1 Introduction

1.1 What is machine learning?

According to **Arthur Samuel** is: ”Field of study that gives computers the ability to learn without being explicitly programmed.”

1.2 Classification of machine learning algorithms

- Supervised learning.
- Unsupervised learning.
- Recommender systems.
- Reinforcement learning.

1.3 Supervised Learning

It refers to algorithms that learn x to y or input to output mappings. The key characteristic of supervised learning is that you give your learning algorithm examples to learn from that includes the *right answers*. This it means the correct label y for a given input x.

1.3.1 Housing price prediction

We want to predict housing prices based on the size of the house.

1.3.2 Regression

In this type of supervised learning we are trying to predict a number from infinitely many possible numbers.

1.3.3 Classification

This kind of algorithm predicts categories. Categories don’t have to be numbers.

1.4 Unsupervised Learning

In these kind of algorithms were given data that isn’t associated with any output lables y. Our goal is finding something interesting in unlabeled data. We are not trying to supervised the algorithm to give some quote right answer for every input, instead we asked the algorithm to figure out all by itself.

1.4.1 Formal definition

Data only comes with inputs x, but not output labels y. Algorithm has to find a structure.

1.4.2 Clustering algorithms

It takes data without label and tries to automatically group them into clusters.

1.4.3 Anomaly detection

Find unusual effects.

1.4.4 Dimensionaly reduction

This algorithm let you take a big data set and compress it to a much smaller data set lossing little information as possible.

2 Linear Regression

This model just means fitting a straigh line to your data. It is called regression model because it predicts numbers as the output. There are infinitely many possible numbers that the model could output.

2.1 Terminology

2.1.1 Training set

Data used to train the model.

2.2 Notation

- $x =$ input variable.
- $y =$ output or target variable.
- $m =$ number of training examples.
- (x, y) single training example.
- $(x^{(i)}, y^{(i)})$ i^{th} training example with $1 \leq i \leq m$.

For representing the learning function we use

$$f_{w,b}(x) = wx + b$$

2.3 Linear Regression with one variable or univariate linear regression

Is a single input variable or feature x.

2.4 Cost function

It will tell us how well the model is doing. We have a training set that contains input features x and output targets y. The model we are going to use to fit this training set is this linear function:

$$f_{w,b}(x) = wx + b$$

w and b are called parameters of the model.
b is also called the y intercep because crossed the vertical axis. The value of w give us the slope of the line. In linear regression we want to do is to choose values for w and b so that straight line you get from the function f somehow fits the data well.

$$\hat{y} = f_{w,b}\left(x^{(i)}\right)$$

The **cost function** takes the prediction $\hat{y}$ and compares it to the target y by taking $\hat{y} - y$ this called the error. The final **cost function** is:

$$J(w, b) = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

this **cost function** is called squared error.

2.4.1 Recap

- model: $f_{w,b} = wx + b$
- parameters: w, b
- cost function: $J(w, b) = \frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}\left(x^{(i)}\right) - y^{(i)}\right)^2$
- goal: $\min\limits_{w,b} J(w, b)$