Syllabus | CS 598/ STAT 542

Practical Statistical Learning

Logistics

Instructor	**Alexandra Chronopoulou**
Email	achronop@illinois.edu
Course Website	Coursera @ https://www.coursera.org/
Discussion Forum	Campuswire @ campuswire.com
Assignment Submission	Gradescope @ gradescope.com
Office Hours	On Zoom - Information and Details can be found on Coursera
TAs	TBD

Course Description

This course provides an introduction to modern techniques for statistical analysis of complex and massive data. Examples of these include model selection for regression, classification, nonparametric models such as splines and kernel models, regularization, dimension reduction, and clustering analysis. Applications are discussed as well as computation and theoretical foundations.

Course Prerequisites

Knowledge of basic multivariate calculus, statistical inference, and linear algebra. You should be comfortable with the following concepts: probability distribution functions, expectations, conditional distributions, likelihood functions, random samples, estimators, and linear regression models.

Approved Prerequisites: CS 498, or CS 410, or CS 412, or CS 445, or STAT 410.

Course Learning Outcomes

By the end of the course, you will be able to:

- Use a broad range of methods and techniques in machine learning.
- Have a deeper understanding of major algorithms and techniques in machine learning.
- Build analytics pipelines for regression problems.

- Build analytics pipelines for classification problems.
- Build analytics pipelines for recommendation problems.

Textbook & References

There is no required textbook for this course.

- Instructor's Notes: Link
- *"An Introduction to Statistical Learning with Applications in R"*, by James, Witte, Hastie and Tibshirani (basic)
- *"An Introduction to Statistical Learning with Applications in Python"*, by James, Witte, Hastie, Tibshirani, and Taylor (basic)
- *"The Elements of Statistical Learning: Data Mining, Inference, and Prediction"*, by Hastie, Tibshirani, and Friedman (more advanced)
- You may also want to view the Data School YouTube videos associated with the ESL book, as an additional resource.

Course Software

In this course, you are welcome to use R or Python to complete the course coding assignments and/or projects.

Coursework

The course is structured in 16-week modules. Each week is dedicated to a specific topic and will comprise of a mix of the following items that will be posted on *Coursera*:

- Video Lectures
  Each week a topic will be presented through a collection of short video lectures.
- Graded Quizzes
  There are **10** *quizzes* associated with the lectures during the semester. The quizzes should be done individually and have unlimited attempts and there is no time limit on how long you take to complete each attempt. Each quiz is worth 4% of your final course grade.
- Coding Assignments
  During the semester, there will be **5** *coding assignments* where you will need to practice coding some of the methods discussed in the course. The coding assignments may be completed in groups or individually. Each coding assignment is worth 6% of your final course grade.
- Projects
  Instead of exams, there will be **4** small projects during the semester. Those will be focused on analyzing a data set and creating a report of your analysis. The coding assignments may be completed in groups or individually. Each project is worth 7.5% of your final course grade.

A detailed Lectures Plan with dates can be found in the end of the Syllabus, and on Coursera.

Grading

Grading Scheme	
Graded Quizzes	40%
Coding Assignments	30%
Projects	30%

Letter Range	Percentage
A-/A; no A+	90.00 – 100.00
B-/B/B+	80.00 – 89.99
C-/C/C+	70.00 – 79.99
D-/D/D+	60.00 – 69.99
F	≤ 59.99

Student Code and Policies

A student at the University of Illinois at the Urbana-Champaign campus is a member of a University community of which all members have at least the rights and responsibilities common to all citizens, free from institutional censorship; affiliation with the University as a student does not diminish the rights or responsibilities held by a student or any other community member as a citizen of larger communities of the state, the nation, and the world. See the University of Illinois Student Code for more information.

Academic Integrity

It is expected that all students will support the idea of academic integrity and be responsible for the integrity of their work. The university has a published policy on academic integrity that may be found at https://provost.illinois.edu/policies/policies/academic-integrity/

These standards will be enforced and infractions of these rules will not be tolerated in this course. Sharing, copying, or providing any part of a homework solution or code is an infraction of the University's rules on academic integrity. We will be actively looking for violations of this policy in homework and project submissions. Any violation will be punished as severely as possible with sanctions and penalties typically ranging from a failing grade on this assignment up to a failing grade in the course, including a letter of the offending infraction kept in the student's permanent university record.

Again, a good rule of thumb: Keep every typed word and piece of code your own. If you think you are operating in a gray area, you probably are. If you would like clarification on specifics, please contact the course staff.

Special Accommodations

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES, you may call (217) 333-1970, e-mail disability@illinois.edu or go to the DRES website.

The instructor reserves the right to make any changes she considers academically advisable. Such changes, if any, will be announced via email and class announcements. It is your responsibility to keep track of the proceedings.

CS 598 PSL - Schedule

Week	Topic	Concepts	Coding Assignments	Quiz	Project
1	Introduction	Introduction to Statistical Learning, Variance and bias trade-off, Model evaluation	CA 1 assigned	Q1	
2	Linear Regression	Linear regression review, Model assessment, and Some practical issues.	CA 1 due	Q2	
3	Variable Selection and Regularization	Subset selection, Ridge regression, Lasso, Principal components regression	CA 2 assigned	Q3	
4	Nonlinear Regression	Polynomial regression, Splines, Local smoothers	CA 2 due	Q4	P1 assigned
5	Project Work Week	Work on the project			
6	Regression Trees & Ensemble	Regression Trees, Pruning, Tree ensemble	CA 3 assigned	Q5	P1 due
7	Classification: Introduction	Overview, Classification theory, and evaluation, Naïve Bayes classifiers, Linear discriminant analysis	CA 3 due	Q6	P2 assigned
8	Logistic Regression	Logistic regression with regularization		Q7	
9	Support Vector Machine	Support vector machines		Q8	
10	Classification Trees & Boosting	Classification Trees, Pruning, Combination of trees	CA 4 assigned	Q9	P2 due
11	Recommender System	Recommender systems	CA 4 due	Q10	P3 assigned
12	Project Work Week	Work on the project			
13	Clustering Analysis	K-means, Spectral clustering, Clustering evaluation	CA 5 assigned		P3 due/ P4 assigned
14	Fall Break				
15	Latent Structure Models	Model-based clustering, EM algorithm, Latent Dirichlet Allocation model	CA 5 due		
16	Project Work Week	Work on the final project		All quizzes due	P4 due

CS 598 PSL - Schedule

Quiz	Release Date	Suggested Deadline	Hard Deadline
Quiz #1	First day of class	Monday of Week 3	Sunday of Week 16
Quiz #2	First day of class	Monday of Week 4	Sunday of Week 16
Quiz #3	First day of class	Monday of Week 5	Sunday of Week 16
Quiz #4	First day of class	Monday of Week 6	Sunday of Week 16
Quiz #5	Friday of Week 2	Monday of Week 7	Sunday of Week 16
Quiz #6	Friday of Week 2	Monday of Week 8	Sunday of Week 16
Quiz #7	Friday of Week 2	Monday of Week 9	Sunday of Week 16
Quiz #8	Friday of Week 3	Monday of Week 10	Sunday of Week 16
Quiz #9	Friday of Week 3	Monday of Week 12	Sunday of Week 16
Quiz #10	Friday of Week 3	Monday of Week 13	Sunday of Week 16

Coding	Release Date	Hard Deadline
Coding #1	First day of class	Monday of Week 2
Coding #2	First day of class	Monday of Week 4
Coding #3	Friday of Week 2	Monday of Week 7
Coding #4	Friday of Week 3	Monday of Week 11
Coding #5	Friday of Week 3	Monday of Week 15

Project	Release Date	Hard Deadline
Project 1	First day of class	Monday of Week 6
Project 2	Friday of Week 2	Monday of Week 10
Project 3	Friday of Week 3	Monday of Week 13
Project 4	Friday of Week 3	**Sunday** of Week 16