

Data Science Fundamentals

Spring 2021

Week 8 Announcements

Exam coming up next Thursday. Project 3 has been assigned! It will be due on June 1. HW3 is also now out. It will be due on June 3rd.

Course Information

- Lecture: Tuesdays and Thursdays, 10:00AM-Noon PST.
- Discussion: Fridays, 12:00pm-2:00pm. Link [Here](#)
- Office Hour: Mondays, 8:00PM-10PM PST Link [Here](#)

Description

A fundamental question that will be addressed is: given data arising in real-world, how does one analyze that data so as to understand the corresponding phenomenon. The course teaches critical concepts and skills in computer programming related to statistical inference, in conjunction with hands-on analysis of real-world datasets, including economic data, health data, geographical data, and social networks.

We will cover topics in machine learning and data analytics. We learn about the two basic kinds of statistical models, which have classically been used for prediction. We also cover clustering methodologies. We then cover Feature Selection, feature Engineering, and Data Pipelines. We explore more sophisticated model evaluation approaches (cross-validation and bootstrapping) with the goal of understanding how we can make our models as generalizable as possible.

Sections

- 1 Introduction: What Is Data Science? Modeling.
- 2 Statistical Inference, Exploratory Data Analysis, and the Data Science Process
- 3 Machine Learning Algorithms
- 4 Spam Filters, Naive Bayes, and Wrangling
- 5 Logistic Regression
- 6 Time Stamps and Financial Modeling.
- 7 Extracting Meaning from Data
- 8 Recommendation Engines: Building a User-Facing Data Product at Scale
- 9 Data Visualization
- 10 Causality
- 11 Data Engineering