

Data Science Decal

Fall 2017

Introduction:

Welcome to the Data Science Decal! In this course, you will discover how to analyze and manipulate data in Python, go over (and implement!) fundamental and practical statistical and machine learning algorithms, as well as learn how to ask the right questions in order to tackle data-driven problems. The course content targets an audience who has experience programming and understand calculus, though motivated and interested students without a strong technical background are encouraged to apply. No matter your background, data science and machine learning are infiltrating your field. More and more data is being created in many different places, and we hope to give you the skills to begin to make sense of it all.

Contact information:

Website : <https://ml.berkeley.edu/decals/DSD>

Email : dsd@ml.berkeley.edu

* The website will be our main form of communication and will contain links to extra content and assignments. Please check it regularly.

Units:

2 Units

Time and location:

Time : Tuesday 5:00-7:00 PM

Location : Genetics and Plant Biology 100

Office hours:

TBD; OH for facilitators are also available by appointment

Prerequisites:

This class is a projects-based class with a machine learning bias. You are expected to have some programming or statistics backgrounds and so the material will be of greatest benefit to sophomores or those who have taken CS61A, DATA 8, STAT 133, or equivalent. However, the first week of class will be a python bootcamp. By the end of it, you can determine whether you are comfortable continuing through the course.

Note that this is not an easy class. The student facilitators intend to provide you with a comprehensive guide to data analysis with the goal of preparing you for industry and, if demonstrated superb interest, future machine learning competitions.

Grading:

- 60% Projects (20% each)
- 40% Homeworks (10% each)
- Sufficient attendance (see attendance section below)

In order to pass the class, you must meet the attendance requirement and **earn at least a 70% cumulative score** on the projects and homework

Projects:

There will be 3 projects in this course, due roughly every 3 weeks. The purpose of the projects is to give you hands on experience manipulating, analyzing, and modeling data, with an emphasis on written explanation and communication. You will work in groups of 3-4.

Homeworks:

There will be 4 homework assignments, assigned in between projects. Homeworks will be completed individually, with an emphasis on coding.

Attendance

We will be keep track of attendance. In order to pass the course, **you must come to AT LEAST 75% of the lectures (that is, 9 lectures at minimum)**. After your 3rd missed day of class, excused or unexcused, you will automatically be assigned a “no pass”.

Class Schedule:

Week 1 - Introduction to the class; Python bootcamp

Week 2 - Data science 101: cleaning, manipulating, and visualizing data

Week 3 - Linear regression: estimation, interpretation, and testing

Week 4 - Logistic regression for classification; regularization

Week 5 - Decision trees and random forests; bias/variance tradeoff

Week 6 - Good practices for machine learning; support vector machines

Week 7 - Neural networks: feedforward networks; implementation with Keras

Week 8 - More neural networks: convolutional and recurrent networks

Week 9 - Amazon web services and working with big data

Week 10 - Dimensionality reduction: principal component analysis, SVD, nonlinear methods

Week 11 - Clustering: k-means, k-medoids, hierarchical, graphs

Week 12 - Special topics TBD

Extra Resources:

Most of the material will be in slides and demos. With each lecture, however, we will provide you with a plethora of extra resources for those interested in learning far more than what we can cover. Stay tuned!

Assignment Schedule:

* Assignments will be due at 11:59 PM on their respective due dates. No late work will be accepted.

Homework 1 - “Python and Data Visualization”

Assigned: 09/12/17

Due : 09/19/17

Project 1 - “Data Science 101: data munging, visualization, and regression”

Assigned: 09/19/17

Due : 10/03/17

Homework 2 - “Hyper-parameter tuning with SVMs”

Assigned: 10/03/17

Due : 10/10/17

Project 2 - “Classification: logistic regression with regularization, and trees”

Assigned: 10/10/17

Due : 10/24/17

Homework 3 - “James’ Keras”

Assigned: 10/31/17

Due : 11/07/17

Project 3 - “BYO feed-forward, convolutional, and recurrent networks”

Assigned: 11/07/17

Due : 11/21/17

Homework 4 - “Clustering”

Assigned: 11/21/17

Due : 11/28/17