Course Overview

Deep Learning

- It is now time to learn about one of the most fundamental libraries of using Python for Quantitative Analysis - NumPy!
- NumPy is already included in the environment file provided, but just in case

conda install numpy

Machine Learning Overview

Machine Learning

- Let’s discuss some basic machine learning concepts to set a foundation for future lectures!
- We’ll talk about supervised learning, unsupervised learning, reinforcement learning, evaluation methods, and more.

Machine Learning

- Unlike typical computer programs, machine learning techniques will iteratively learn from data.
- ML algorithms can find insights in data, even if they aren’t specifically instructed what to look for in the data.

Machine Learning

- In this course we'll discuss three major types of machine learning algorithms
    - Supervised Learning
    - Unsupervised Learning
    - Reinforcement Learning
- We'll also touch on other topics, such as word embeddings with Word2Vec

Supervised Learning

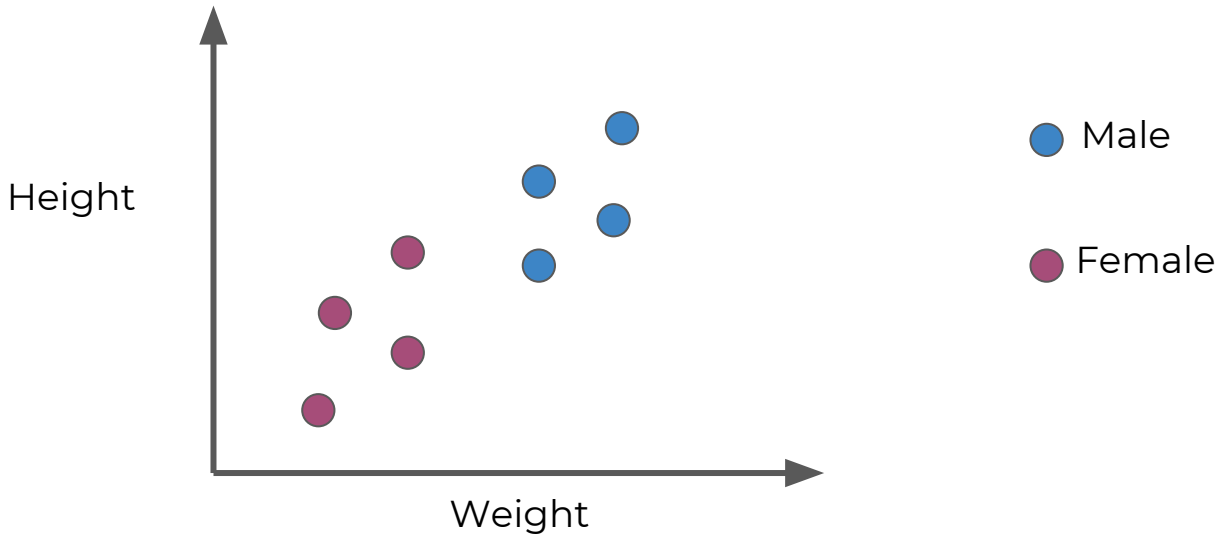
- Supervised Learning uses **labeled** data to predict a label given some features.
- If the label is continuous its called a regression problem, if its categorical it is a classification problem.

Supervised Learning

- Supervised Learning - Classification
  - Features: Height and Weights
  - Label: Gender
  - Task: Given a person’s height and weight, predict their gender.

Supervised Learning

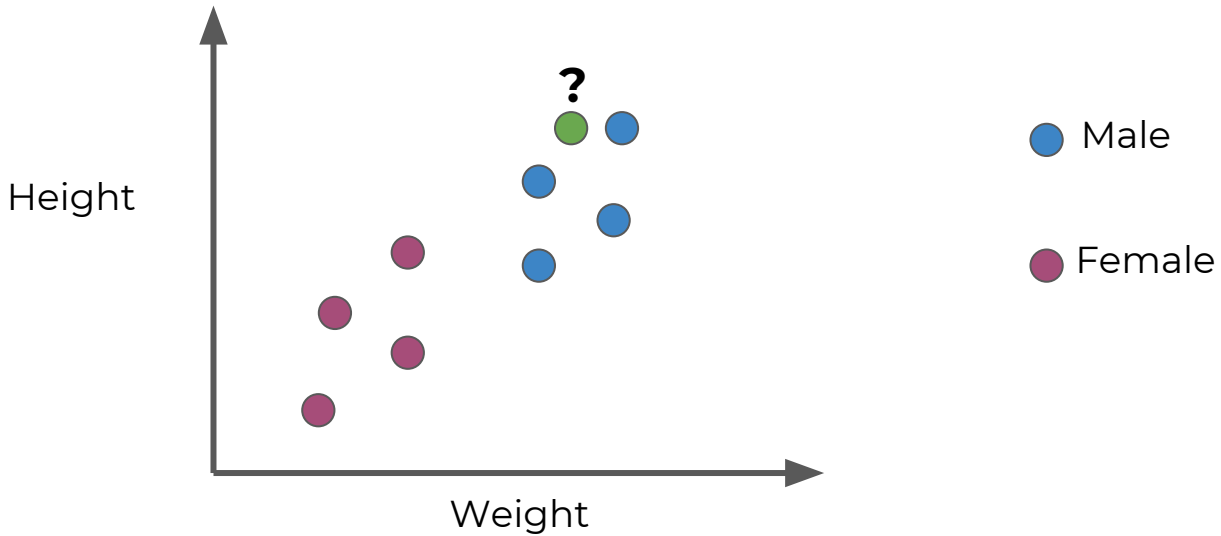
- Supervised Learning - Classification



PIERIAN DATA

Supervised Learning

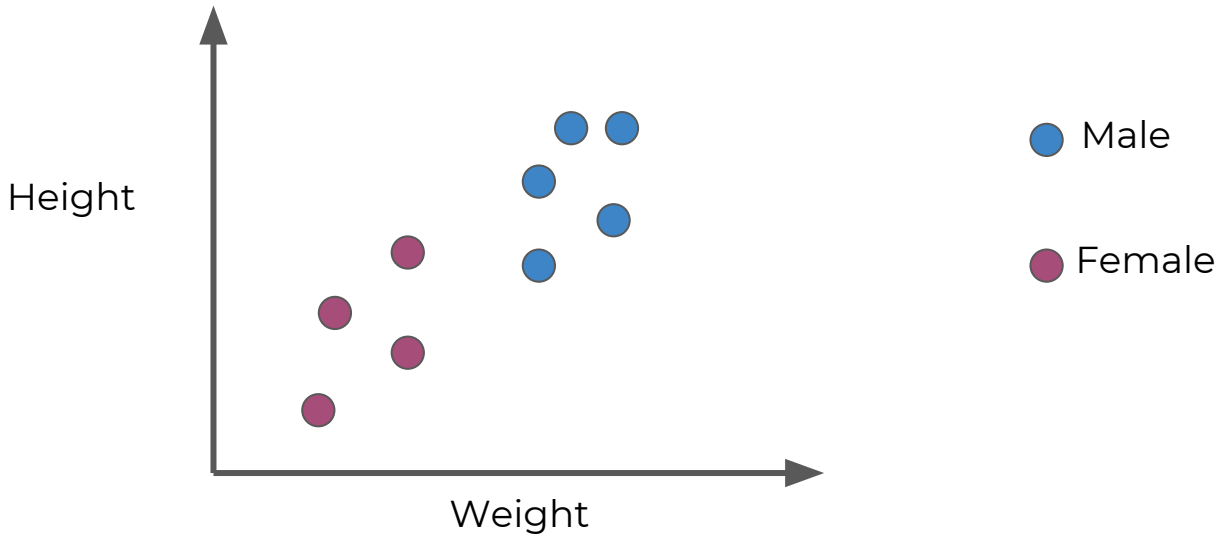
- Supervised Learning - Classification



PIERIAN DATA

Supervised Learning

- Supervised Learning - Classification

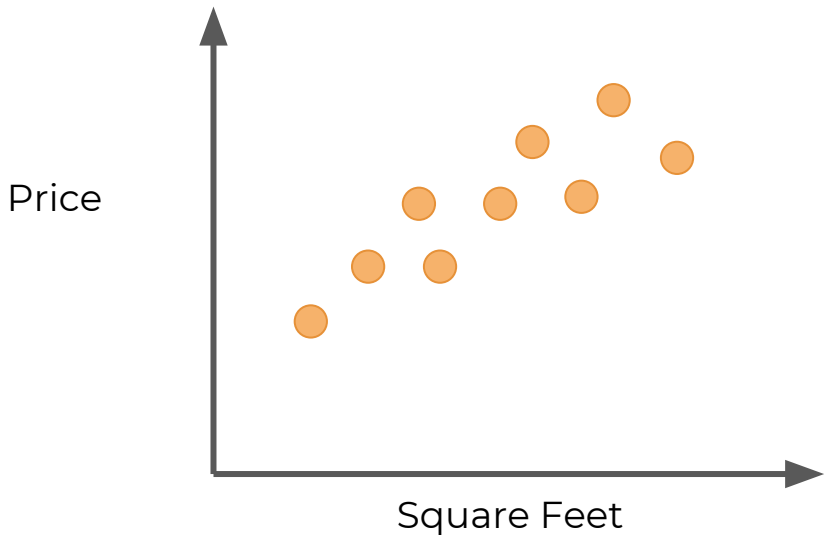


Supervised Learning

- Supervised Learning - Regression
  - Features: Square Footage, Rooms
  - Label: House Price
  - Task: Given a house’s size and number of rooms, predict the selling price.

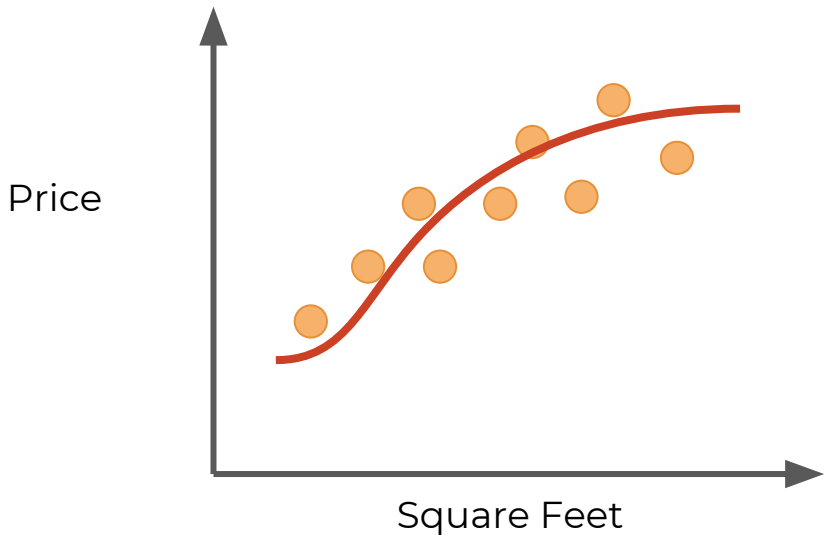
Supervised Learning

- Supervised Learning - Regression



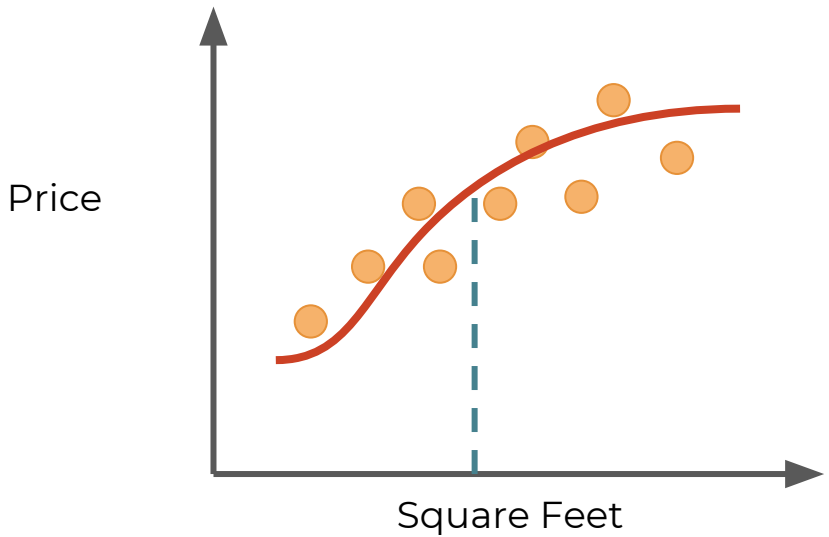
Supervised Learning

- Supervised Learning - Regression



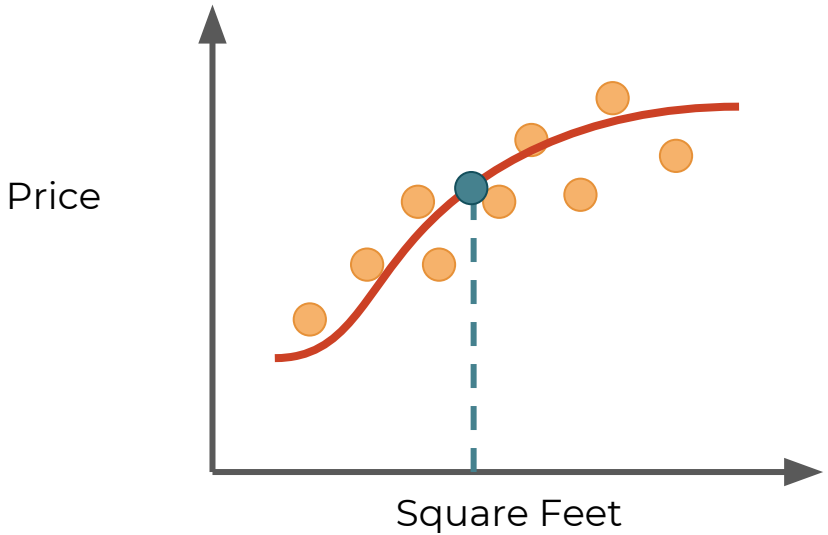
Supervised Learning

- Supervised Learning - Regression



Supervised Learning

- Supervised Learning - Regression



Supervised Learning

- Supervised Learning has the model train on historical data that is already labeled (e.g. previous house sales).
- Once the model is trained, it can then be used on new data, where only the features are known, to attempt prediction

Unsupervised Learning

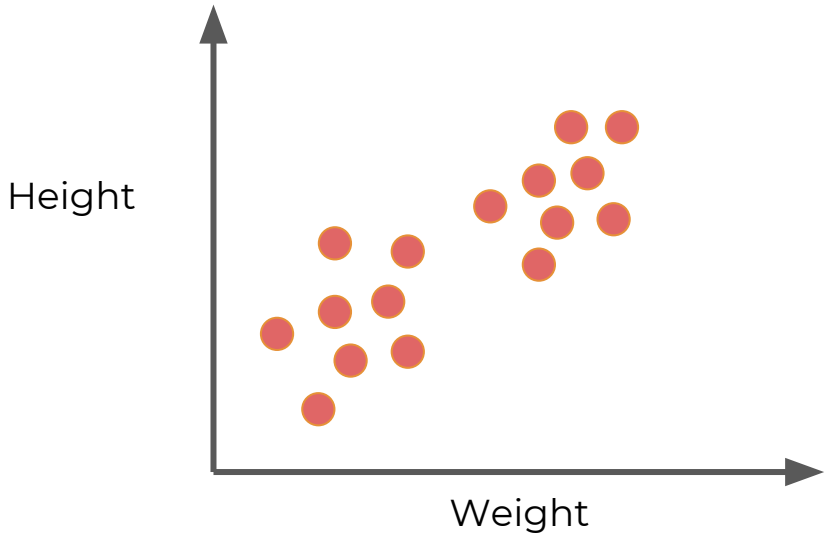
- But what if you don’t have historical labels for your data? (You only have features)
- Since you have no “right answer” to fit on, you need to look for patterns in the data and find a structure.

Unsupervised Learning

- Unsupervised Learning - Clustering
    - Features: Heights and Weights for breeds of dogs.
    - Label: **No Label for unsupervised!**
- Task: Cluster together the data into similar groups. It is then up to the data scientist to interpret the clusters.

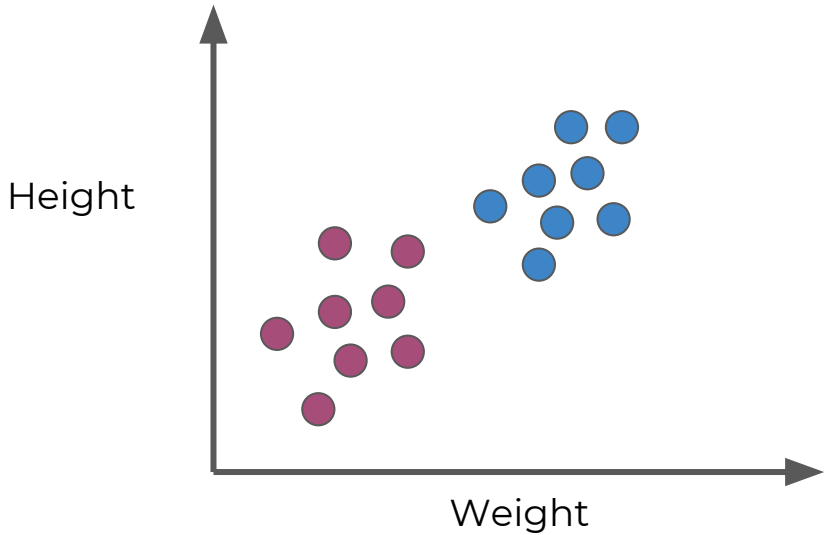
Supervised Learning

- Unsupervised Learning - Clustering



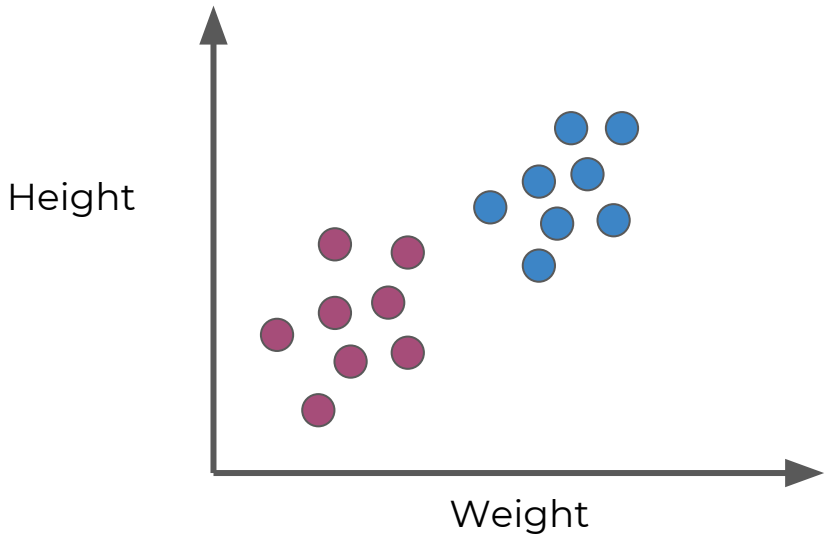
Supervised Learning

- Unsupervised Learning - Clustering



Supervised Learning

- Unsupervised Learning - Clustering



Clustering won't be able to tell you what the group labels should be.

Only that the points in each cluster are similar to each other based off the features.

Reinforcement Learning

- What about machine learning tasks like have a computer learn to play a video game, drive a car, etc... ?
- Reinforcement learning works through trial and error which actions yield the greatest rewards.

Reinforcement Learning

- Components
  - Agent-Learning/Decision Maker
  - Environment - What Agent interacts with
  - Actions - What the Agent can do

Reinforcement Learning

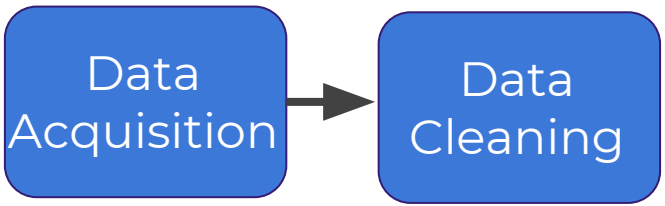
- The agent chooses actions that maximize some specified reward metric over a given amount of time.
- Learning the best policy with the environment and responding with the best actions.

Deep Learning

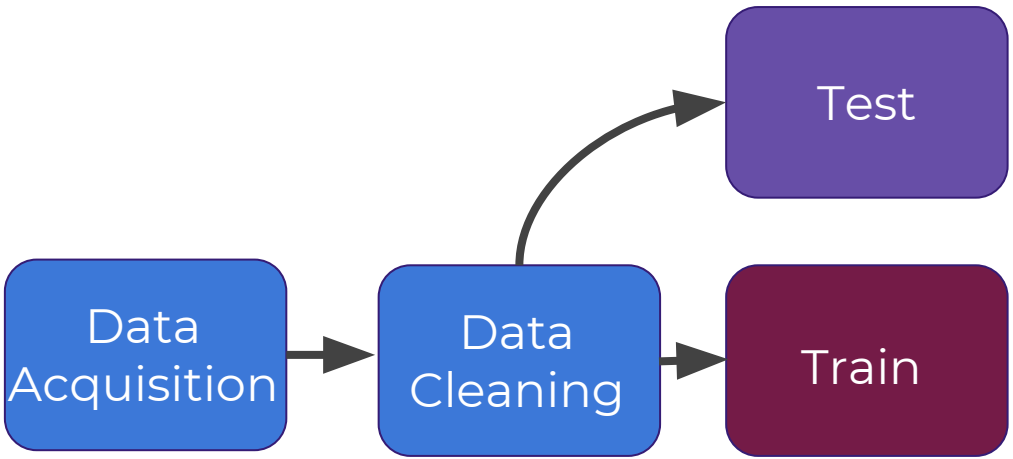
- Let’s walk through the basic machine learning process for a supervised learning problem.
- Afterwards we’ll discuss some key differences for unsupervised learning, as well discuss hold out data sets.

Acquire Data from Some Source

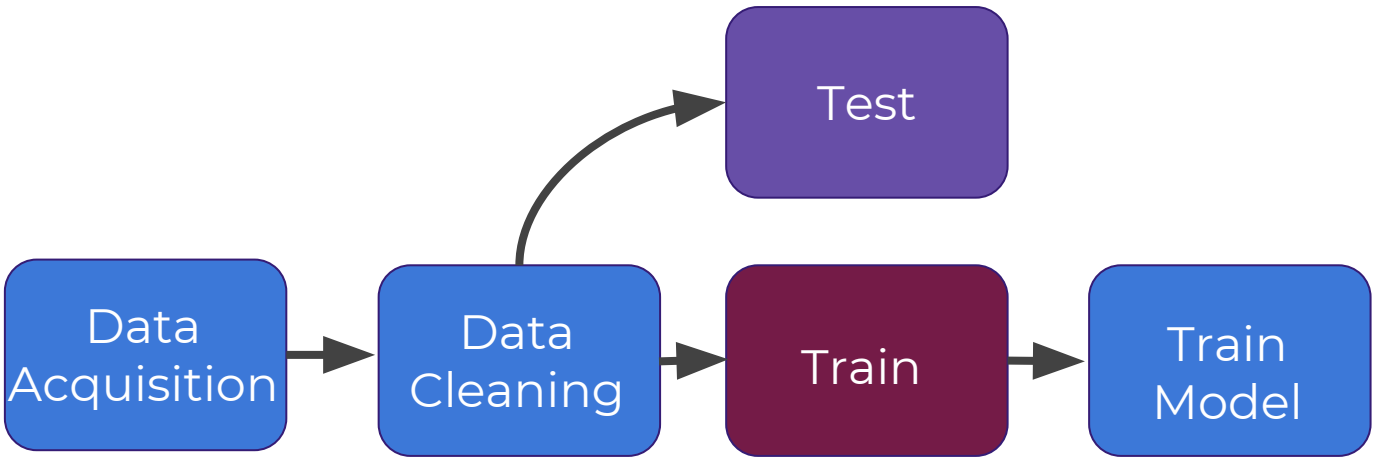
Clean and Organize the Data



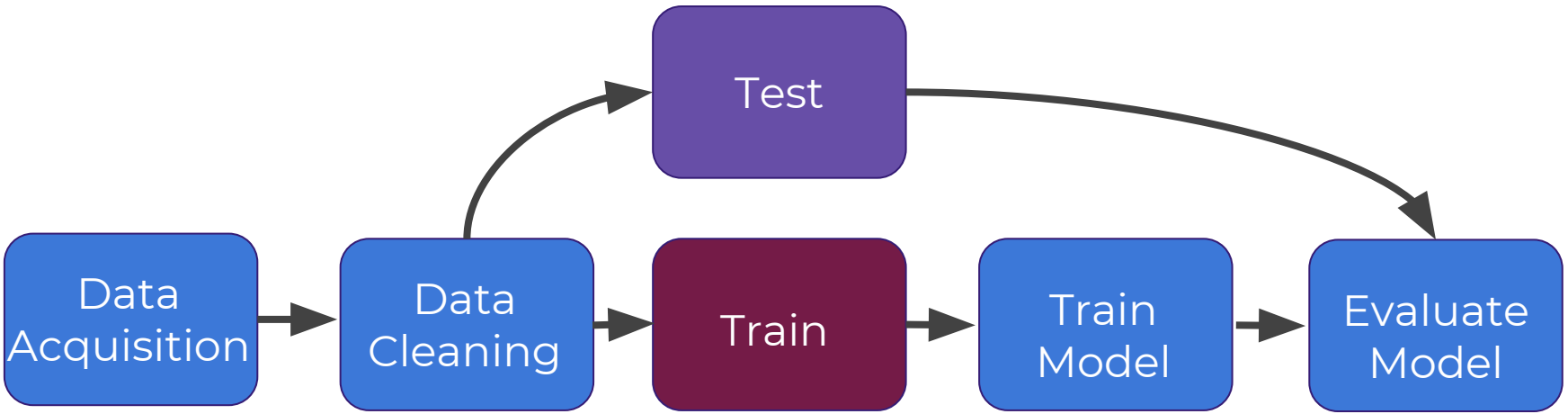
Train Test Split



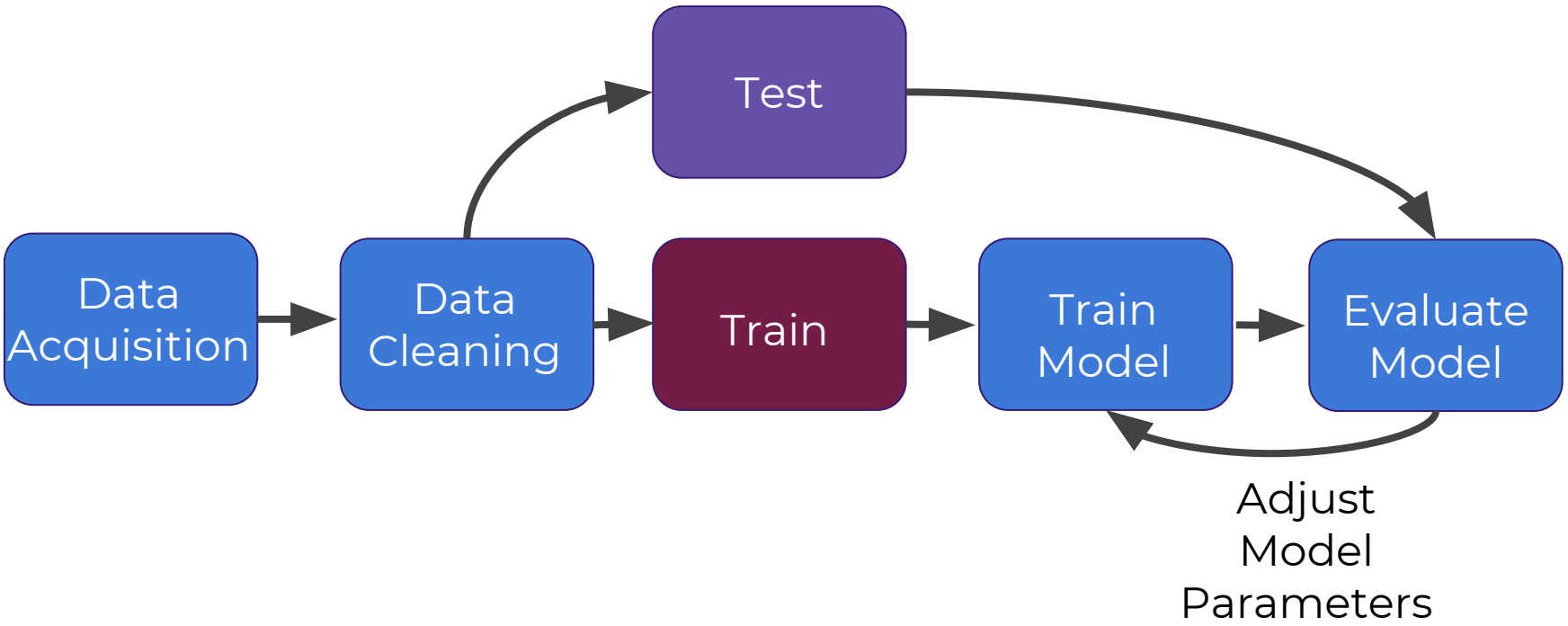
Train/Fit Model on Training Data



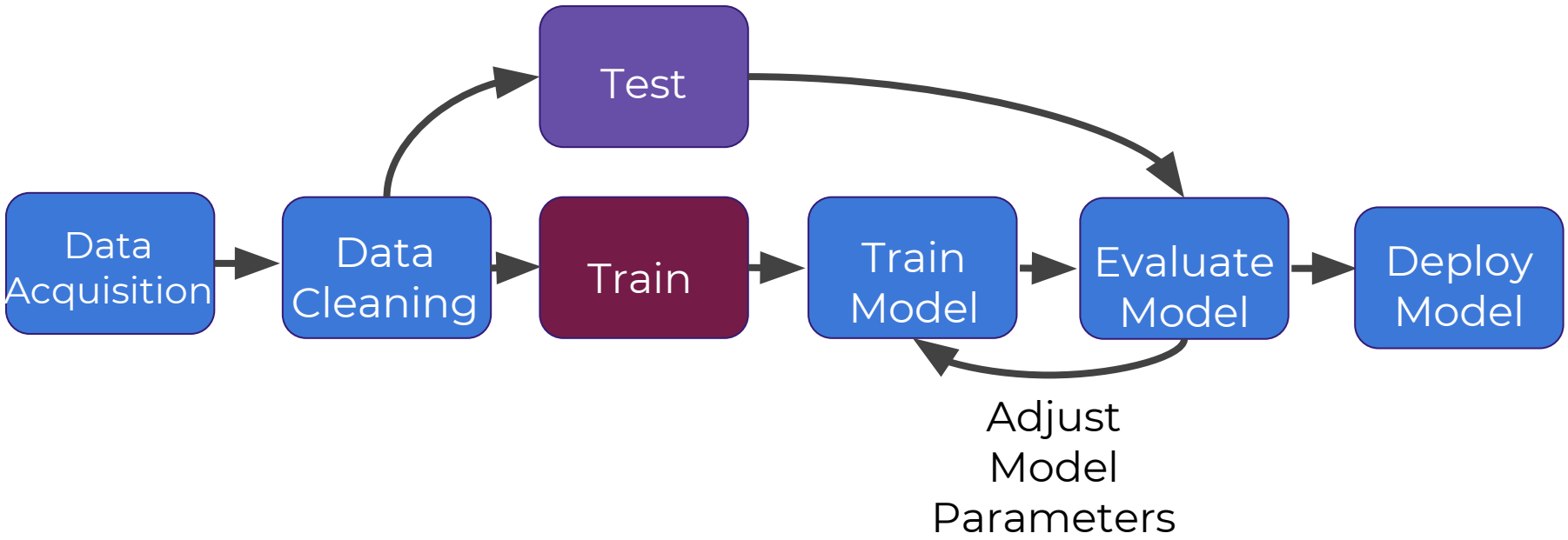
Evaluate Model on Test Data



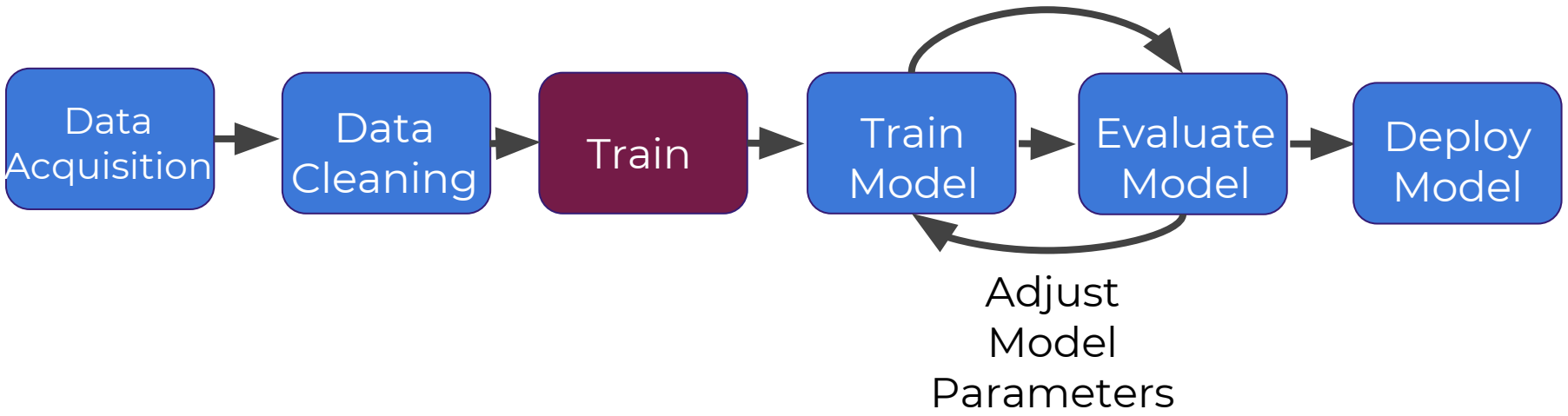
Adjust Model Parameters



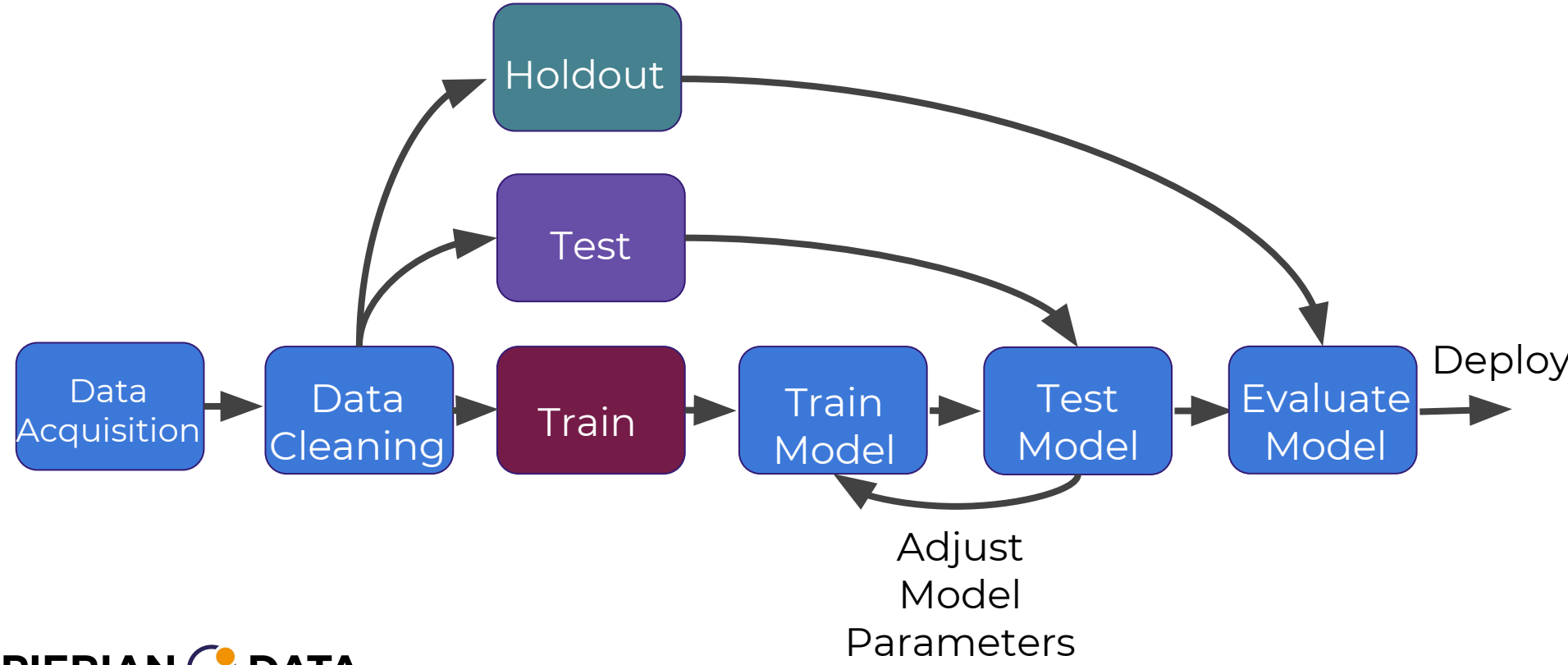
Deploy Model on New Incoming Data



Unsupervised Learning



Hold Out Sets



Deep Learning

- Finally let’s quickly discuss model evaluation, we’ll dive into more details for certain problems later on in the course.

Deep Learning

- Supervised Learning -Classification Eval
  - Accuracy , Recall, Precision
  - Accuracy - Correctly Classified divided by total samples.
- Which metric is the most important depends on the specific situation

Deep Learning

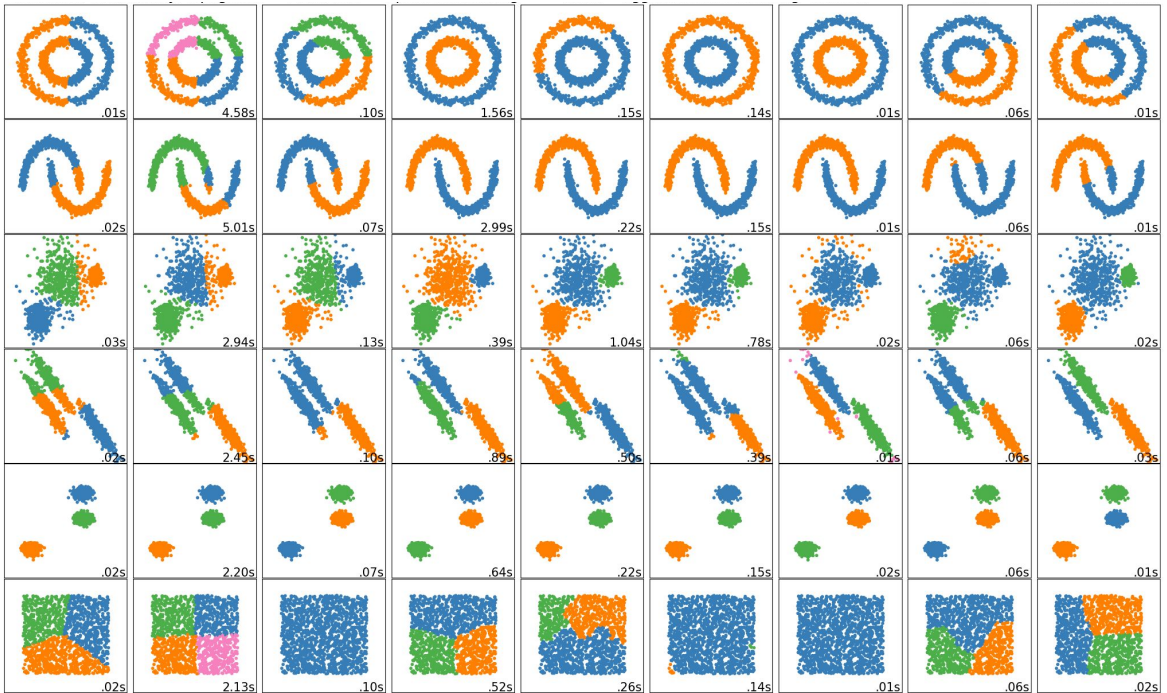
- Supervised Learning -Regression Eval
  - MAE, MSE, RMSE
  - All are measurements of: On average, how far off are you from the correct continuous value.

Deep Learning

- Unsupervised Learning - Evaluation
    - Much harder to evaluate, depends on overall goal of the task
    - Never had “Correct Labels” to compare to
    - Cluster Homogeneity, Rand Index

Deep Learning

- Unsupervised Learning - Evaluation



PIERIAN DATA

Deep Learning

- Reinforcement Learning - Evaluation
    - Usually more obvious, since the “evaluation” is built into the actual training of the model.
    - How well the model performs the task its assigned.

Deep Learning

- Review
  - Machine Learning
  - Types of Machine Learning
  - Machine Learning Process
  - Evaluation Metrics

Crash Course Basics

Deep Learning

- Very Basic Overview
    - NumPy
    - Pandas
    - Matplotlib and Pandas Data Viz
    - SciKit Learn Preprocessing

Deep Learning

- **Not** meant to be a comprehensive overview!
- Check out:
    - Python for Data Science and Machine Learning Bootcamp

NumPy Quick Overview

Pandas Quick Overview

Data Visualization Quick Overview

PIERIAN DATA

SciKit-Learn Quick Overview

PIERIAN DATA

Crash Course Review Exercises

Crash Course Exercise Solutions

PIERIAN DATA