Machine Learning in Python

Spring 2020

Instructor:	Ramzi Saouma	**Time:**	$00:00 - 00:00$
Email:	Ramzi@Saouma.org	**Office Hours:**	.

Course Pages: Course website to be updated here.

Main References: This is a list of recommended texts which we will use continuously during the course. Some texts demand a higher level of mathematical pre-requisites as compared to the rest, but all convey the essential concepts and ideas eloquently.

- Witten, Gareth James, Robert Tibshirani, and Trevor Hastie, *An Introduction to Statistical Learning: With Applications in R*, Springer, 2017.

- Wes Mckinney, *Python for Data Analysis, 2e*, O'Reilly, 2017.

- Peter Flach, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, Cambridge University Press, 2012.

- Tony Guida, *Big Data and Machine Learning in Quantitative Investment*, Wiley, 2019.

Recommended Pre-Course Reading: The following article presents a broad overview of the developments and milestones for the past 50 years of data science.

David Donoho, *50 years of Data Science.* Based on a presentation at the Tukey Centennial workshop, Princeton NJ, September 18 2015. Available online at:
https://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf

Objectives: This course is primarily designed for graduate students to gain exposure to the concepts and methods revolving around statistical and machine learning, with emphasis on real-life applications to finance. Broadly speaking, this course is designed to enable students to appreciate the broad range of machine learning methodologies as well as the techniques to implement them in Python. During the first four weeks each session will be split in two parts. The first part is dedicated to Python's essentials, while the second part focuses on Machine learning concepts. Once we reach week 5, students should have acquired enough skills in Python to start applying it to Machine learning.

Prerequisites: An undergraduate-level understanding of probability, statistics, graph theory, algorithms, and linear algebra is assumed.

Tentative Course Outline:

| Week 1:
Part I: Introduction to Python

- Getting familiar with Jupyter

- Semantics and Objects

- Types and Operators

- Control Flow

Part II: Introduction to Machine Learning

- The Anti-Spam Case

- Introducing Case I: Trader Profiling

| Week 2: Part I: Python Data Structures

- Sequences

- Functions

Part II: Linear Regressions

- Simple Linear Regression

- Multiple Linear Regression

- Potential Problems

| Week 3:
Part I: Introducing NymPy: Arrays and Vectorised Computation
Part II: Classification Methods

| Week 4:
Part I: Getting Started with Pandas
Part II: Dreaming of Algorithms

| Week 5:
Data Visualisation in Python: Applied Plotting, Charting Data Representation in Python

| Week 6:

- Applied Supervised Learning

- Basic Classification and Regression Problems
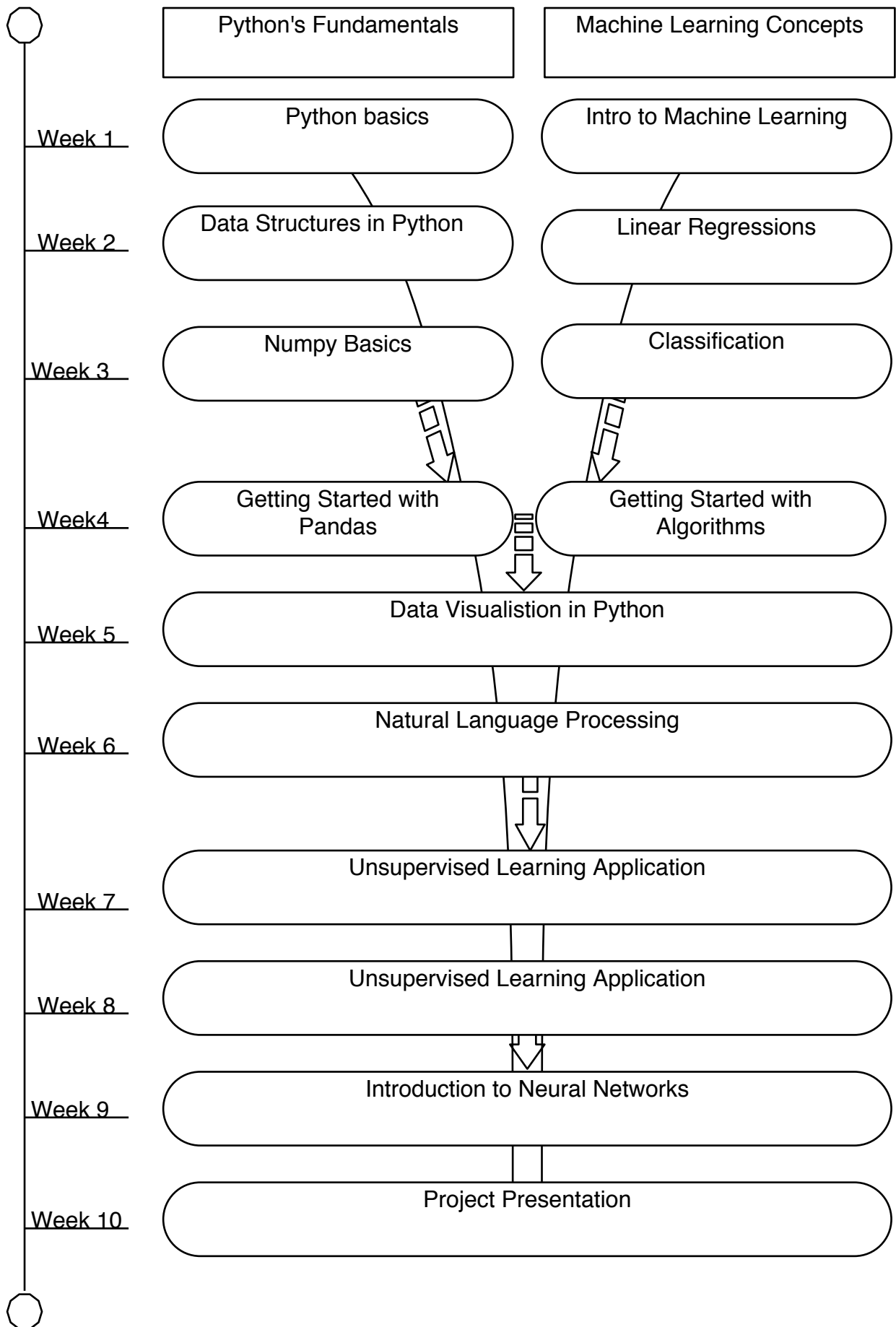
| Week 7:
Advanced Supervised Methods

| Week 8:
Model Evaluation

| Week 9:
Case Study: Predicting House Prices

| Week 10:
Final Project Presentations

	Python's Fundamentals	Machine Learning Concepts
Week 1	Python basics	Intro to Machine Learning
Week 2	Data Structures in Python	Linear Regressions
Week 3	Numpy Basics	Classification
Week4	Getting Started with Pandas	Getting Started with Algorithms
Week 5	Data Visualistion in Python	
Week 6	Natural Language Processing	
Week 7	Unsupervised Learning Application	
Week 8	Unsupervised Learning Application	
Week 9	Introduction to Neural Networks	
Week 10	Project Presentation	

Grading Policy: Homework (50%), Final Project (50%)