitting a 10 page final report (by May 1st). The proposal should contain a detailed problem statement that includes questions of interest and a description of what methods will be used and how they will be used to answer questions of interest or solve the problem. More details will be given in class.

Dates:
Homework due dates and midterm dates are tentatively as follows. Please refer to in-class announcements for official dates.

Week	Description	Week	Description
2	Project dataset selection	9	HW4; take-home midterm due
3	HW1	10	Project proposal
4	Project EDA	11	HW5
5	HW2	13	HW6
7	HW3	15	Project poster session
8	Take-home midterm assigned	May 1st	Project final report

Grading:
Grades will be assigned on the basis of the following components.

- Homework (40%)
- Group project (30%)
- Take-home midterm exam (25%)
- Participation (5%)

Grading Scale:
Standard OSU grading scale applies.

Schedule of Topics:

Week	# Lectures	Topic	Description
1	1–2	Overview	
1	1	Linear regression	
2–3	5	Classification	Logistic regression, Gaussian LDA, algorithms for fitting
3–4	2	Resampling methods	Cross-validation, bootstrap
5–6	6	Model selection and regularization	Overfitting, variable selection, penalization, ridge regression, sparse linear models, LASSO, coordinate descent
7–8	5	Basis function models	Basis expansions, smoothing splines, additive models, backfitting, sparse additive models
8-9	4–5	Support vector machines and kernels	Max margin classification, separating hyperplanes, the kernel trick, comparison with earlier methods, nonlinear decision boundaries, perceptron algorithm
10	3	Tree-based methods	Classification and regression trees, variable importance measures
11–12	4	Bagging and boosting	Bagging, random forests, boosting, forward stage-wise additive modeling
12–13	3	Neural networks	Feed-forward NN (multilayer perceptrons), convolutional neural networks, back propagation
13–14	4	Unsupervised learning	k-means, Gaussian mixture models and EM, PCA
$*$	2	Deep learning	

*time permitting

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 Accomodations:

The University strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability (including mental health, chronic or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: slds@osu.edu; 614-292-3307; slds.osu.edu; 098 Baker Hall, 113 W. 12th Avenue.

Mental Health:
As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a students ability to participate in daily activities. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know are suffering from any of the aforementioned conditions, you can learn more about the broad range of confidential mental health services available on campus via theOffice of Student Lifes Counseling and Consultation Service (CCS) by visiting ccs.osu.eduor calling 614-292-5766. CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on call counselor when CCS is closed at 614-292-5766and 24 hour emergency help is also available through the 24/7 National Suicide Prevention Hotline at 1-800-273-TALK or at `suicidepreventionlifeline.org`.

Sexual Misconduct/Relationship Violence:
Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories (e.g., race). If you or someone you know has been sexually harassed or assaulted, you may find the appropriate resources at http://titleix.osu.eduor by contacting the Ohio State Title IX Coordinator, Kellie Brennan, at titleix@osu.edu.

Diversity at Ohio State:
The Ohio State University affirms the importance and value of diversity in the student body. Our programs and curricula reflect our multicultural society and global economy and seek to provide opportunities for students to learn more about persons who are different from them. We are committed to maintaining a community that recognizes and values the inherent worth and dignity of every person; fosters sensitivity, understanding, and mutual respect among each member of our community; and encourages each individual to strive to reach his or her own potential. Discrimination against any individual based upon protected status, which is defined as age, color, disability, gender identity or expression, national origin, race, religion, sex, sexual orientation, or veteran status, is prohibited.

Statistics 6500: Statistical Machine Learning

Rationale

Statistical/machine learning is increasingly in high demand across many disciplines. This demand is motivated by the desire to use statistical models to learn key relationships from data and apply such data-derived knowledge to understand, predict, manipulate and control the world around us. The broad usefulness of these statistical/machine learning techniques has been driven by technological shifts that have enabled a boom in both the collection of data and the capacity of computing systems to process these large datasets.
Much of this demand comes from industry where there is a need for talent with the skills to apply techniques that can enable automated learning from large datasets in a statistical fashion. Consequently, this demand is reflected in frequent requests from both masters students in the statistics department and graduate students from across campus for a course that can go beyond what is currently offered. For instance, many of the students who take STAT 6450 or 6560 have expressed interest in such a course. Both of these courses regularly enroll 40+ students each. However, the current courses on statistical learning that are offered (e.g., STAT 7620, and STAT 4620) are either too advanced or too basic to meet this demand. The aim of the proposed course is to train graduate students in modern statistical/machine learning methods while still remaining accessible to students with a broad set of backgrounds. The topic of statistical/machine learning has reached a point of maturity where it is no longer limited to a few domain experts, but rather used by a wide range of fields. Hence it is imperative that Ohio State University graduates aiming to join industry have a solid training in statistical/machine learning methods.

Course objectives and expected learning outcomes

- Students will understand the statistical learning framework, including core concepts such as loss, learning, and generalization; they will be able to judge when the framework is applicable and be able to formulate problems within this framework.
- Students will recognize the role of statistical models that are appropriate for a variety of statistical learning problems; they will understand the assumptions, formulation, and evaluation of these models
- Students will understand the rationale and algorithms behind statistical learning methods, and they will know the merits and limitations of these methods.
- Students will be able to quantitatively evaluate and compare different statistical learning methods.

- Students will be able to apply statistical learning methods for data analysis and be able to interpret the results in the context of the application.

Relationship to other courses

The Statistics Department currently offers an advanced Ph.D. elective course on statistical learning (STAT 7620) that focuses more on conceptual and theoretical aspects rather than practical applications. The course is typically offered biannually and is regularly at enrollment capacity. Many external as well as Masters of Applied Statistics students often attempt to enroll in the course, but find that the level of the course exceeds their background. The departments of Computer Science & Engineering and Electrical & Computer Engineering have their own courses which cover topics in statistical/machine learning. *CSE 5523: Machine Learning and Statistical Pattern Recognition* covers similar topics to this proposed course, however its primary audience consists of graduate students in CSE and as such requires a sequence of computer science courses that makes the course difficult to access for non-CSE students. ECE 7868: Pattern Recognition and Machine Learning covers a narrower selection of topics in statistical learning, mainly focusing on classification. It is only offered on even years and would be of limited use for our students.

RE: Concurrence for STAT 6500

Sivilotti, Paul
Sent: Friday, May 04, 2018 11:28 AM
To: Lee, Yoonkyung

Hi Yoon--

The CSE curriculum committee met this morning and approved the concurrence request for STAT 6500 "Statistical Machine learning".

As you know, machine learning is an active area of curricular development for CSE. We anticipate expanding our offerings in that area, for both CSE majors and for non-CSE students, resources permitting. We look forward to continued collaboration with Stats.

Best wishes,
--paul

-----Original Message-----
From: Lee, Yoonkyung
Sent: Wednesday, May 2, 2018 4:45 PM
To: Sivilotti, Paul
Subject: RE: Concurrence for STAT 6500

Hi Paul,

We understand how busy the end of the term could be. Thank you for letting us know. We will wait.

Yoon

_____
From: Sivilotti, Paul
Sent: Tuesday, May 01, 2018 8:45 AM
To: Lee, Yoonkyung
Subject: Re: Concurrence for STAT 6500

Hi Yoon--

We are having a curriculum committee meeting later this week to discuss... So you will have a response soon! The end-of-term crunch pushed this item back, so my apologies for the delay.

Regards,
--paul

> On Apr 30, 2018, at 10:52 PM, Lee, Yoonkyung <yklee@stat.osu.edu> wrote:
>
> Dear Paul,
>
> This is a friendly reminder that we are waiting for your concurrence on STAT 6500 Statistical Machine Learning in a couple of days. Let me know if you have any questions.
>
> Thanks,
> Yoon
>
> _____
> From: Soundarajan, Neelam
> Sent: Thursday, April 19, 2018 10:25 AM
> To: Lee, Yoonkyung

> Cc: Sivilotti, Paul
> Subject: RE: Concurrence for STAT 6500
>
> Dear Yoon,
>
> This has to be discussed by our Curriculum Comm. which is chaired by Paul Sivilotti. I have forwarded it to him.
>
> Best,
>
> --Neelam
>
> _____
> From: Lee, Yoonkyung
> Sent: Wednesday, April 18, 2018 10:26 PM
> To: Soundarajan, Neelam
> Subject: Concurrence for STAT 6500
>
> Dear Neelam,
>
> We would like to request your concurrence on a new graduate course, STAT 6500 Statistical Machine Learning. This course is designed for our masters students with interests in methods and algorithms for supervised and unsupervised learning from a statistical perspective and will serve as an elective for our graduate students. Attached please find the sample syllabus and our rationale for the course proposal.
>
> Please let me know if you have any questions. We would appreciate getting your response within two weeks.
>
> Thank you!
>
> Yoon
>
>
> --
> Yoonkyung Lee
> Professor of Statistics
> Professor of Computer Science and Engineering (by courtesy) The Ohio State University
>
>

--

Prof. Paul A. G. Sivilotti Computer Science and Engineering
The Ohio State University 2015 Neil Ave., Columbus OH, 43210
614.292.5835, Fax 292.2911 http://www.cse.ohio-state.edu/~paolo

Reply Reply All Forward Chat

FW: Concurrence for STAT 6500

Lee, Yoonkyung

To: Lee, Yoonkyung

Tuesday, April 24, 2018 3:47 PM

From: Anderson, Betty Lise
Sent: Tuesday, April 24, 2018 2:59 PM
To: Lee, Yoonkyung
Subject: Re: Concurrence for STAT 6500

Yoon,

I checked with the relevant people in ECE and ECE concurs with this course. Note we will possibly be proposing something on machine learning but only for ECE students and at the 4000- level. In fact I think you have already concurred with it.

Have a great summer!

> On Apr 18, 2018, at 10:12 PM, Lee, Yoonkyung <yklee@stat.osu.edu> wrote:
>
> Dear Betty,
>
> We would like to request your concurrence on a new graduate course, STAT 6500 Statistical Machine Learning. This course is designed for our masters students with interests in methods and algorithms for supervised and unsupervised learning from a statistical perspective and will serve as an elective for our graduate students. Attached please find the sample syllabus and our rationale for the course proposal.
>
> Please let me know if you have any questions. We would appreciate getting your response within two weeks.
>
> Thank you!
>
> Yoon
>
> --
> Yoonkyung Lee
> Professor of Statistics
> Professor of Computer Science and Engineering (by courtesy)
> The Ohio State University