STAT4620 – 2 CREDIT HOURS
Introduction to Statistical Learning

Term: Fall, 2019
Instructor: Matthew T. Pratola
Email: mpratola@stat.osu.edu
Location: WeFr 12:40-1:35pm Journalism Building 270
Office Hours: We 2:00-3:00pm CH204D
TA/Grader: Xiaohan Fu (fu.688@osu.edu)
DALC Hours: `https://data-analytics.osu.edu/dalc`
Project Presentations Day 1: Wednesday December 4th, 12:40-1:35pm.
Project Presentations Day 2: Monday December 9th, 4:00-5:45pm.
Course Website: Carmen (Canvas)
Long course title: Introduction to Statistical Learning

Course description:
This course provides an introduction to the principles of statistical learning and standard learning techniques for regression, classification, clustering, dimensionality reduction, and feature extraction. An outline of topics is:

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

Expected Learning Outcomes:
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.

Course Prerequisites:
C- or better in STAT3302 (Statistical Modeling for Discovery II)

Textbooks: The required textbook for the course is *An Introduction to Statistical Learning with Applications in R* by G. James, D. Witten, T. Hastie and R. Tibshirani. The book is available for purchase at the official University bookstore (`ohiostate.bkstore.com`) and is also available for free online in PDF format at `http://faculty.marshall.usc.edu/gareth-james/ISL/`.

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

Students are responsible for all material covered in class, in the assigned readings and in homework problems. As an introductory course, the quantity of material covered in the lectures is extensive. *It is highly recommended that you do not fall behind.*

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 datasets. The grade for the analysis portion of each assignment will be based on both the accuracy and appropriateness of the analysis, as well as the clarity of the description of the analysis and results.

Tentative due dates for the assignments are shown below. The assignments and assignment solutions will be posted on Carmen (Canvas). You must show all your work for all homework problems; do NOT just write the final answer.

You are encouraged to discuss homework problems with each other in general terms, but you must write your own homework solutions. Homework reports must be submitted in hardcopy. Late submissions will NOT be accepted. Academic misconduct of any sort will NOT be tolerated. Please review OSU's policies at http://studentlife.osu.edu/csc/.

Project:
Students will be responsible for completing a team project. Each team will consist of 3 students. 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 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. The course project will serve as a cumulative evaluation of your learning in lieu of a final examination. A basic calculator will be necessary for all exams (cell phone calculators, iPad's, etc., are not permitted). Cellphones must be silenced during class and are not allowed to be on the desk or otherwise accessible during exams. No make-up exams will be given.

Dates:
Homework due dates and midterm dates are tentatively as follows (please refer to in-class announcements for official dates):
HW1 09/06; HW2 09/20;
Midterm I 09/27 (in class);
HW3 10/09; HW4 10/25;
Midterm II 10/30 (in class);

HW5 11/08; HW6 11/22;
Project Report 12/04
Last time I checked, the last day to drop the course without a 'W' appearing on your record is September 13th and the last day to drop a course without petition is October 25th. However, please refer to the OSU registrar's office for official drop guidelines in case these dates change.

Grading:
The final course grade will be based on homework assignments, a project report and project presentations, and two midterms. The weights for each component of the grades are: **15%HW, 25%Midterm 1, 25% Midterm 2, 35% Project**.

(Tentative) Schedule of Topics:

Class	Date	Section	Topic
1	21-Aug	Ch1, Ch2, 3.1,3.2,3.5	Introduction; Linear Regression: SLR,MLR,Geometry and Loss
2	23-Aug	3.2,3.5	Linear Regression continued; Weighted Least Squares
3	28-Aug	3.3 and notes (7.1,7.2)	Beyond Linear Regression [**HW1 posted**]
4	30-Aug	5.1	Cross-Validation
5	4-Sep	4.1,4.2,4.3	Classification: Logistic Regression
6	6-Sep	4.4,4.5	Classification: Linear Discriminant Analysis (LDA) [**HW2 posted**]
7	11-Sep	5.2	The Bootstrap
8	13-Sep	5.3	Cross-Validation & Bootstrap Examples
9	18-Sep	6.1,6.2.1	Regularization: Ridge Regression
10	20-Sep	6.2.2,6.2.3	Regularization: The LASSO
11	25-Sep	6.3	Regularization: Dimension Reduction/PCA/PCR
-	27-Sep	**Midterm 1**	**Midterm 1** [**HW3 posted**]
12	2-Oct	6.4	High Dimensional Data Analysis and the Curse of Dimensionality
13	4-Oct	7.3,7.4	Spline Regression
14	9-Oct	7.5	Smoothing Splines [**HW4 posted**]
-	11-Oct	Holiday	Holiday
15	16-Oct	7.6	Local Regression
16	18-Oct	7.7	Generalized Additive Models (GAM's) (regression & classification)
17	23-Oct	8.1	Regression and Classification Trees
18	25-Oct	8.2	Trees: Bagging, Boosting [**HW5 posted**]
-	30-Oct	**Midterm 2**	**Midterm 2**
19	1-Nov	8.2	Trees: Random Forests
20	6-Nov	10.3	Clustering: K-means
21	8-Nov	10.3	Clustering: K-means & Hierarchical [**HW6 posted**]
22	13-Nov	10.3	Clustering: Hierarchical
23	15-Nov	-	$\star$Additional Topic
23	20-Nov	-	$\star$Additional Topic
23	22-Nov	-	`R` package Caret
-	27-Nov	Holiday	Holiday
-	29-Nov	Holiday	Holiday
26	4-Dec	-	Project Presentations
27	9-Dec	-	Project Presentations

$\star$ time permitting.

Data Analytics Learning Center: GTA / Grader Office Hours:
Graduate teaching assistants (GTAs) for Stat 3201, 3202, 3301, 3302, 3303 and 4620 will hold their office hours in the Data Analytics Learning Center (DALC) in Pomerene 151. The hours during which the GTA/grader for our course will hold office hours in PO 151 can be found at the top of the syllabus. You can meet with the GTA for our course in the DALC during his or her office hours to discuss questions you have about the course material, homework assignments, R, etc.

You are welcome to stop by the DALC when it is open even if it is not currently being staffed by the GTA for our course, e.g. if you are looking for a place to study or work on an assignment for one of the supported courses. If the DALC is staffed by a GTA for another Statistics course when you stop by, he or she will help you if possible, but may not be able to answer all of your questions.

A complete list of hours during which the DALC will be staffed by GTAs for Statistics Department courses can be found at `https://data-analytics.osu.edu/dalc`.

In rare situations due to last minute emergencies, the GTA assigned to the DALC may not be able to attend his or her office hours. If the DALC is closed when the schedule indicates it should be open, we recommend waiting for a few minutes. If no one shows up in a reasonable amount of time, please email your instructor to let us know about the problem. You can also contact your GTA to see about arranging a make-up time to meet.

Academic Misconduct:

ACADEMIC MISCONDUCT OF ANY SORT WILL *NOT* BE TOLERATED.

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.