# UCLA

**Computer Science Department**

**Data Science Fundamentals (CS/ECE M148) Spring 2021**

**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:** Majid Sarrafzadeh (majid@cs.ucla.edu)

**Teaching Assistants**:

* Orpaz Goldstein ([orpgol@cs.ucla.edu](mailto:orpgol@cs.ucla.edu))
* Sajad Darabi (sajad.darabi@cs.ucla.edu)
* Lionel Levine ([lionel@cs.ucla.edu)](mailto:lionel@cs.ucla.edu)),

**Lectures:** Tuesday/Thursday 10:00 am - 11:50 am, [Zoom](https://ucla.zoom.us/my/majid1)

**Discussions:**

* Sajad Darabi Sec (1A) F 10:00 am - 11:50 am [Zoom](https://ucla.zoom.us/j/5622766807?pwd=bXFXUU0xc3dGdVVKeXlkQ3FvbUdTUT09)
* Lionel Levine Sec (1B) F 12:00 pm - 1:50 pm [Zoom](https://ucla.zoom.us/j/8323477632)
* Orpaz Goldstein Sec (1C) F 2:00 pm - 3:50 pm [Zoom](https://ucla.zoom.us/j/94436154375)

**Office hours (**Lionel Levine**):** Monday 8pm-10pm, [Zoom](https://ucla.zoom.us/j/8323477632)

**Office hours** (Orpaz Goldstein): Tuesday 8AM-10AM [Zoom](https://ucla.zoom.us/j/94436154375)

**Office hours** (Sajad Darabi): Wednesday 5PM-7PM [Zoom](https://ucla.zoom.us/j/5622766807?pwd=bXFXUU0xc3dGdVVKeXlkQ3FvbUdTUT09)

**Office hours (**Majid Sarrafzadeh**):** Tuesday 12:00 pm - 1:00 pm, [Zoom](https://ucla.zoom.us/my/majid1)

All the zoom links are also posted on **CCLE “Zoom Links”** section.

**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. Please make sure to sign up for Piazza forum at <http://piazza.com/ucla/spring2021/csm148>

**Grading:** Homework 15% 3 homework assignments  
 Late MIDTERM 40% 9th week of the class (90 minutes)  
 Projects (4) 40% Project 0 (2%), 3 more projects (10%, 10%, 18%)

Attendance 5%

Late policy: No late projects / homework will be accepted

There will be alternative plans for International students, who already have approval from the instructor, in terms of midterm and attendance. Project and Homework deadline for international students is all in pacific time (same as the rest of the class).

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.

**Gradescope entry code: 2RDN7K**  ([How to Enroll](https://help.gradescope.com/article/gi7gm49peg-student-add-course#:~:text=Existing%2520users%253A%2520If%2520you%2520already%2520have%2520a%2520Gradescope%2520account%252C%2520log,Then%2520enter%2520your%2520course%2520code.))

You may discuss problems with friends, but you must write your solutions individually. I expect all students to follow the [UCLA Student Conduct Code](http://www.deanofstudents.ucla.edu/Portals/16/Documents/144317_condeofcoduct_MarchV2_final%252520with%252520links.pdf), which prohibits cheating, fabrication, and multiple submissions.

**Schedule:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Topics** | **Assignment** | **Note** |
| Week 1 | | | |
| Mar 30th | Introduction |  |  |
| April 1st | Intro to Python | Project 0 Posted – Intro to Python |  |
| April 2nd | Discussion |  |  |
| Week 2 | | | |
| April 6th | Problem formulation and Data Wrangling | Lab 1: Intro to Python (numpy, graphing libraries, program structure, Jupyter Notebook) | Project 0 due |
| April 8th | Experimental Data Analysis | Lab 2: Python: sklearn, matplotlib  Project 1 Posted  HW1 Posted |  |
| April 9th | Discussion |  |  |
| Week 3 | | | |
| April 13th | kNN and Linear Regression | Lab 3: Scikit-learn for Simple Linear Regression | Project 1 due |
| April 15th | Linear Regression/Hypothesis Testing | Project 2 Posted | HW1 due |
| April 16th | Discussion |  |  |
| Week 4 | | | |
| April 20th | Multi and Poly Linear Regression | Lab 4: Multiple Linear Regression and Cross Validation |  |
| April 22nd | Model Selection & Cross Validation/ Regularization: Ridge & Lasso | HW2 posted |  |
| April 23rd | Discussion |  |  |
| Week 5 | | | |
| April 27th | Logistic Regression | Lab 5: Logistic Regression |  |
| April 29th | Logistic Regression / Multi-class Logistic Regression |  |  |
| Apr 30th | Discussion |  |  |
| Week 6 | | | |
| May 4th | kNN classification and SVM | Lab 6: KNN Classification & Imputation | Project 2 due |
| May 6th | Decision Trees | Project 3 posted | HW2 due |
| May 7th | Discussion |  |  |
| Week 7 | | | |
| May 11th | Bagging | Lab 7: Trees and Random Forests |  |
| May 13th | Neural Network Design | HW3 Posted |  |
| May 14th | Discussion |  |  |
| Week 8 | | | |
| May 18th | NN: Design/Back Propagation/ Regularization | Lab 8: Boosting |  |
| May 19th | Unsupervised (clustering) |  |  |
| May 20th | Discussion |  |  |
| Week 9 | | | |
| May 25th | Visualization |  |  |
| May 27th | Exam (there will be an evening session for students in other time zones) |  |  |
| May 28th | Discussion | Exam Solutions |  |
| Week 10 | | | |
| June 1st | Ethics and Gender Bias in Data Science |  | Project 3 due |
| June 3rd | Case Study |  | HW3 due |
| June 4th | - |  |  |

**\*The project/Homework deadlines are 12:01 am (For example, Project 1 is due on Tuesday April 13th – a minute after midnight on Monday the 12th) and is accepted all the way till 9:59 am before the class starts on Tuesday.**

**SHORT VERSION**

**Violence and Discrimination Resources**

UCLA prohibits gender discrimination, sexual harassment, domestic and dating violence, sexual assault, and stalking. If you have experienced any of these, there are a variety of campus resources to assist you, including a confidential hotline where you can talk to someone 24/7: (310) 825-0768.

Title IX is a federal civil rights law in the United States that was passed as part of the Education Amendments of 1972. Therefore **international students have access to the same services** at the Title IX office as non-international students.

Please note that **faculty and TAs are** **responsible employees**, which means faculty, TAs, and other UC employees are required under the UC Policy on Sexual Violence and Sexual Harassment to inform the Title IX Coordinator—a non-confidential resource —should they become aware that you or any other student has experienced sexual violence or gender discrimination.

**Confidential Resources:** You can make an appointment with a **CARE** Advocate ([careprogram.ucla.edu](http://careprogram.ucla.edu/)) or call their 24/7 hotline: (888) 200-6665. CARE offers free and confidential services for students, staff and faculty impacted by sexual assault, relationship violence and stalking. Counseling and Psychological Services **(CAPS)** ([counseling.ucla.edu](http://counseling.ucla.edu/)) also offers confidential crisis counseling and 24/7 support at (310) 825-0768. You can also receive confidential off-campus emergency medical services, advocacy, and counseling at the **Rape Treatment Center UCLA Medical Center Santa Monica**. Their 24/7 hotline is (424) 259-7208.

**Non-Confidential Reporting:** You can report sexual violence or sexual harassment directly to the University's Title IX Office, 2255 Murphy Hall, [titleix@conet.ucla.edu](mailto:titleix@conet.ucla.edu), (310) 206-3417. Reports to law enforcement can be made to UCPD at (310) 825-1491. Please note that any appointment you make with a Title IX representative is NOT confidential.