Elements of Data Science: A First Course
Spring 2022

Time: Tuesday 6:10pm-8:45pm

Instructor: Haiyuan Wang Ph.D.

Textbook: Python Data Science Handbook by Jake VanderPas

Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow:

Concepts, Tools, and Techniques to Build Intelligent Systems (2nd Edition) by Aurelien Geron

Prerequisite(s):

- Introductory programming class as well as basic familiarity with Python 3.
- Basic familiarity with the command line.

Course Description

This course is designed as an introduction to elements that constitute the skill set of a data scientist. The course will focus on the utility of these elements in common tasks of a data scientist, rather than their theoretical formulation and properties. The course provides a foundation of methodology with applied examples to analyze large engineering, business, and social data for data science problems. Hands-on experiments with Python will be emphasized.

Topics include:

- Python Data Science Tools
- Data Cleaning, Exploration and Visualization
- Hypothesis Testing and Statistical Modeling
- Classification, Regression and Clustering
- Dimensionality Reduction and Topic Modeling
- Model Evaluation and Model Selection
- Feature Engineering and Feature Selection
- Natural Language Processing
- Data processing and delivery using ETL and APIs
- Dealing with Time Series Data
- Recommendation Engines

Assignments and Grading

Weekly Quiz	20%
Homework Assignments (Four, equally weighted at 7.5% each)	30%
Two Projects	30%
Final Exam	20%
TOTAL	100%

Quality of Performance	Letter Grade	Range %	GPA/Quality Pts.
Excellent - work is of exceptional quality	A+	99 - 100	4.33
	A	93 - 98.99	4.0
	A-	90 - 92.99	3.67
Good - work is above average	B+	87 - 89.99	3.33
Satisfactory	B	83 - 86.99	3.0
Below Average	B-	80 - 82.99	2.67
Poor	C+	77 - 79.99	2.33
	C	73 - 76.99	2.0
	C-	70 - 72.99	1.67
	D	65 - 69.99	1.0
	D-	60 - 64.99	0.67
Failure	F	< 60	0.0

Weekly Outline

Week	Topic	Readings	ToDo
1	Introduction to Data Science Problems and Tools		
2	Python and Numpy		
3	Pandas, Visualization and Data Exploration		
4	Hypothesis Testing		
5	Intro to Machine Learning		
6	Machine Learning Models / Project 1		
7	Model Evaluation and Selection		
8	Data Cleaning and Feature Engineering		

9	Joining Data, Dimensionality Reduction and Imbalanced Classes		
10	NLP, Sentiment Analysis and Topic Modeling		
11	Clustering and Recommendation Systems / Project 2		
12	Timeseries, Data Processing and Delivery		
13	Databases and Review / Final		