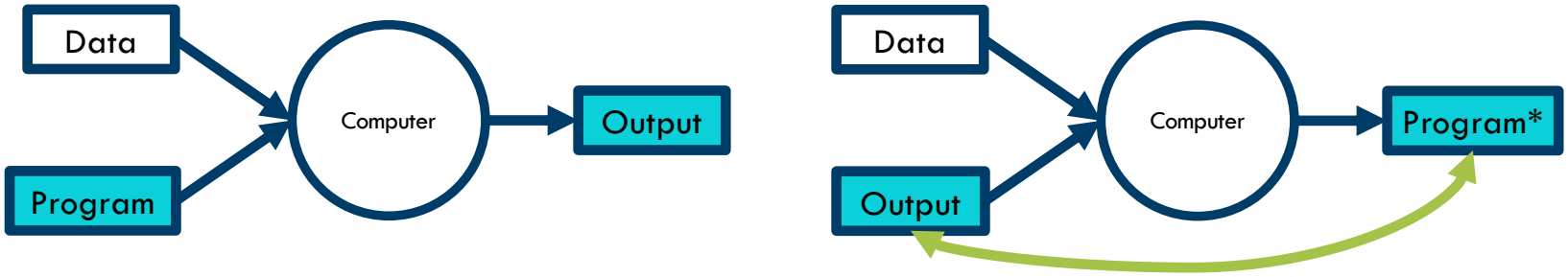
COMP 2537 and COMP 2800
Winter 2023

Python and Linear Regression

What is Machine Learning (ML)

- ML is a collection of algorithm techniques used to design systems that learn from data
- We use these systems to generate predictions and deduce patterns
- Compare traditional programming (left) with machine learning (right)



* also known as the "model"

ML algorithms fall into 2 broad categories

- *Supervised* learning algorithms
- **Trained** with labeled data, i.e., data composed of examples of the desired answers
- Using the labeled data, we can make predictions about unlabeled data, i.e., what kind of "thing" it is

- *Unsupervised* learning algorithms
- **Applied** to unlabeled data
- The goal is to find relationships in the unlabeled data

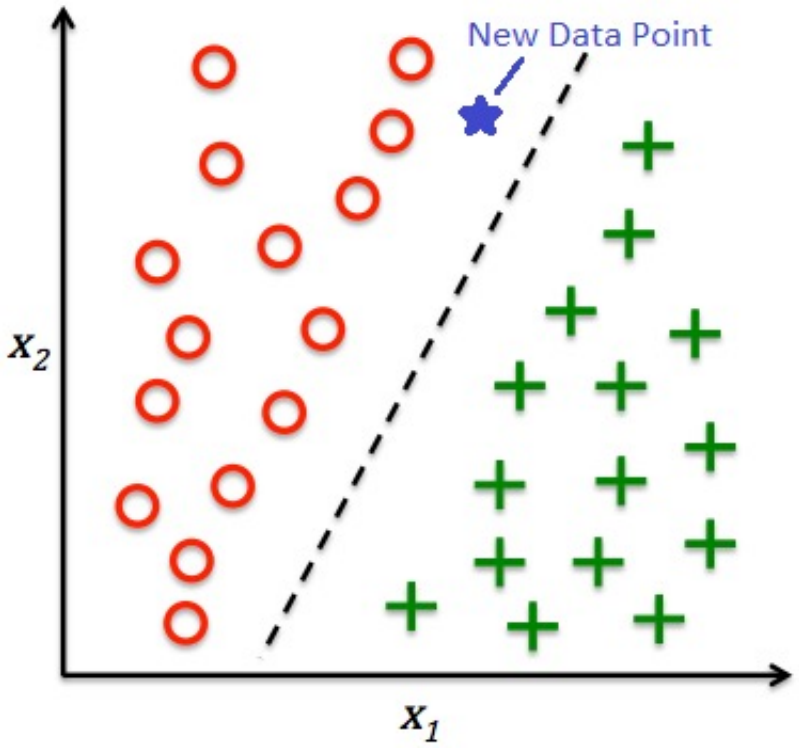
What sorts of problems can we solve?

We can solve **3** main kinds of problems with ML:

1. *Classification*: identifying the set of categories a new observation belongs to based on a set of training data. Is this thing an A or a B?
2. *Clustering*: grouping similar data points into intuitive groups. How is this organized?
3. *Regression*: estimating the relationship between variables to help forecast the future. How much or how many will there be?

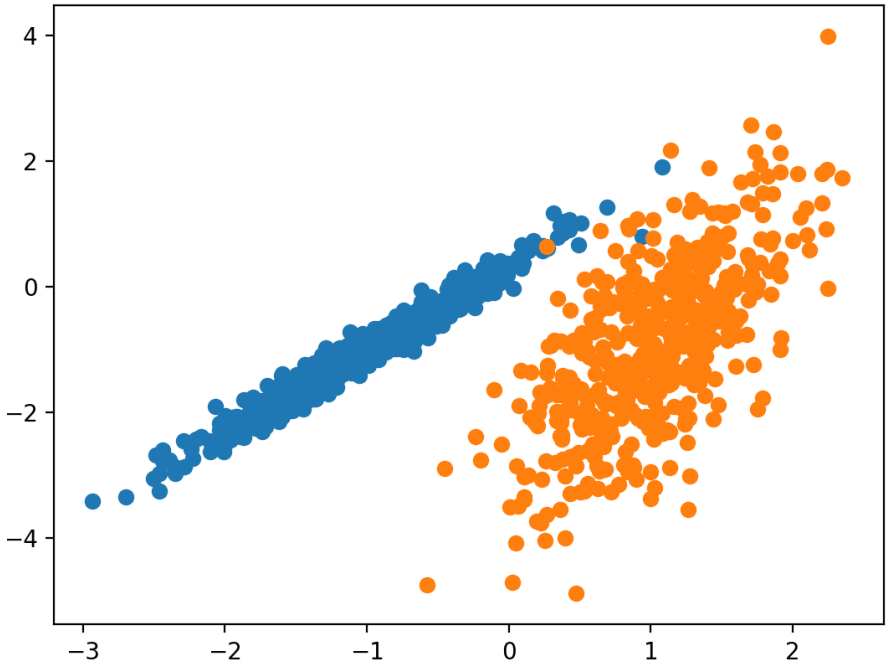
Classification

- Q: Is this tumor cancerous?
- Q: Who will win the upcoming election?
- Q: What kind of tree is this?
- The outcome of a classification problem is a discrete value that predicts the class to which an observation belongs
- The outcome can also be a continuous value, i.e., if candidate A is predicted to win with a probability of 0.65, the continuous value is 0.65 and the discrete class value is "win"
- Examples include logical regression and k-nearest neighbours



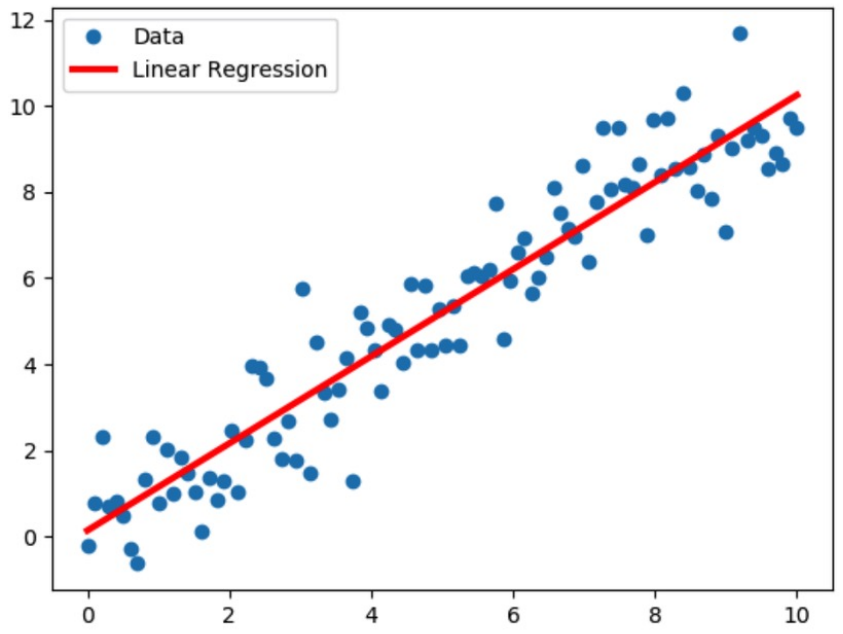
Clustering

- Q: Which viewers like the same kinds of movies?
- Q: Which models of hard drive fail in the same way?
- Given a set of data, clustering helps us decide how they are organized by segregating them into natural "clumps" with similar traits
- Clustering is an example of unsupervised learning
- An example is clustering using k-means



Regression

- Q: What are the predicted sales numbers for an item during the next quarter?
- Q: What will the temperature be next Friday?
- Q: What is the lifespan of a particular model of tire?
- The outcome is always a continuous output variable
- An example is linear regression
- Classification and regression use **labeled data**



Let’s get to work and do some programming

Flex your fingers

We’re about to do some stuff