Chapter 1
Introduction

COMP3314
Machine Learning

Outline

- Motivation
- Types of ML
  - Supervised Learning
    - Classification
    - Regression
  - Reinforcement Learning
    - Chess
  - Unsupervised Learning
    - Clustering
    - Dimensionality Reduction
- Terminology and Notation
- Roadmap
  - Preprocessing
  - Learning
  - Evaluation and Prediction
- Python
  - Installation
- Linear Algebra Review
- References

Motivation

- Nowadays large amount of structured and unstructured data is available
- ML algorithms can turn this data into knowledge
  - Powerful open source libraries available to do this

- In this course you will understand how these algorithms work
- You will also learn how to utilize them to make predictions

Motivation

- ML algorithms are self-learning
  - Automatically derive knowledge from data to make predictions
    - No need for humans to manually derive rules
    - ML offers a more efficient alternative for capturing the knowledge in data to gradually improve the performance of predictive models
- ML becomes increasingly relevant in CS research
  - More importantly
    - Plays an ever greater role in our everyday lives

How do you use machine learning everyday?

Examples of Machine Learning

- Basket analysis
- Credit scoring
- Medical diagnosis
- Biometrics
- Object recognition

Machine Learning Definition

- Subfield of Artificial Intelligence (AI)
- Arthur Samuel (1959)
  - Field of study that gives computers the ability to learn without being explicitly programmed
- Tom Mitchell (1998)
  - A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E

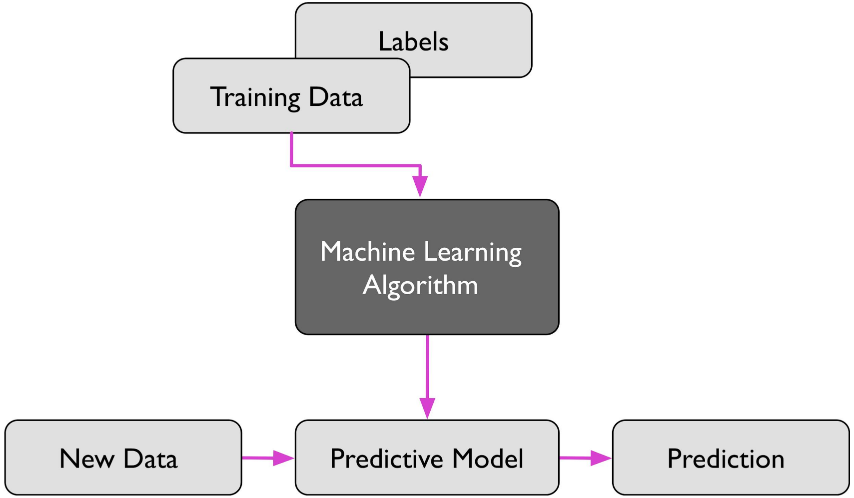
Types of machine learning

- In the following we will consider three types of machine learning

- Supervised Learning
  - Labeled data
  - Direct feedback
  - Predict outcome/future
- Unsupervised Learning
  - No labels/targets
  - No feedback
  - Find hidden structure in data
- Reinforcement Learning
  - Decision process
  - Reward system
  - Learn series of actions

Supervised Learning

- Learn from labeled training data
  - Make predictions about unseen / future data
- Supervised refers to a set of samples where the desired output signals (labels) are already known



Supervised Learning: Classification vs. Regression

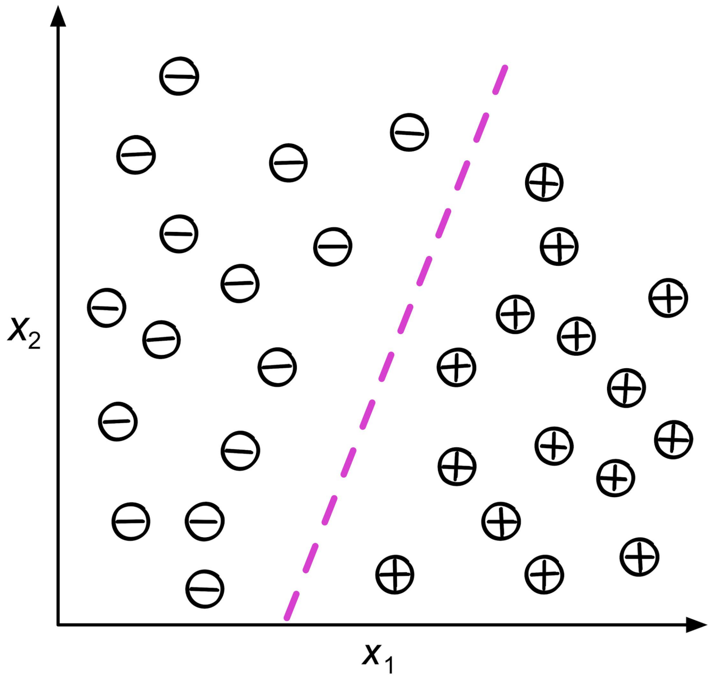
- Two subcategories of supervised learning
  - Classification
    - A supervised learning task with discrete class labels
      - E.g., spam email classifier
  - Regression
    - Outcome is a continuous value
      - E.g., student exam score prediction

Supervised Learning - Classification

- Goal: Predict class labels of new instances, based on past observations
- Class labels are discrete, unordered values
- Two subcategories of classifiers:
    - Binary classification
        - Only two possible class labels can be assigned
            - E.g., spam vs. non-spam emails
    - Multiclass classification
        - Any fixed number >2 of class labels can be assigned
            - E.g., handwritten digit recognition

Classification - Example

- Given 30 training samples
  - 15 labeled as negative class
  - 15 labeled as positive class
- Let each sample have 2 dimensions
- Classifier will learn the decision boundary
  - Represented as a dashed line
  - Able to separate the two classes

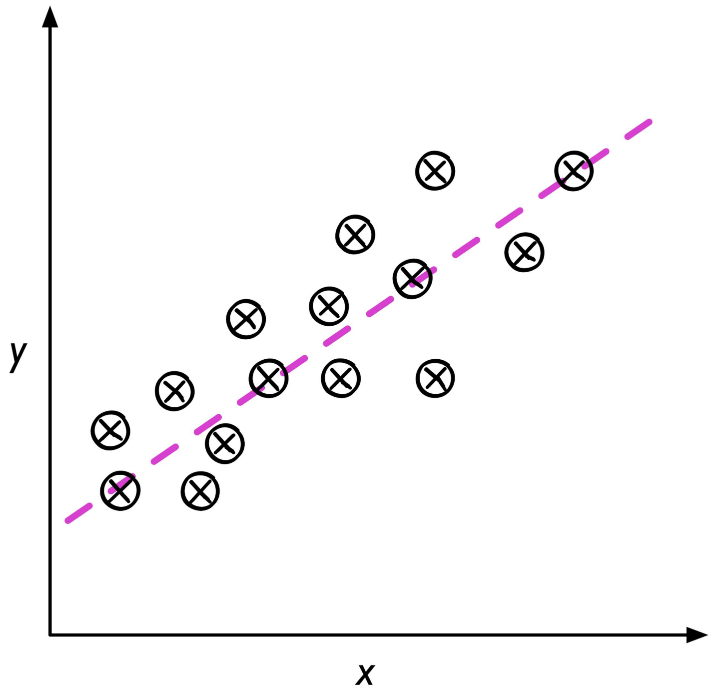


Regression

- Prediction of continuous outcome
  - The term regression was devised by Francis Galton in his article Regression towards Mediocrity in 1886
- Example:
  - Predicting the exam scores given time spent studying

Regression - Example

- Given
  - Predictor variable x
  - Response variable y
- I.e., 1D data set
- Fit a line to it minimizing the distance between sample points and the fitted line
  - Average squared distance is most commonly used
- Use the line to predict outcome of new data

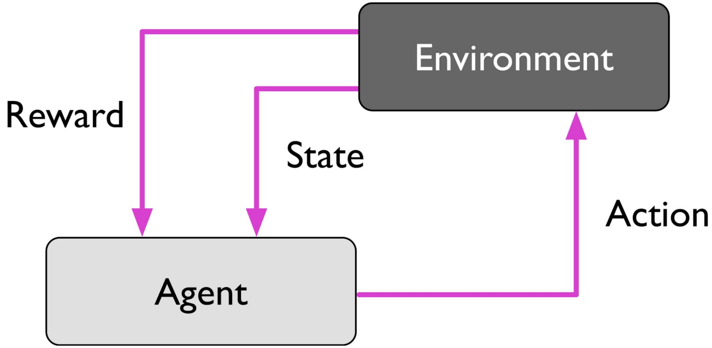


Quiz

- Consider the following supervised ML tasks. Label each task with *Classification Task* or *Regression Task*
  a. You are working for an investment bank and your task is to predict investors sentiment for certain stocks by analyzing popular online investments forums
  b. You are working for a property agency and your task is to predict the housing price for a property based on past data that the agency has available in their database
  c. Your task is to analyze a video stream of the western harbour tunnel and count how many Tesla pass by every day

Reinforcement Learning

- The system (aka agent) improves its performance based on interactions with an environment
- Trial-and-Error approach
  - Learning by doing
- The agent receives feedback (reward) from the environment
  - This reward is not the correct ground truth
    - It is a sample experience
  - Extensive interaction with the environment allows agent to learn a series of actions that maximizes this reward



Reinforcement Learning - Example: Chess

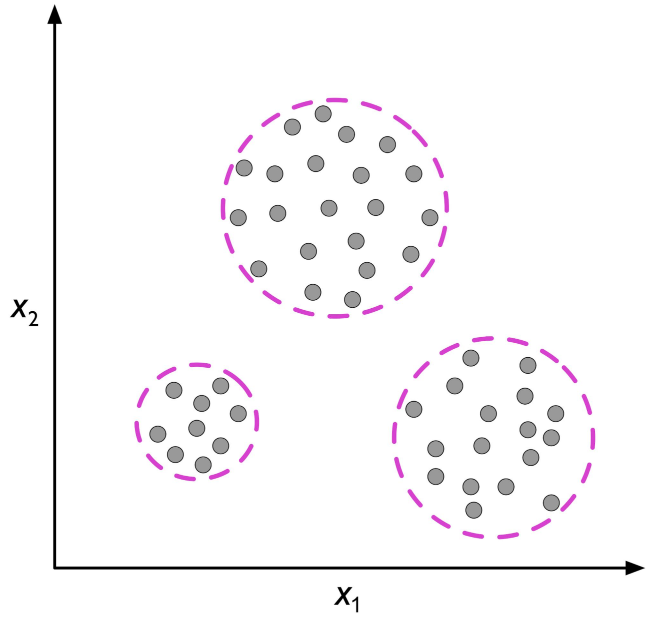
- Agent decides upon a series of moves depending on state of board
  - Environment is the board
  - Reward can be defined as win or lose at the end of the game
- Outcome of each move results in different state of the environment
  - Removing an opponent's chess piece from the board or threatening the queen is associated with a positive event
  - Losing a chess piece to the opponent is associated with a negative event
- Note: Not every turn results in the removal of a chess piece
  - Reinforcement learning is concerned with learning the series of steps by maximizing a reward based on immediate and delayed feedback

Unsupervised Learning

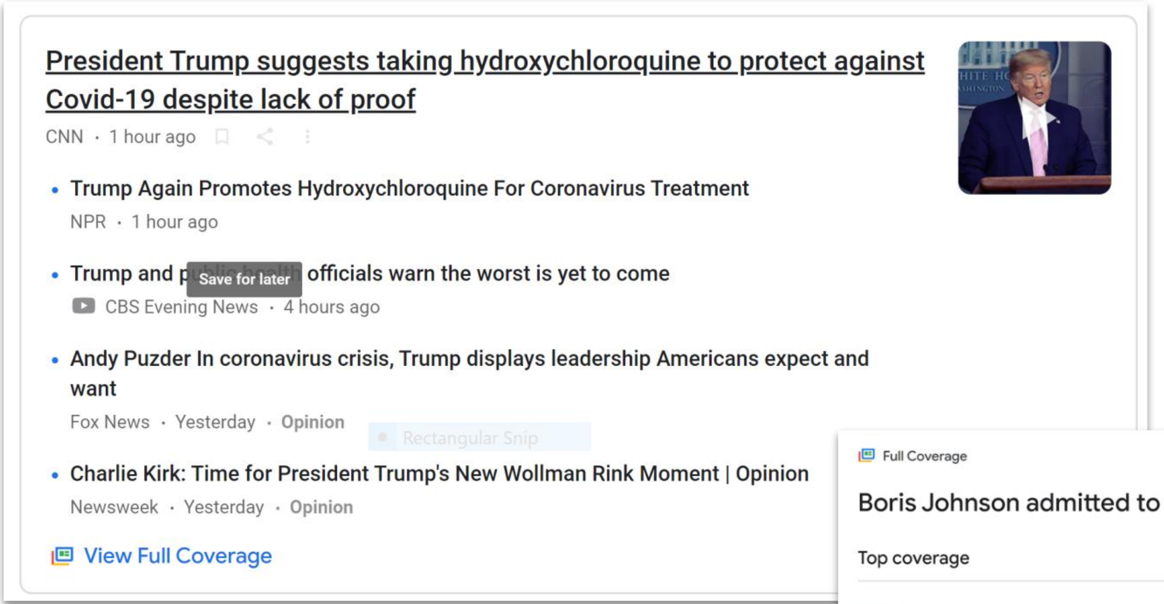
- Unlabeled data / data of unknown structure
- Explores the structure of data
  - Extract meaningful information without guidance of known outcome variable / reward function
- Examples
  - Clustering
  - Dimensionality reduction

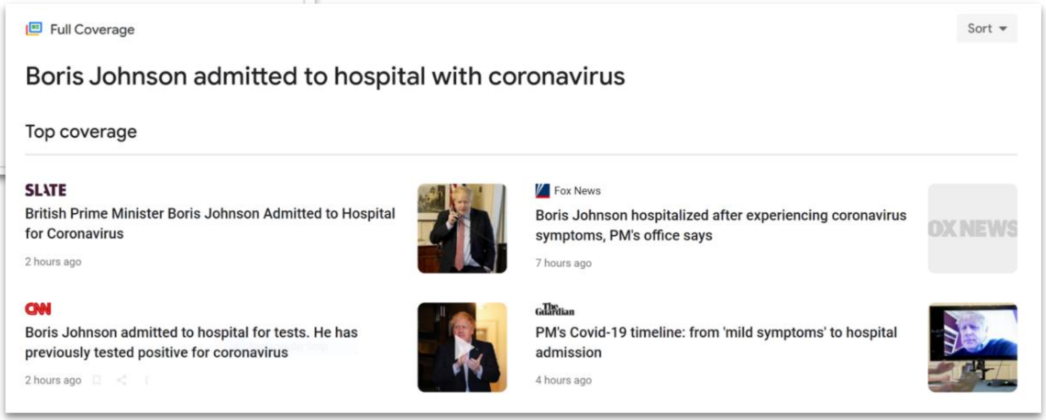
Clustering

- Exploratory data analysis technique
- Organizes information into meaningful subgroups (clusters) without having any knowledge of group memberships
- Each cluster defines a group of objects that share a certain degree of similarity but are more dissimilar to objects in other clusters



Clustering - Example





Clustering - More Examples

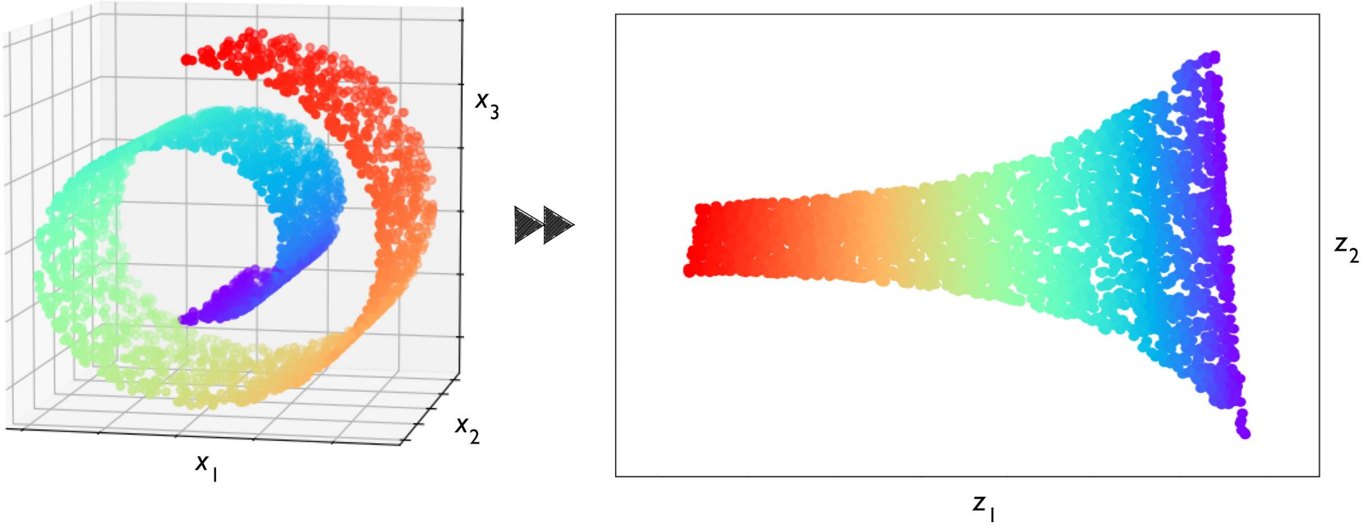
- Human genetic clustering
- Sequence clustering
- Social network analysis
- Market research
- Grouping of shopping items

Dimensionality Reduction

- Often we are working with data of high dimensionality
  - I.e., each observation comes with a high number of measurements
- High dimensional data can present a challenge
  - Computational performance
  - Predictive performance
  - Visualization
- Dimensionality reduction is a commonly used approach in feature preprocessing
  - Compress data onto a smaller dimensional subspace
  - Retaining most relevant information

Dimensionality Reduction - Example

- High-dimensional feature set can be projected onto 1D, 2D or 3D feature spaces
  - 3D to 2D example



Types of machine learning

- In the following we will consider three types of machine learning

Supervised Learning	Labeled data Direct feedback Predict outcome/future
Unsupervised Learning	No labels/targets No feedback Find hidden structure in data
Reinforcement Learning	Decision process Reward system Learn series of actions

Terminology and Notations

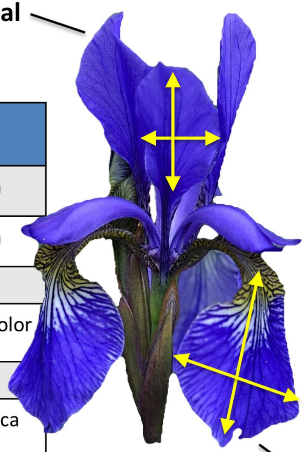
- Iris flower data set contains measurements of 150 Iris flowers from three different species
  - Setosa, Versicolor, and Virginica
- Introduced in Fisher's 1936 paper The use of multiple measurements in taxonomic problems
- Row
  - A single flower sample
- Column
  - Flower features (measurements in centimeters)

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			...		
50	6.4	3.5	4.5	1.2	Versicolor
			...		
150	5.9	3.0	5.0	1.8	Virginica

Features
(attributes, measurements, dimensions)

Class labels
(targets)



Terminology and Notations

- We will use a matrix and vector notation to refer to our data
- Each sample is a separate row in a feature matrix **X**, where each feature is stored as a separate column
- Iris dataset example
  - 150 samples and four features are written as a 150 x 4 matrix **X**

$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & x_4^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & x_4^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{(150)} & x_2^{(150)} & x_3^{(150)} & x_4^{(150)} \end{bmatrix}$$

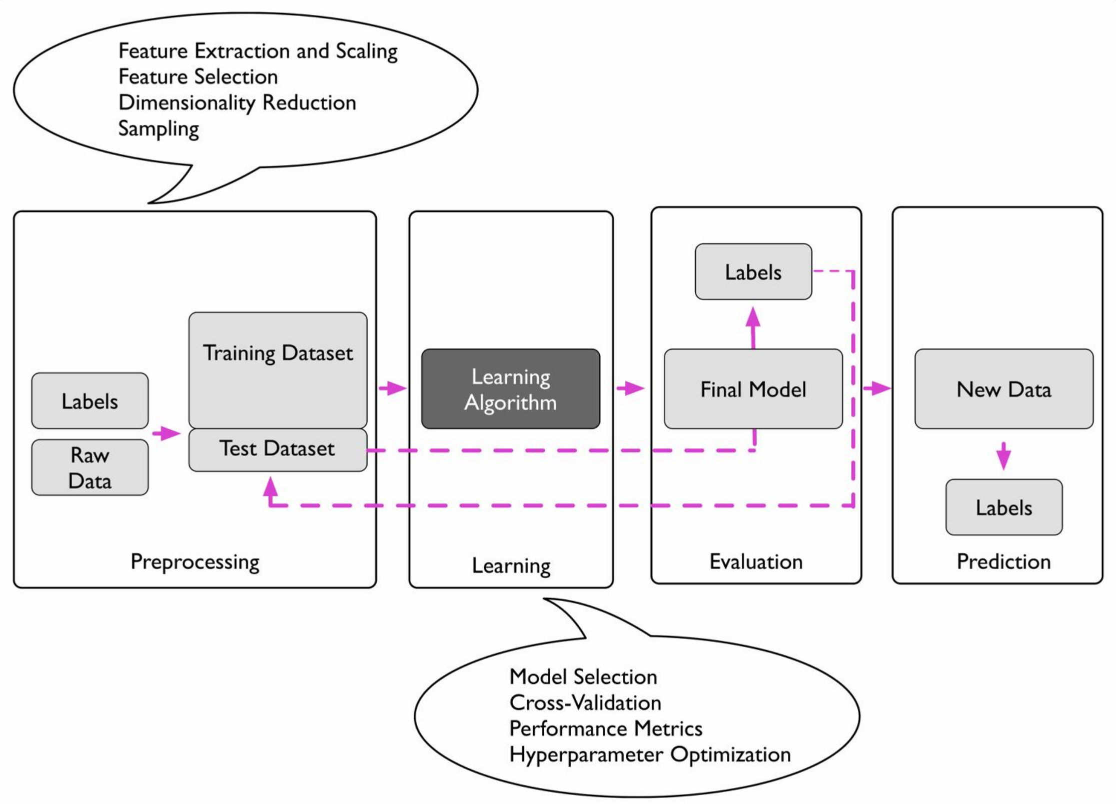
Terminology and Notations

- We will use the superscript *i* to refer to the *i*th training sample, and the subscript *j* to refer to the *j*th dimension of the dataset
  - For example $x_j^{(i)} = x_1^{(150)}$ refers to the first dimension of the flower sample 150
- We use lowercase, bold-face letters to refer to vectors and uppercase, bold-face letters to refer to matrices
- Note that each row in the iris dataset **X** can be written as a four-dimensional row vector and each feature dimension is a 150-dimensional column vector

$$\boldsymbol{x}^{(i)} = \begin{bmatrix} x_1^{(i)} & x_2^{(i)} & x_3^{(i)} & x_4^{(i)} \end{bmatrix} \qquad \boldsymbol{x}_j = \begin{bmatrix} x_j^{(1)} \\ x_j^{(2)} \\ \vdots \\ x_j^{(150)} \end{bmatrix}$$

Roadmap

- Typical workflow for using ML in predictive modeling



Preprocessing

- Preprocessing of the data is a crucial steps in any ML application
- Feature selection, extraction and scaling
  - Select and extract useful features from raw data
  - Many algorithms also require that the selected features are on the same scale
- Dimensionality reduction
  - May improve
    - Computational performance
    - Predictive performance
- Sampling
  - Randomly divide the dataset into a separate training and test set to determine whether our algorithm not only performs well on the training set but also generalizes well to new data
  - Keep the test set until the very end to evaluate the final model

Learning

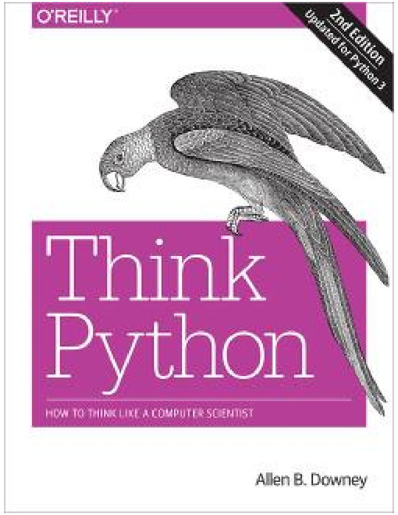
- Model selection
  - Compare algorithms and select the best performing model
- Cross-validation
  - How de we know which model performs well on the final test dataset if we don’t use this test set for model selection?
    - Cross-validation splits the training dataset further into training and validation subsets
- Performance metric
  - Decide upon a metric to measure performance
- Hyperparameter optimization
  - Fine-tune parameters of the model based on performance on validation set

Evaluation and Prediction

- After model selection and training we use the test dataset to estimate how well it performs on unseen data
  - Estimate the generalization error
- If we are satisfied with its performance, we can now use this model to predict new data

Python

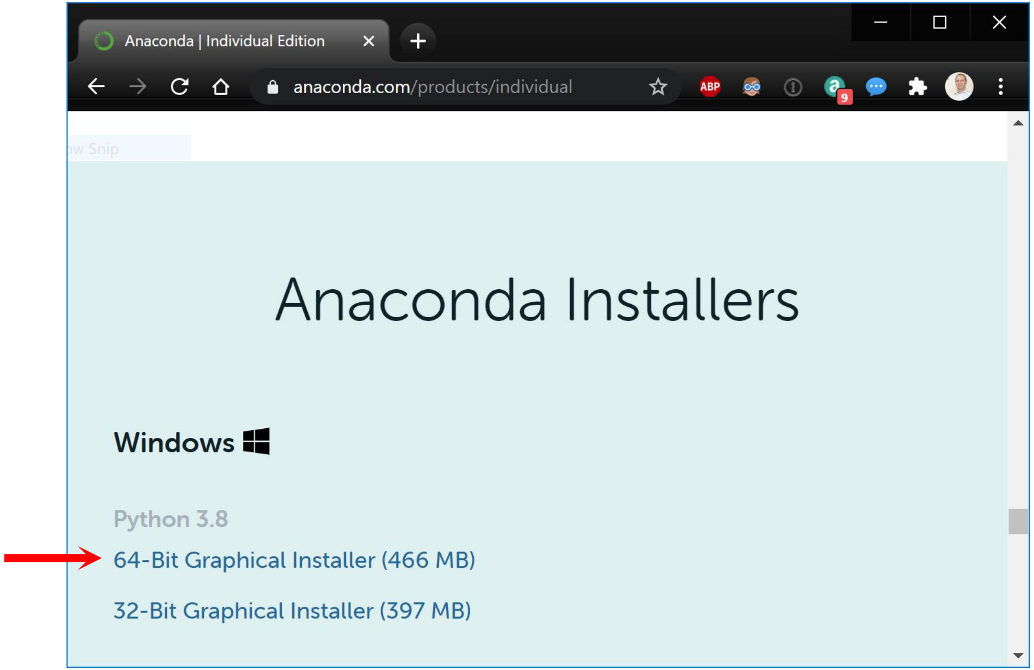
- We assume that you are familiar with the basics of python
  - Recommended textbook

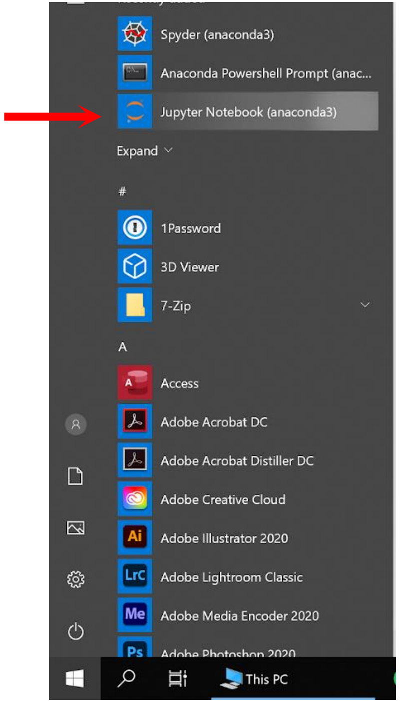


Free PDF [here]

Programming Environment: Local

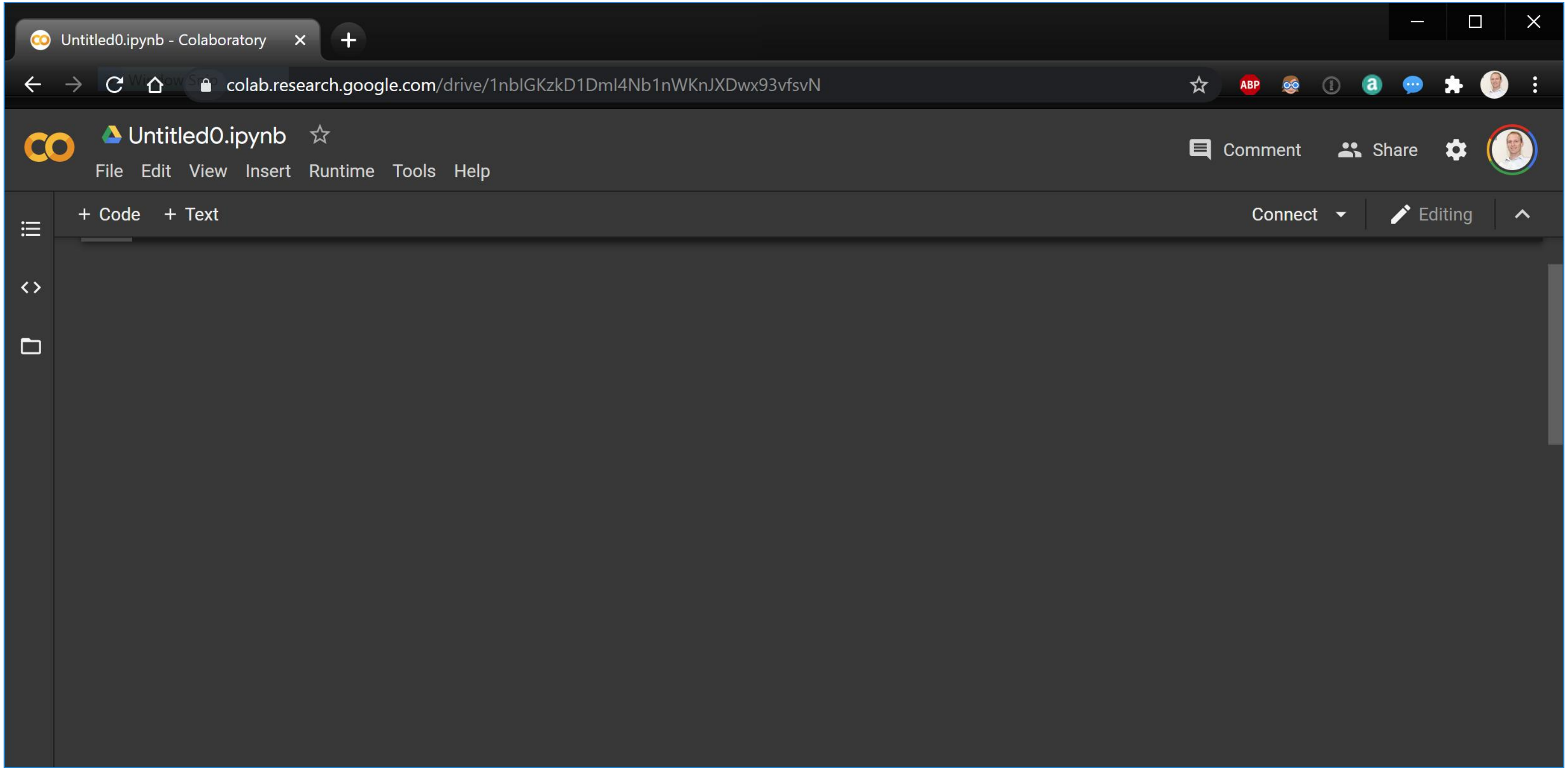
- In this course we are going to use
    - Python 3, NumPy, MatPlotLib, SciPy and Jupyter





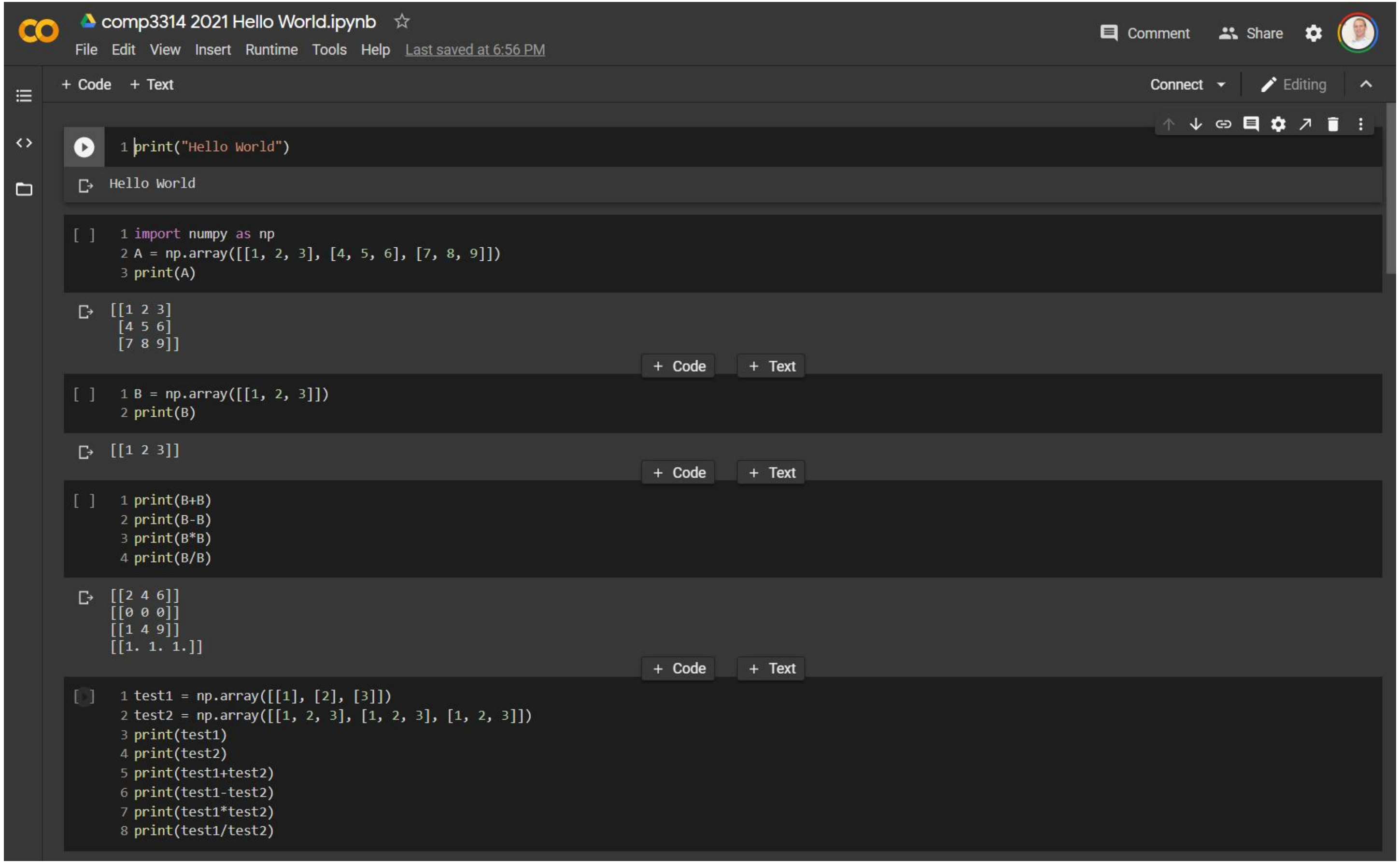
Programming Environment: Cloud

Google CoLab



Installation

Code is here



Linear Algebra Review

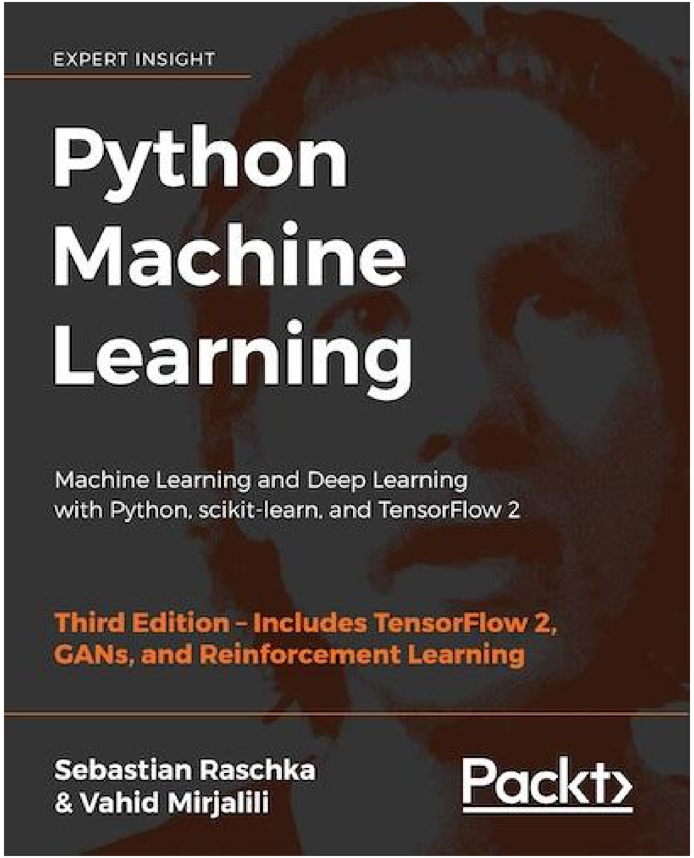
- We will only use basic concepts from linear algebra
- However, if you need a quick refresher, please take a look at Zico Kolter's excellent videos

Python Review

- We assume that your are familiar with the libraries/tools, follow these links if you need a refresher
  - NumPy
  - Pandas
  - Matplotlib
  - Jupyter

References

- Materials in this chapter are based on
  - Book
  - Code



References

- Some materials in this chapter are based on
  - Book
  - Code