UCLA
Computer Science Department

Data Science Fundamentals (CS M148)   Winter 2023

Course Description: 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.

Requisites: COM SCI 31 or COMPTNG 10A, 10B, and one course from C&EE 110, EC ENGR 131A, MATH 170A, 170E, or STATS 100A.

Instructor: Baharan Mirzasoleiman (baharan@cs.ucla.edu)
Teaching Assistants:
- Lionel Levine (lionel@cs.ucla.edu)
- Siddharth Joshi (sjoshi804@gmail.com)
- Wenhan Yang (hangeryang18@g.ucla.edu)
- Yihe Deng (yihedeng@ucla.edu)
- Yu Yang (yuyang@cs.ucla.edu)

Lectures: Monday/Wednesday 8:00 am - 9:50 am, Kaplan 150, Zoom
Discussions:
- Lionel Levine Sec (1A) F 10:00 pm - 11:50 pm, Royce Hall 156
- Siddharth Joshi Sec (1B) F 12:00 pm - 1:50 pm, Royce Hall 154
- Wenhan Yang Sec (1D) F 4:00 pm - 5:50 pm, Franz Hall 1260
- Yihe Deng Sec (1C) F 2:00 pm - 3:50 pm, Dodd Hall 78
- Yu Yang Sec (1E) F 12:00 pm - 1:50 pm, Boelter Hall 2760

Office hours (Lionel Levine**):** Thursday / 1-2pm, Zoom (see canvas)
Office hours (Siddharth Joshi): Wednesday / 2:30-3:30pm, Zoom (see canvas)
Office hours (Wenhan Yang): Thursday / 2-3pm, Zoom (see canvas)
Office hours (Yihe Deng): Tuesday / 1-2pm, Zoom (see canvas)
Office hours (Yu Yang): Monday / 1-2pm, Zoom (see canvas)
Office hours (Baharan Mirzasoleiman**):** Monday 10am-10:30am, Zoom
All the zoom links are also posted on **Canvas**.

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 Canvas 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. Please make sure to sign up for Piazza forum at https://piazza.com/ucla/winter2023/csm148

Grading:

Homework	20%	3 homework assignments
Midterm	20%	5th week of the class (90 minutes)
Projects (3)	20%	3 projects
Final	40%	Final exam week

Late policy: No late projects / homework will be accepted

You must complete the assignments and projects **entirely on your own**. You are NOT allowed to discuss your solutions with others or see another student's solutions. Gradescope is used to submit assignments.

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:

Date	Topics	Assignment	Ungraded Labs
		Week 1	
Jan 9	Introduction		
Jan 11	Data Collection, Cleaning, Bias in the Data		Lab 1: Intro to Python (numpy, graphing libraries, program structure, Jupyter Notebook)
Jan 13	Discussion	Project 1 Posted	
		Week 2	
Jan 16	Martin Luther King, Jr. holiday		Lab 2: Python: sklearn, matplotlib
Jan 18	kNN/Linear Regression		Lab 3: Scikit-learn for Simple Linear Regression
Jan 20	Discussion	Project 1 due Project 2 Posted	
		Week 3	
Jan 23	Multi and Poly Linear Regression/ Model Selection & Cross Validation		
Jan 25	Hypothesis Testing		Lab 4: Multiple Linear Regression and Cross Validation
Jan 27	Discussion	HW1 Posted Project 2 due	
		Week 4	
Jan 30	Regularization: Ridge & Lasso Logistic Regression		Lab 5: Logistic Regression

Feb 1	Logistic Regression / Multi-class Logistic Regression		
Feb 3	Discussion	HW1 due HW2 posted	
	Week 5		
Feb 6	kNN classification and SVM		Lab 6: KNN Classification & Imputation
Feb 8	Decision Trees		
Feb 10	Discussion	HW 2 due	
	Week 6		
Feb 13	Midterm Exam Review		
Feb 15	Midterm Exam		Lab 7: Trees and Random Forests
Feb 17	Discussion	HW 3 posted	
	Week 7		
Feb 20	Presidents' Day holiday		Lab 8: Intro to NN
Feb 22	Missing Data & Imputation		
Feb 24	Discussion		
	Week 8		
Feb 27	Perceptron and MLP		
March 1	Neural Network Design		Lab 9: Regularization with NN
March 3	Discussion	HW 3 due Project 3 posted	
	Week 9		
March 6	NN: Design/Back Propagation/ Regularization/ Gradient Descent		Lab 10: Clustering
March 8	Unsupervised (clustering)/Review		
March 10	Discussion		
	Week 10		
March 13	Model Interpretation		
March 15	Extra: Learning from large datasets		

March 17	Discussion	Project 3 due	

*The project/Homework deadlines are due before the class starts.