CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

Practical Machine Learning

Practical Machine Learning

Lecture: Introduction to Machine Learning

Ted Scully

Basic Terminology Machine Learning Problem

- **Features (often referred to as attributes or variables)** below are Outlook, Temp, Humidity and Windy
- The **Class (Label)** is Play (for regression often referred to as regression target)
- **We refer to an instance as one row from the dataset.**
- **Inference –** Model takes in unseen feature vector and produces a classification.

Tennis Dataset

ID	Outlook	Temp	Humidity	Windy	Play?
A	sunny	hot	high	false	no
B	sunny	hot	high	true	no
C	overcast	hot	high	false	yes
D	rainy	mild	high	false	yes
E	rainy	cool	normal	false	yes
F	rainy	cool	normal	true	no
G	overcast	cool	normal	true	yes
H	sunny	mild	high	false	no
I	sunny	cool	normal	false	yes
J	rainy	mild	normal	false	yes
K	sunny	mild	normal	true	yes
L	overcast	mild	high	true	yes
M	overcast	hot	normal	false	yes
N	rainy	mild	high	true	no

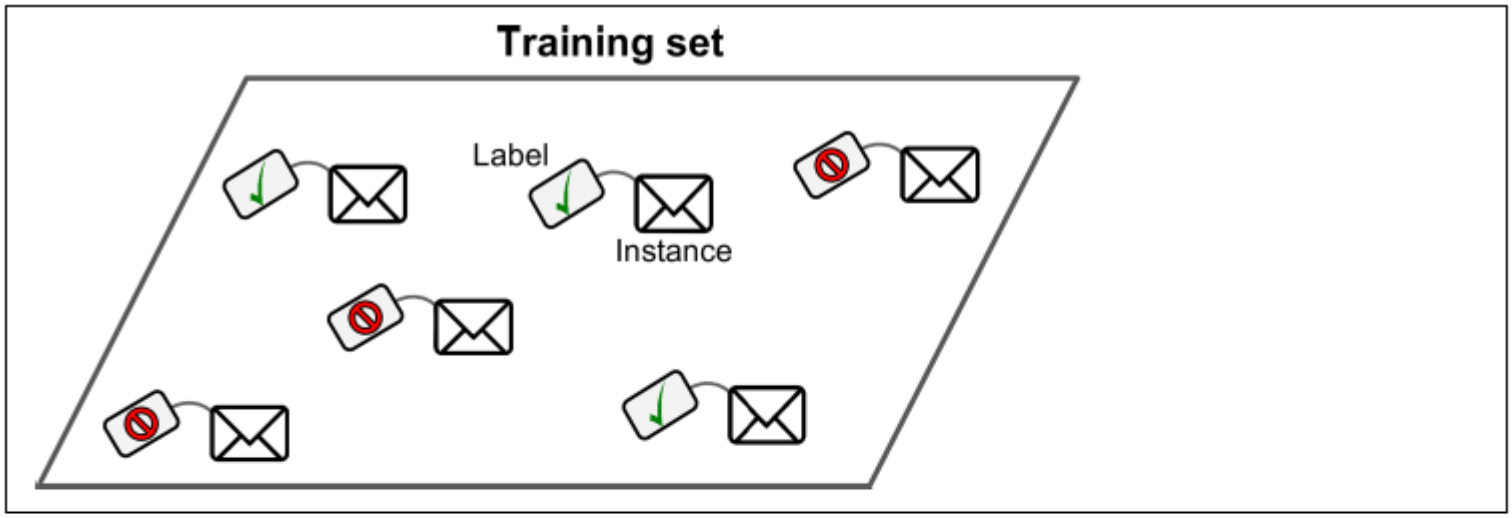
Categories of Machine Learning Algorithms

- Machine learning algorithms can be divided into five main categories
  - Supervised Learning Algorithms
  - Unsupervised Learning Algorithms
  - Semi Supervised Learning Algorithms
  - Self Supervised Machine Learning Algorithms
  - Reinforcement Learning Algorithms

Supervised Learning Algorithms

- Supervised learning algorithms used **labelled training data** to learn.

- In other words, the training data you feed to the algorithm includes the desired solutions, called labels.

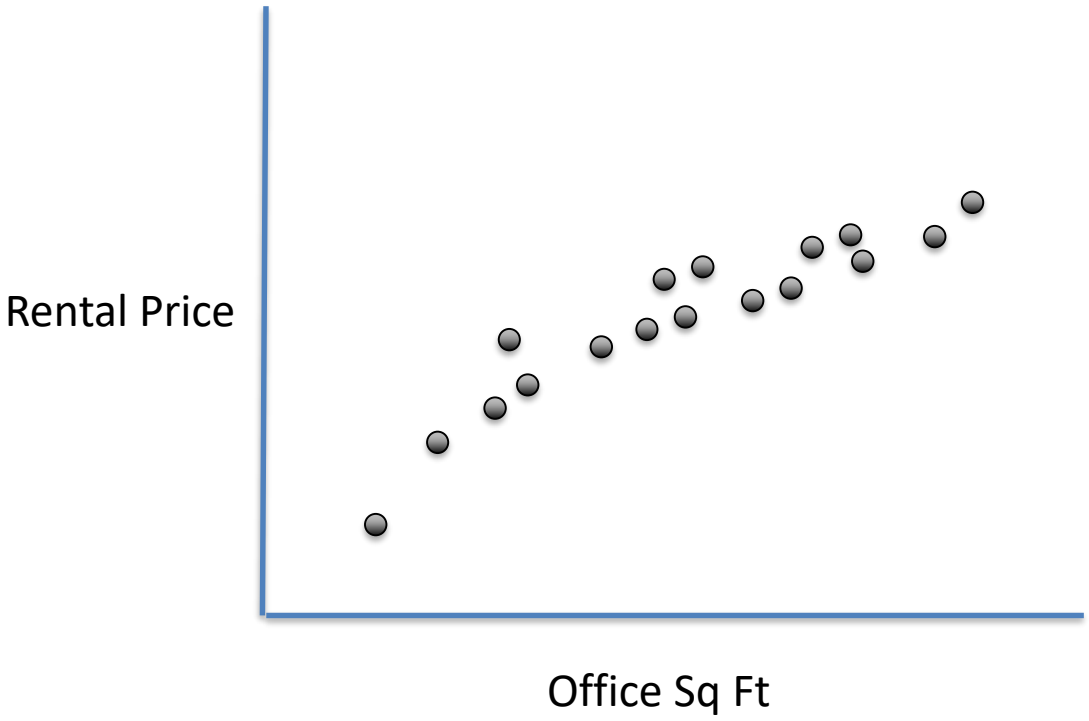


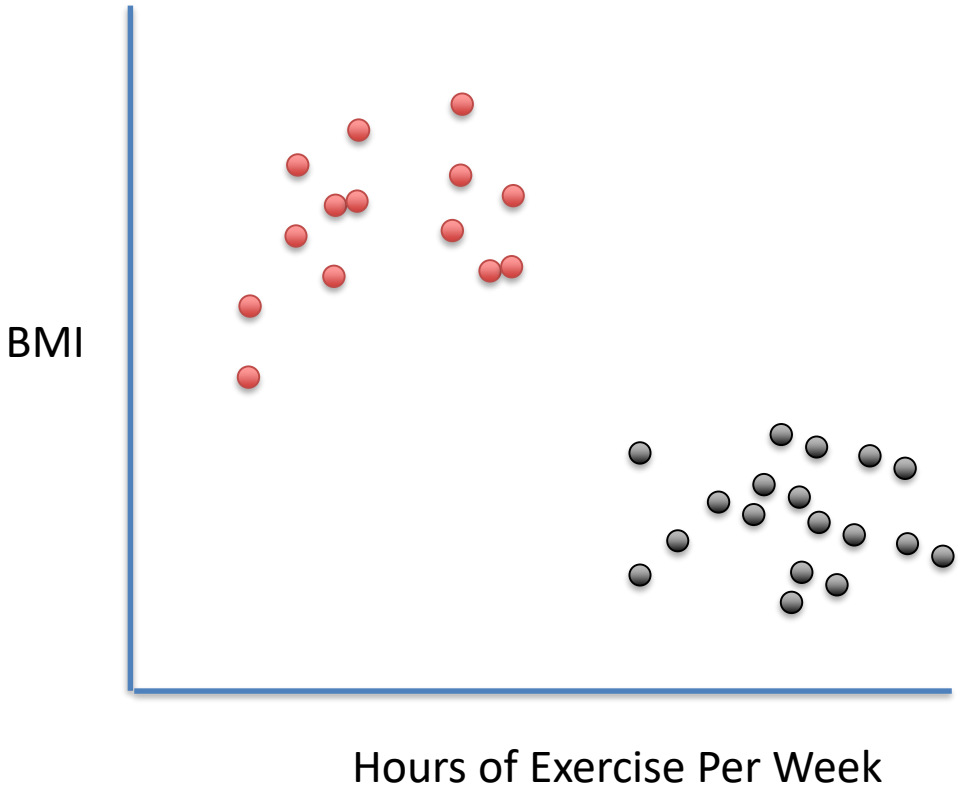
Supervised Learning Algorithms

- Supervised learning can be subdivided into either classification or regression algorithms.

- In **classification** the objective is to correctly **predict the category** that new objects or cases belong to based on specific attributes they have. This decision is based on previous data that you have already observed.

- **Regression** is similar except that rather than predicting a category we want to **predict a numerical value**. For example, predict the concentration of a drug based on a chemical analysis or predict a persons lifespan based on information about their health and lifestyle.

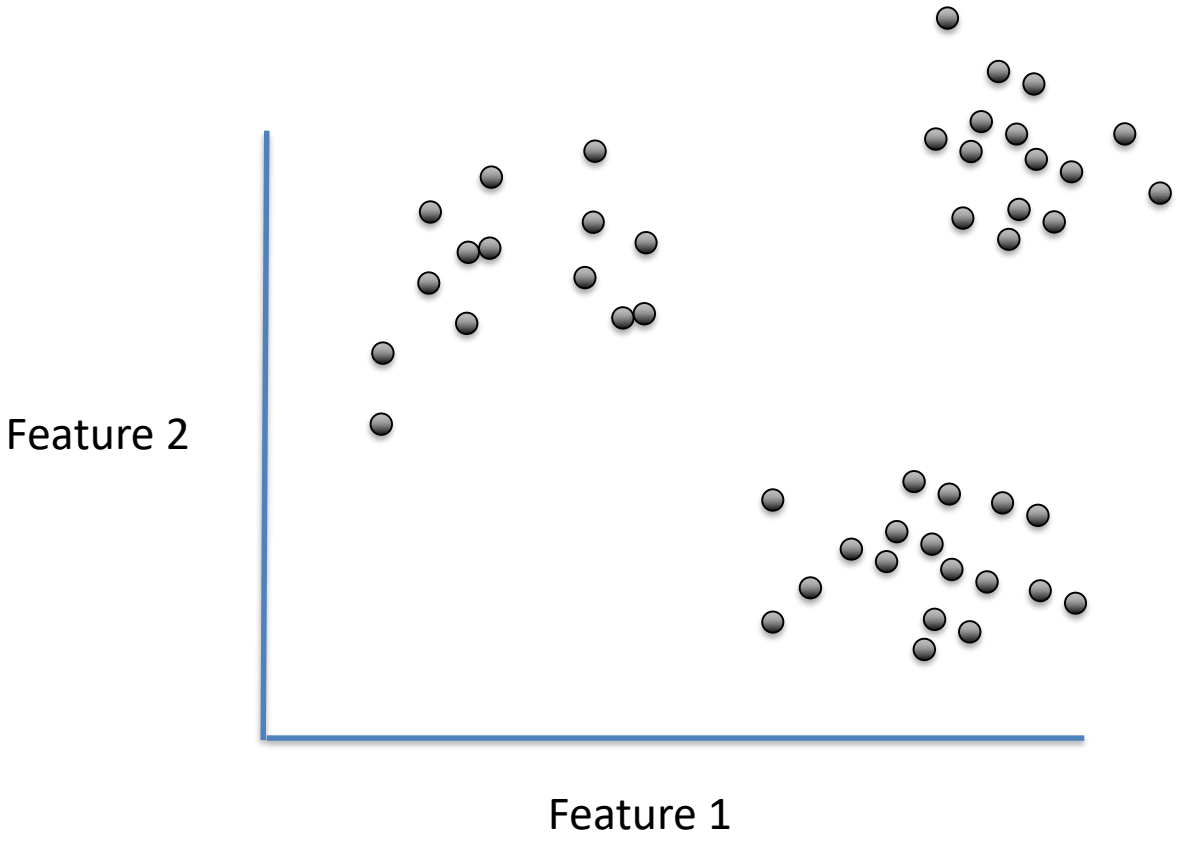

Rental Price
Office Sq Ft

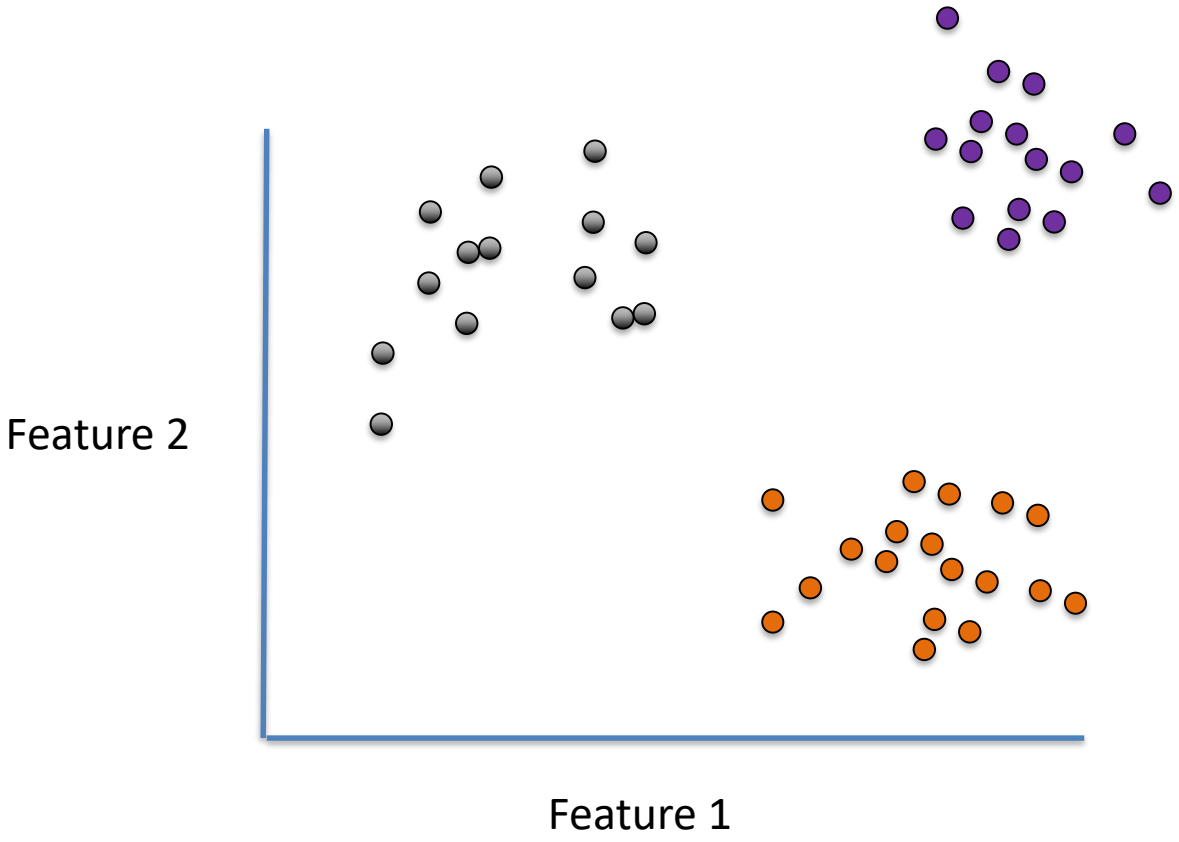

BMI
Hours of Exercise Per Week

Unsupervised Learning Algorithms

- In unsupervised learning the algorithm is not provided with any labelled training data and **must learn patterns from the data**.

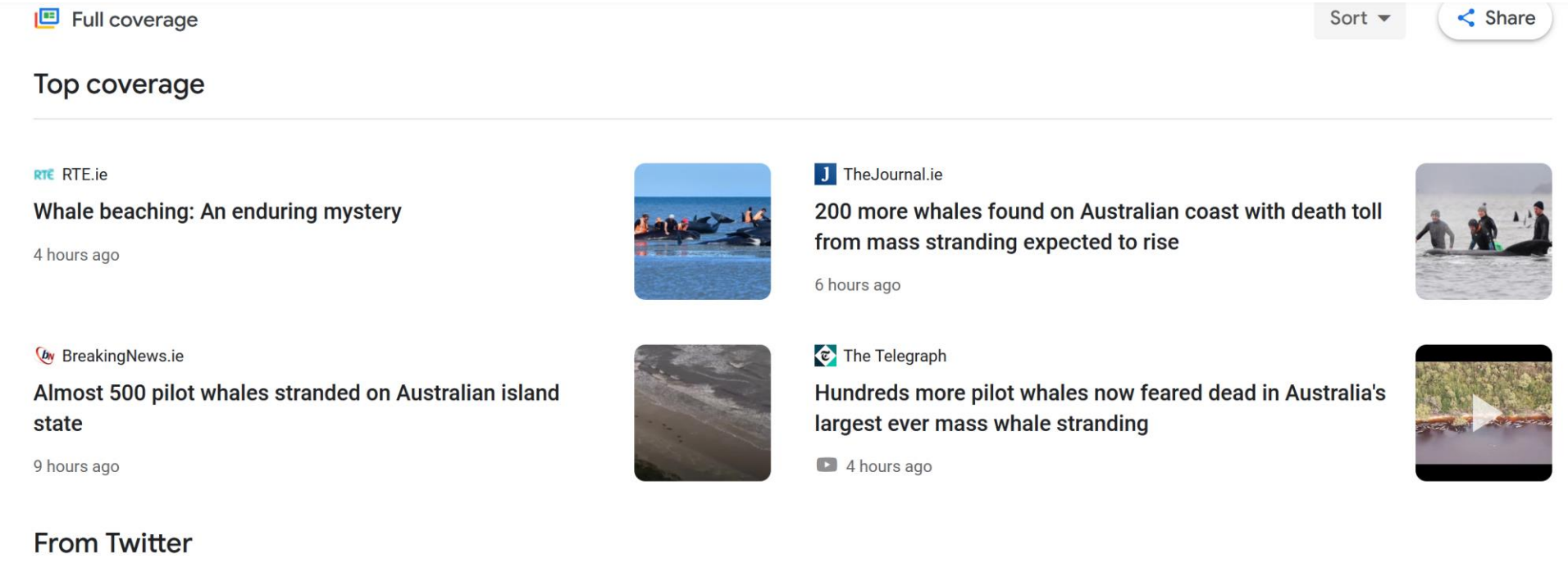
- Unsupervised algorithms seek out **similarity between pieces of data** in order to determine whether they can be characterized as forming a group.
  - These groups are termed clusters.
  - There are a broad range of clustering machine learning techniques.
  - Example- K Means Clustering (is told in advance how many clusters it should form -- a potentially difficulty)


Feature 2
Feature 1


Feature 2
Feature 1

Applications

- Google news uses clustering to group new articles related to their content. In this case articles related to a the recent whale beaching in Australia

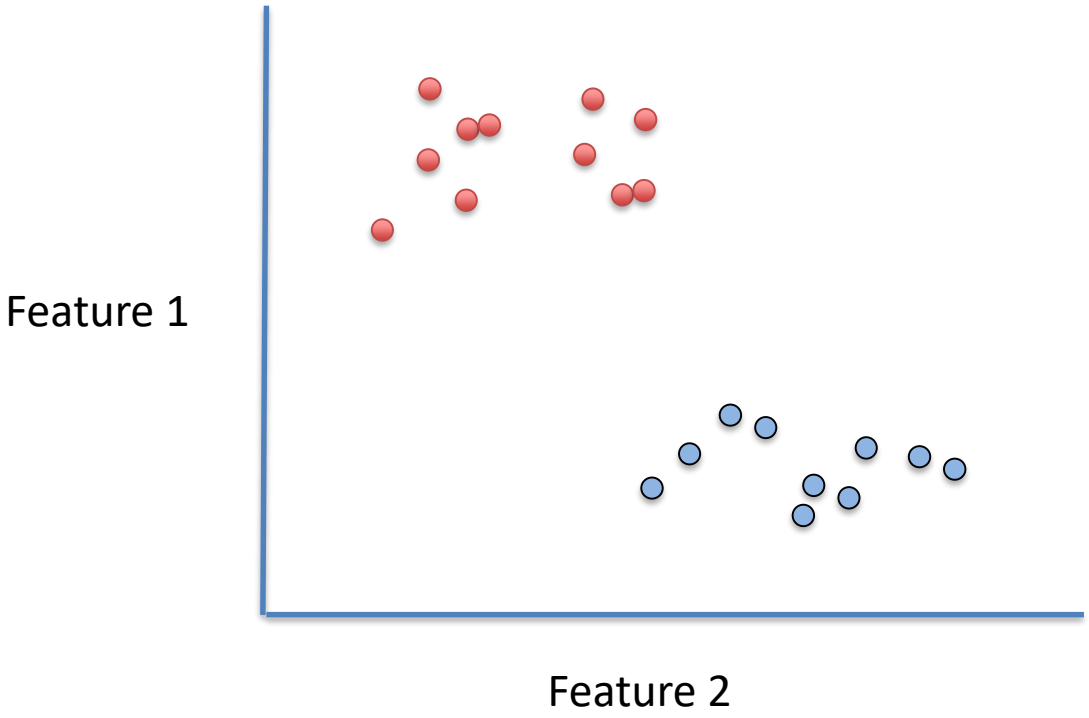


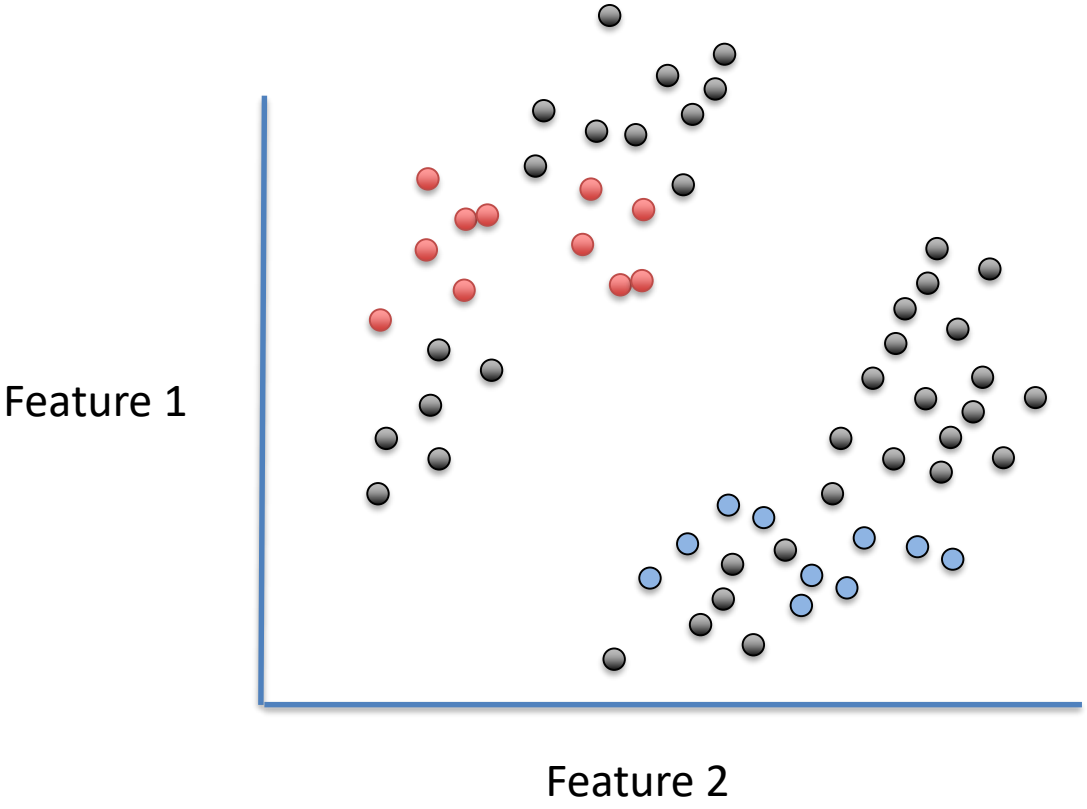
Semi-Supervised Learning

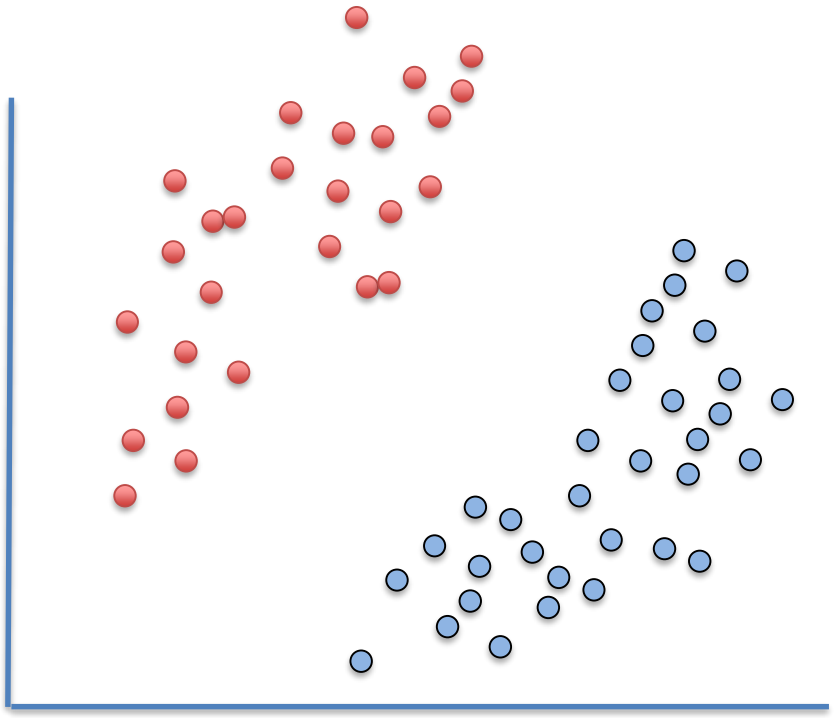
- The semi-supervised learning approach to machine learning **combines** supervised and unsupervised learning techniques.

- Remember supervised learning uses a labelled training set, while unsupervised learning techniques use unlabelled data.

- A semi-supervised approach **utilises both labelled and unlabelled data** for training.
  - Normally a small amount of labelled data is used along with a large amount of unlabelled data


Feature 1
Feature 2


Feature 1
Feature 2


Feature 1
Feature 2

Self Supervised Machine Learning Algorithms

- **Self supervised** learning occurs when we have an unlabelled dataset.
- However, the learning process we use is a **supervised learning process**.
- This seems counterintuitive! How can we use supervised learning techniques when the data isn't labelled???
- This is a little more challenging to grasp if you have limited exposure to ML.
- A good example of self supervised learning is a standard generative adversarial network (see next few slides).
- [Generative Adversarial Networks can be a little difficult to grasp so don't worry if you don't fully understand this now. We will be covering it in much more detail in the Deep Learning module next semester]

Real Dataset

Real Dataset

Artificial Dataset

Generator Network

The objective of the generator is to produce new images from the same distribution as the original dataset.

When the generator starts generating images initially they will be very poor. They won't look at all like the images from the real dataset.

We need to teach the generator how to produce between images.

To do this we will use a supervised learning algorithm.

But don't supervised learning algorithms take in labelled data?

Real Dataset

Class Label 1

Artificial Dataset

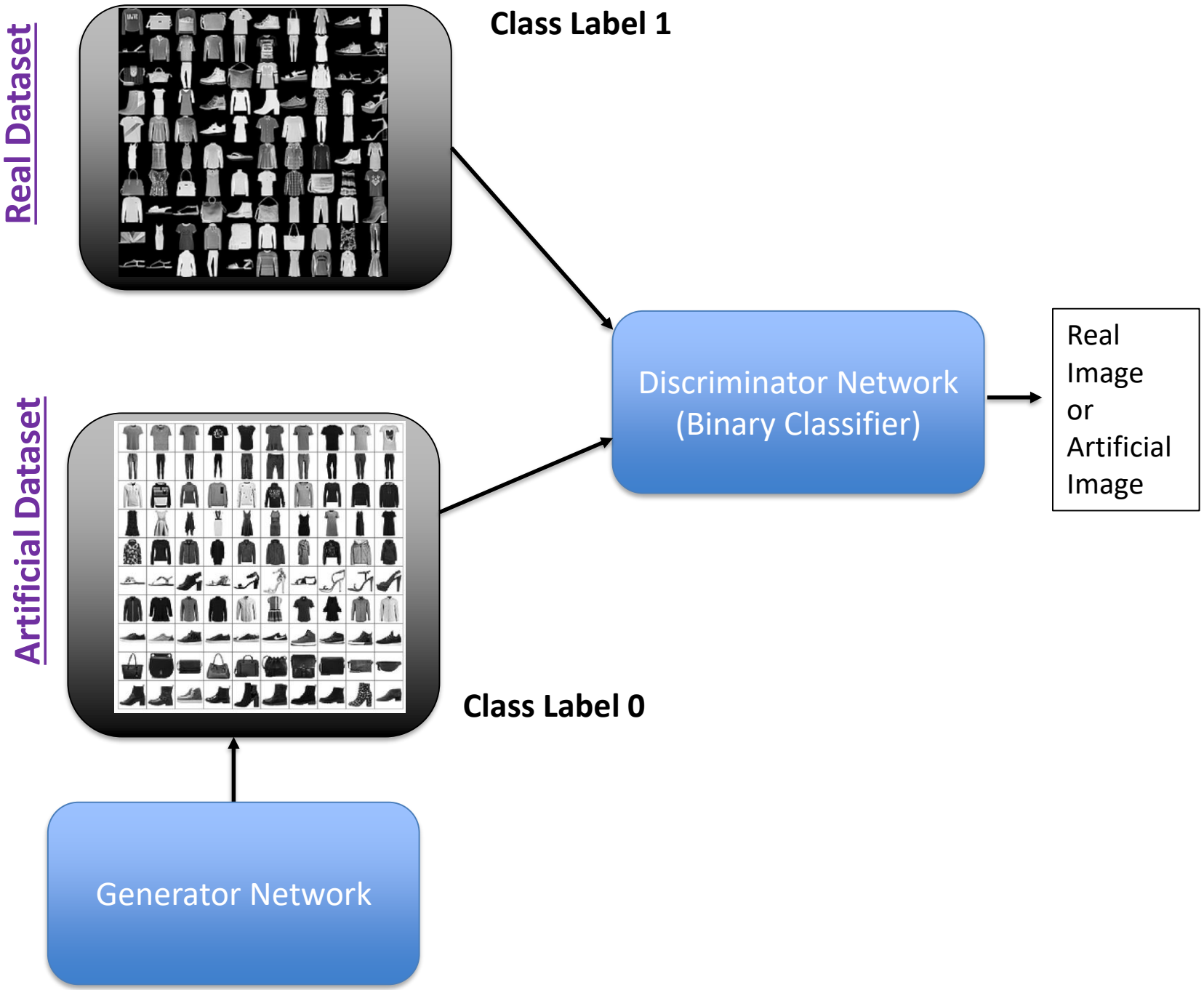
Class Label 0

Generator Network

To do this we will use a supervised learning algorithm. But don't supervised learning algorithms take in labelled data?

Correct.

So we label all the images in the true dataset as 1 and all the images in the artificial dataset as 0.


Real Dataset
Class Label 1
Artificial Dataset
Discriminator Network
(Binary Classifier)
Real
Image
or
Artificial
Image
Class Label 0
Generator Network

Reinforcement Learning Algorithms

- The objective of Reinforcement learning algorithms is to utilize observed rewards to **learn an optimal (or near-optimal) set of actions (or policy)** for each state in a given environment.

- Unlike most other forms of machine learning the learner **knows nothing about their environment** and just has a set of available actions that it can take.

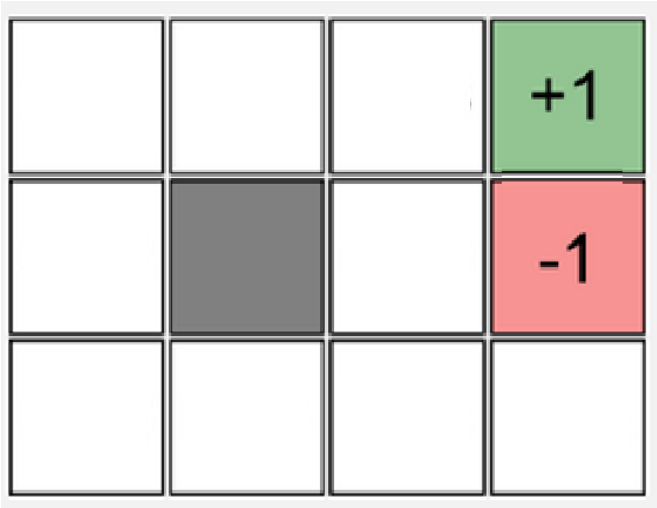
- The learner is not told which actions to take, but instead must discover which actions yield the most reward by trying them in the environment.

- This process continues until it reaches a positive or negative terminal or goal state. It then positively or negatively reinforces the actions that led to that state.

- Through repeated interaction with it's environment the agent develops a policy, which **defines what action the agent should choose when it is in a given situation.**

Reinforcement Learning Algorithms

- Take the simple environment on the left. An agent knows nothing about this environment but wants to navigate to the successful state.



- They must start by exploring their environment. They have a certain number of actions available to them. (Up, down, left or right). Each action will take them to a new date on the grid.

- The intended direction of movement occurs with a probability of 0.8. With a 0.2 probability you will make a move at right angle to the intended direction. The terminal states have rewards of +1 and -1 respectively.

- **A reinforcement learning program will play this game many times and will develop and optimal policy over time.**

Reinforcement Learning Algorithms

- Take the simple environment on the left. An agent knows nothing about this environment but wants to navigate to the successful state.
- They must start by exploring their environment. They have a certain number of actions available to them. (Up, down, left or right). Each action will take them to a new date on the grid.
- The intended direction of movement occurs with a probability of 0.8. With a 0.2 probability you will make a move at right angle to the intended direction. The terminal states have rewards of +1 and -1 respectively.
- **A reinforcement learning program will play this game many times and will develop and optimal policy over time.**