

Syllabus of STAT154

(Modern Statistical Prediction and Machine Learning)

Instructor: Song Mei (songmei@berkeley.edu)

Lectures: T/Th 17:00 - 18:29. Location: Moffitt Library 101.

Instructor office hours: Will announce on course homepage.

GSI: Ryan Theisen (theisen@berkeley.edu)

Lab 101: M 10:00 - 11:59. Location: Moffitt Library 103.

Lab 102: M 14:00 - 15:59. Location: Wheeler 220.

GSI office hours: Will announce on course homepage.

Zoom link (Lectures, Labs, Office hours): <https://berkeley.zoom.us/j/93149999627>

Important websites

Course website (for general logistics): https://www.stat.berkeley.edu/~songmei/Teaching/STAT154_Fall2021/index.html

bCourses (for potentially recordings): <https://bcourses.berkeley.edu/courses/1507964/>

Piazza (for questions): <https://piazza.com/class/kst8dnv06k33vo>

Gradescope (for submitting homeworks): <https://www.gradescope.com/courses/299436>.

Entry code: D5G6WN.

Course introduction

This course will focus on statistical learning methods and data analysis skills. Upon completing this course, the students are expected to be able to 1) build baseline models for real world data analysis problems; 2) implement models using programming languages; 3) draw conclusions from models.

Topics

Basic topics:

Tasks: Regression. Classification. Dimension reduction. Clustering.

Algorithms: Solving linear systems. Gradient descent. Newton's method. Power iteration for eigenvalue problems. EM algorithms.

Others: Kernel methods. Regularization. Sample splitting. Resampling methods. Cross validation.

Advanced topics:

Statistical learning theory and optimization theory.

Bagging and Boosting. Tree based models. Neural networks. Bayesian models.

Online learning. Bandit problems.

Textbooks

An Introduction to Statistical Learning. Elements of Statistical Learning. (free pdf online)

Other references

Stanford CS229 lecture notes: <http://cs229.stanford.edu/syllabus-fall2020.html>

Berkeley CS189/289 lecture notes: <https://people.eecs.berkeley.edu/~jrs/189/>

Prerequisite

MATH 53 and 54 or equivalents; MATH 110 is highly recommended. STAT 135 or equivalent (DATA/STAT C100 *and* (STAT 134 *or* DATA/STAT C140) will be accepted). STAT 133 preferred. STAT 151A is recommended. Scripting language required and R experience recommended.

Homework/Grading

- Class attendance is required.
- There will be 6-7 homeworks.
- In class mid-term. Date TBA.
- Final exam date: Dec 16, 11:30 am - 2:30 pm.
- Final grade will be Homework \times 50 % + mid-term \times 20 % + final \times 30 %.

Code of conduct; attribution of work

The high academic standard at the University of California, Berkeley, is reflected in each degree awarded. Every student is expected to maintain this high standard by ensuring that all academic work reflects unique ideas or properly attributes the ideas to the original sources.

These are some basic expectations of students with regards to academic integrity: Any work submitted should be your own individual thoughts, and should not have been submitted for credit in another course unless you have prior written permission to re-use it in this course from this instructor.

All assignments must use “proper attribution,” meaning that you have identified the original source and extent of words or ideas that you reproduce or use in your assignment. This includes drafts and homework assignments! If you are unclear about expectations, ask your instructor.

Do not collaborate or work with other students on assignments or projects unless the instructor gives you permission or instruction to do so.

Disability accommodations

If you need an accommodation for a disability, if you have information you wish to share with the instructor about a medical emergency, or if you need special arrangements if the building needs to be evacuated, please inform the instructor as soon as possible.

If you are not currently listed with DSP (the Disabled Students' Program) and believe you might benefit from their support, please apply online at <https://dsp.berkeley.edu/>.