Machine Learning Tutorial

Week1-Part 1

Introduction

Yasin Ceran

Anaconda Python

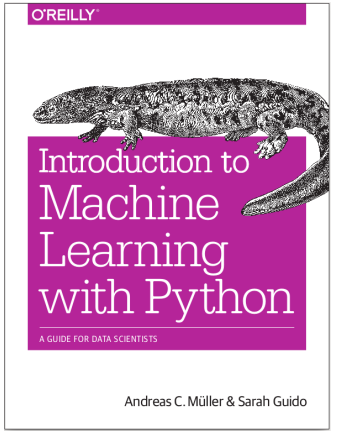
- First, download Anaconda. We recommend downloading Anaconda's latest Python 3 version. https://www.anaconda.com/distribution/
- Second, install the version of Anaconda which you downloaded, following the instructions on the download page.
- Congratulations, you have installed Jupyter Notebook!

Python Knowledge

- Familiarity with Python programming and basic use of NumPy, pandas and matplotlib.
- A good reference is the Python Data Science Handbook by Jake VanderPlas.
- It's online for free and available as a notebook at the link below. I highly recommend going through it before starting the class.

https://github.com/jakevdp/PythonDataScienceHandbookß

Book

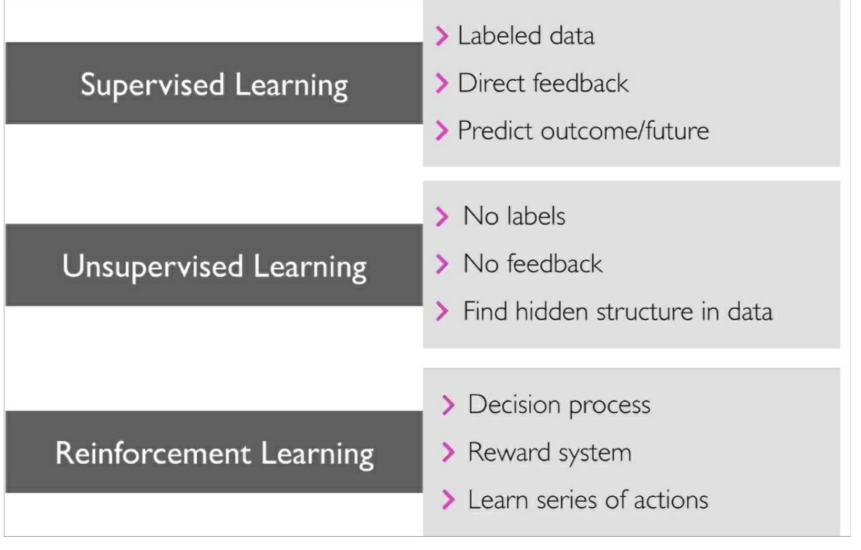


Most of the slides and lecture content is based on the class notes and the textbook of Dr. Andreas Mueller.

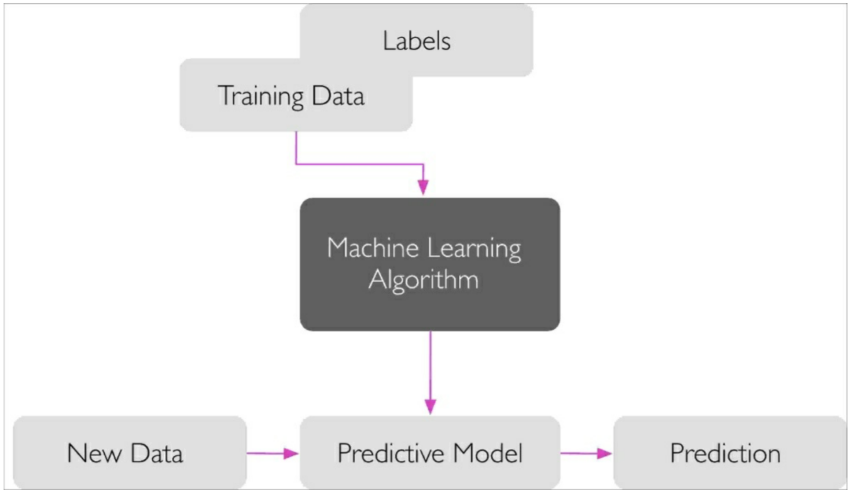
What is Machine Learning

- Large amount of structured and unstructured data
- Machine Learning helps capturing the knowledge from the data to improve the performance of predictive models and make data-driven decisions

Types of Machine Learning



Supervised Learning



Supervised Learning

$$(x_i, y_i) \propto p(x, y) \text{ i.i.d.}$$

$$x_i \in \mathbb{R}^p$$

$$y_i \in \mathbb{R}$$

$$f(x_i) \approx y_i$$

Classification and Regression

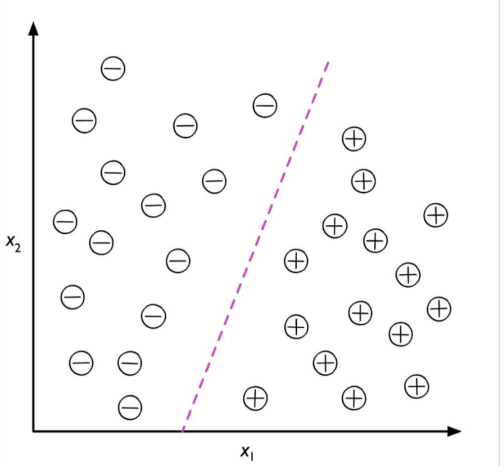
Classification

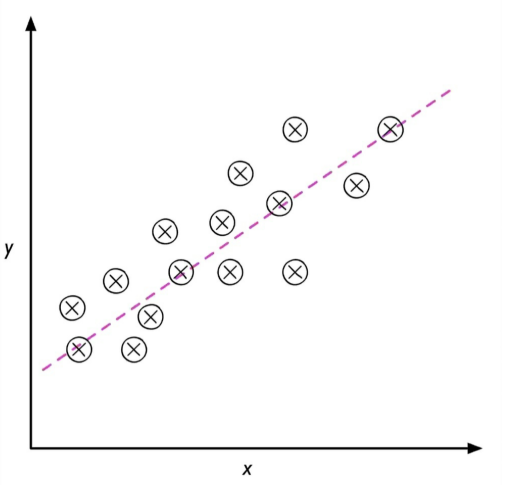
- target y discrete
- Will you pass?

Regression

- target y continuous
- How many points will you get in the exam?

Classification and Regression





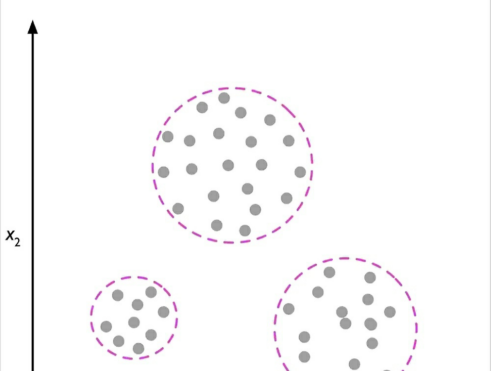
Generalization

Not only $f(x_i) \approx y_i,$

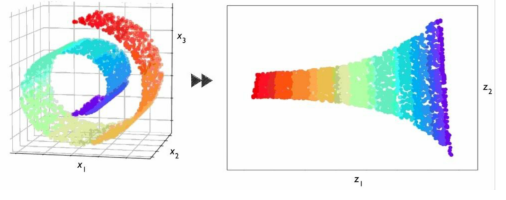
also for new data: $f(x) \approx y$

Unsupervised Learning

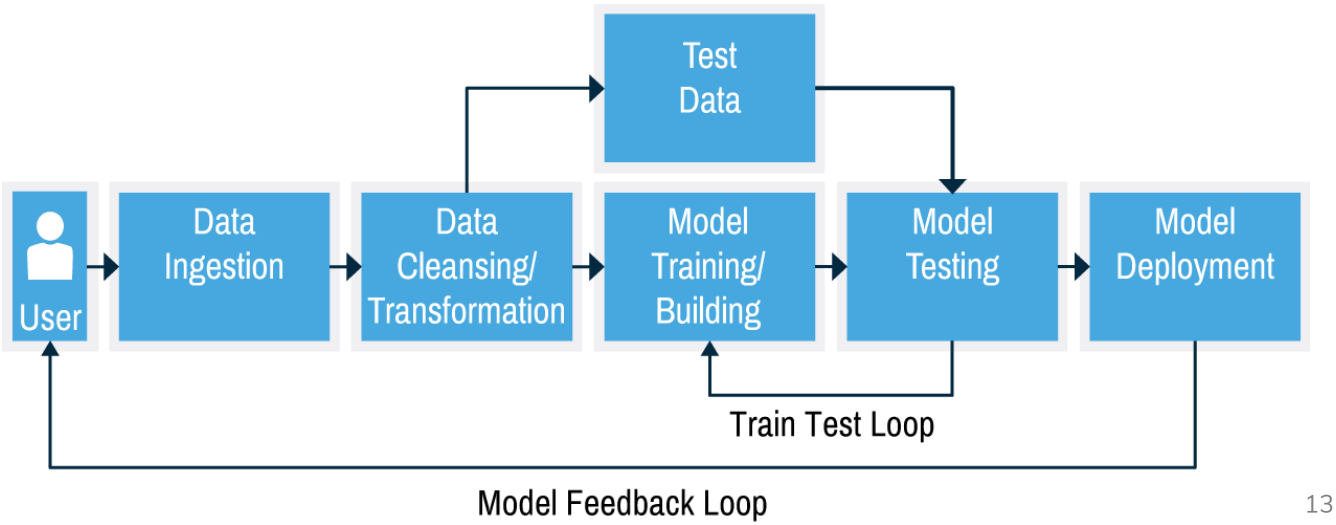
- Clustering is an explanatory data analysis technique



- Dimentionality reduction is used to remove noise and compress data

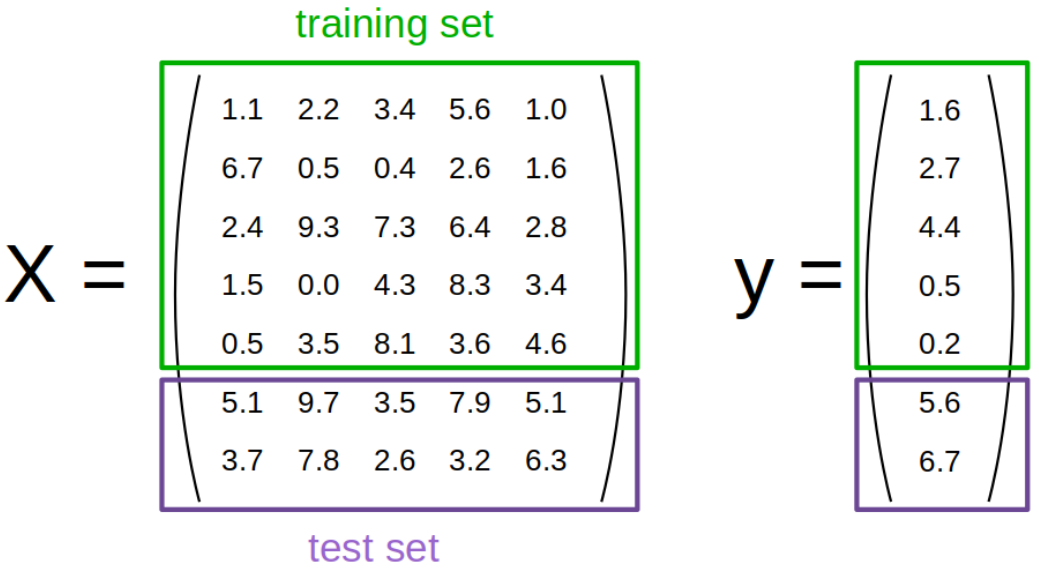


The Machine Learning Work-Flow



Representing Data

Training and Test Data



Jupyter Notebook

Part 1- Data Loading

Questions ?