Statistical Machine Learning

Statistics GR5241/GU4241 — Spring 2020

TTH 2:40 pm–3:55 pm; HyFlex

Before Feburary 1st, **ALL CLASS WILL BE OFFERED ONLINE ONLY. COVID-19 CORONAVIRUS.** The policies set forth in this syllabus are subject to change as we try to determine how best to keep you safe from the COVID-19 coronavirus while we provide the education we promised you.

Instructor:	Xiaofei Shi (xs2427@columbia.edu)	*Office hours:* T & Th 12:30-1:30 pm
TA:	Dingyi Fang (df2709@columbia.edu)	*Office hours:* M 4:00-5:30 pm

Course Description:

The course will provide an introduction to Machine Learning and its core models and algorithms. The aim of the course is to provide students of statistics with detailed knowledge of how Machine Learning methods work and how statistical models can be brought to bear in computer systems—not only to analyze large data sets, but to let computers perform tasks that traditional methods of computer science are unable to address. Examples range from speech recognition and text analysis through bioinformatics and medical diagnosis. This course provides a first introduction to the statistical methods and mathematical concepts which make such technologies possible.

Prerequisite: A previous course in statistics, elementary probability, multivariate calculus, linear algebra and ability to do moderate coding.

Programming: Coursework will require the use of statistical software. While python is recommended, you are free to use any language to finish the homework, including Matlab and R.

Evaluation: Your course grade will be computed as the following weighted average:

- Assignments: 40% weight for 4 assignments
- Exams: 30% (15% weight for the midterm exam, 15% weight for the final exam)
- 25% weight for the course project
- 5% weight for class participation.

I will then convert the course grade to a letter grade. The possible letter grades that will be assigned at this point are A+, A, A-, B+, B, B-, C+, C, C-, D and R.

Tentative course outline:

Date	Topics	
Mon 01/11	**Assignment 1 out**	
Tue 01/12	Introduction	Online
Thu 01/14	MLE and MAP	Online
Tue 01/19	Regression: Linear Regression	Online
Thu 01/21	Regression: Regularized Regression and Logistic Regression	Online
Tue 01/26	Recitation 1 (by Dingyi)	Online
Thu 01/28	Decision Tree	Online
Mon 02/01	**Assignment 2 out, Assignment 2 due**	
Tue 02/02	Support Vector Machines 1	Online
Thu 02/04	Support Vector Machines 2	Online
Tue 02/09	K nearest neighbors	Online
Thu 02/11	Naive Bayes, Generative vs Discriminative	Online
Tue 02/16	Boosting, Surrogate Losses, Ensemble Methods	Online
Thu 02/18	Recitation 2 (by Dingyi)	Online
Mon 02/22	**Assignment 3 out, Assignment 2 due**	
Tue 02/23	Midterm review	Online
Thu 02/25	Midterm Exam	Online
Tue 03/09	Clustering: K means	417 Math Building
Thu 03/11	Clustering: Mixture of Gaussians, Expectation Maximization	Online
Tue 03/16	Optimization Methods: Gradient Descent	417 Math Building
Thu 03/18	Recitation 3 (by Dingyi)	Online
Mon 03/22	**Assignment 3 out, Assignment 4 due**	
Tue 03/23	Optimization Methods: Modern Technique	417 Math Building
Thu 03/25	Neural Networks	417 Math Building
Tue 03/30	Introduction to Deep Learning	417 Math Building
Thu 04/01	Sequence Models: HMMs	417 Math Building
Tue 04/06	Recitation 4	Online
Thu 04/08	Presentation	Online
Mon 04/12	**Assignment 4 due**	
Tue 04/13	Presentation	Online
Thu 04/15	Final Review	417 Math Building
Fri 04/23	**Final Project Due**	

Policy on Collaboration:
You are encouraged to work together on the homework. Discussing the homework problems with one another can be a valuable learning experience. However, it is a violation of the rules on academic integrity to copy another student's solution and submit it as your own. **You should write up your solutions separately, not referring to a common document**. Furthermore, you should not submit any work that you do not fully understand. You should be able to start with a clean sheet of paper and without notes or assistance write out the solution to any homework solution you submit. If you will do that with every homework you submit, the similarity between your solutions and those of other students will not arouse suspicion. More importantly, you will be well prepared for the exams. You are not permitted to use homework solutions for this course from previous years or solutions you find from other sources, including the internet.

You do not need and may not use anything during the exams except writing instruments and a non-programmable calculator. In particular, cell phones must be put away during exams. During the exam, you may not give nor receive assistance from other students. Violation of these rules will be treated seriously according to school policy. The most common penalty for violations of the homework policy is that you will receive a grade of 0 on that homework and an additional one letter reduction in your course grade. The standard penalty for violations during an exam is failure in the course. Second violations or first violations that are accompanied by lies denying what the evidence shows generally result in suspension or dismissal from the program. Persons providing assistance in violation of these rules are subject to the same penalties as persons receiving assistance.

People caught cheating generally tell a story about being unable to keep up with the material. First of all, remember that you can submit homework late. Secondly, if there is something going on in your life that makes it hard for you to keep up, please let me know. That brings me to the following point.

Taking Care of Yourself. It is easy for me to say and hard for all of us, including me, to do, but taking care of your physical and mental health is essential, especially during the COVID-19 pandemic. Life is a marathon, and you need to pace yourself.

Do your best to maintain a healthy lifestyle by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

If you or anyone you know experiences extreme academic stress, difficult life events, or feelings of anxiety or depression, I strongly encourage you to seek support. Counseling and Psychological Services is here to help 24/7, and everything will be confidential: call 212-854-2878 or visit

https://health.columbia.edu/content/counseling-and-psychological-services

In addition, consider reaching out to a friend, faculty or family member you trust for help getting connected to the support. Keep in mind that for serious psychological issues, the first counselor you meet with may not be the right one for you, but this does not mean you should give up on counseling. Keep looking for someone who can help you.

If you or someone you know is feeling suicidal or in danger of self-harm, call immediately, day or night:

Counseling and Psychological Services: 212-854-2878

If the situation is life threatening, call the police:

On campus: Columbia Police: 212-854-2797

Off campus: 911

If you have questions about this advice, your coursework, or anything else about which I might be helpful, please let me know.

Rubric and policies are designed with the experience from MSML and MSCF program at CMU. Materials are based off lectures from Prof. Peter Orbanz and Prof. Cynthia Rush at Columbia University, 10-725 lectures at CMU from Prof. Ryan Tibshirani, and 10-701 lectures at CMU from Prof. Ziv Bar-Joseph, Prof. Pradeep Ravikumar and Prof. Aarti Singh.