Lectures and labs

Lecture 1 (Wednesday, January 24): Introduction
Data collection, cleaning, storage, retrieval, learning, visualization.
Readings: CASI Epilogue, DSB Ch 1-3.
Science and data science, David M. Blei and Padhraic Smyth, PNAS 2017.
Optional: Large-scale physical activity data reveal worldwide activity inequality, Tim Althoff, Rok Sosič, Jennifer L. Hicks, Abby C. King, Scott L. Delp, and Jure Leskovec, Nature 2017.
Lab 1: Tableau and Kaggle
Homework 1: due February 5.

Lecture 2 (Wednesday, January 31): Supervised learning, fitting.
Readings: DSB Ch 4-5.
Optional: CASI Ch 8.
Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States, Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei, PNAS 2017.
Lab 2: Python, NumPy, scikit-learn

Lecture 3 (Wednesday, February 7): Unsupervised learning, clustering.
Readings: DSB Ch 6.
Optional: Robust continuous clustering, Sohil Atul Shah and Vladlen Koltun, PNAS 2017.
Lab 3: TabPy and scikit-learn
Homework 2: due February 19.

Lecture 4 (Wednesday, February 14): Performance measures.
Readings: DSB Ch 7-8.
Optional: CASI Ch 12.
Computer-based personality judgments are more accurate than those made by humans, Wu Youyoua, Michal Kosinski, and David Stillwella, PNAS 2015.

Lecture 5 (Wednesday, February 21): Dimensionality reduction.

Lecture 6 (Wednesday, February 28):

Readings: DSB Ch 9.

Lecture 7 (Wednesday, March 7):

Spring Recess (Monday-Sunday, March 12-18)

Lecture 8 (Wednesday, March 21):

Lecture 9 (Wednesday, March 28):

Lecture 10 (Wednesday, April 4):

Lecture 11 (Wednesday, April 11):

Lecture 12 (Wednesday, April 18):

Lecture 13 (Wednesday, April 25):

Lecture 14 (Wednesday, May 2):

Last day of Spring 2018 classes (Monday, May 7)

Office Hours & Contact Info

CDS 620
Wednesday 12-2pm: Lecturer, Iddo Drori, idrori@nyu.edu
Tuesday, 11am-1pm: Section Leader, Datta Sainath Dwarampudi, ddattasainath@nyu.edu
Friday 2-4pm: Grader, Samhita Damotharan, sd2941@nyu.edu
Thursday 2-4pm: Grader, Sai Anirudh Kondaveeti, sak797@nyu.edu

Textbooks

Computer Age Statistical Inference: Algorithms, Evidence and Data Science, Bradley Efron and Trevor Hastie, Cambridge University Press, 2016.
Data Science for Business, Foster Provost and Tom Fawcett, O'Reilly Media, 2013.
Introduction to Computation and Programming using Python, with Application to Understanding Data, John V. Guttag, 2nd Edition, MIT Press, 2016.

Grading

Homework 30%
Term Project 25%
Midterm 20%
Final Exam 25%