

UCLA

Computer Science Department

CS 97: Introduction to Data Science

Summer 2021

Course Description: This course is an adapted from CS 188: Data Science of Fundamentals. The fundamental question this course aims to address is: given data arising in real-world, how does one analyze that data so as to understand the corresponding phenomenon. This course will cover topics in machine learning, data analytics, and statistical modeling classically employed for prediction. The course will be a blend of theoretical and practical instruction, providing a comprehensive, hands-on overview of the Data Science domain. The course will seek to teach students the data science lifecycle: data selection and cleaning, feature engineering, model selection, and prediction methodologies.

Instructor: Yizhou Sun (yzsun@cs.ucla.edu)

Textbook (Optional):

1. Fundamentals of Machine Learning for Predictive Data Analytics Algorithms, Worked Examples, and Case Studies.
 2. Machine Learning: An Algorithmic Perspective, Second Edition Part of: Chapman & Hall/Crc Machine Learning & Pattern Recognition (21 Books) | by Stephen Marsland.
 3. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc., 2012 | by McKinney, Wes.
 4. Probabilistic programming and Bayesian methods for hackers., 2015 | by Pilon, Cameron Davidson.
-

Course material: Lectures, assignments and solutions will be posted on CCLE or/and Gradescope

Class Communication: Important class announcements will be done through online class forum on Piazza. If you have any questions regarding class materials, they also need to be asked on Piazza.

Grading:

Homework	40%
MIDTERM	30%
Project	30%

You may discuss problems with friends, but you must write your solutions individually. I expect all students to follow the [UCLA Student Conduct Code](#), which prohibits cheating, fabrication, and multiple submissions.

Schedule:

Week 1

	Morning (9:00 -- 12:00)	Afternoon (1:00 -- 4:00)
7/12	Introduction; Know Your Data	Discussion and Lab
7/13	Linear Regression	Discussion and Lab
7/14	Regularization; Model Selection	Discussion and Lab
7/15	Logistic Regression	Discussion and Lab
7/16	kNN and SVM	Lab Touring / Seminar

Week 2

	Morning (9:00 -- 12:00)	Afternoon (1:00 -- 4:00)
7/19	Decision Tree	Discussion and Lab
7/20	PCA/Missing Data & Imputation	Discussion and Lab
7/21	Clustering and K-Means	Discussion and Lab
7/22	Perceptron and NN	Discussion and Lab
7/23	NN: Design/Training/ Regularization	Lab Touring / Seminar

Week 3

	Morning (9:00 -- 12:00)	Afternoon (1:00 -- 4:00)
7/26	Application: Health	Exam (90 mins); Discussion and Lab
7/27	Application: Text	Discussion and Lab
7/28	Application: Image	Discussion and Lab
7/29	Application: Recommender Systems	Discussion and Lab
7/30	Capstone Project Presentation	Capstone Project Presentation