Week 1

Goal: Predicting outputs with a model trained for it. Output can be a value, function, class.

Supervised and unsupervised learning are the two main subclass of ML.

In Supervised learning our training set contains both x (inputs) and y (outputs), but unsupervised learning seeks patterns and structure in x as training data doesn't contain y.

Regression: A technique to predict a output from infinitely many possibilities.

Classification: Predicting a outputs from finite number of categories.

Linear Regression with One variable

$F: F_{w,b}(x) = wx + b$

$\hat{y} = F(x)$

$$\left.\begin{array}{l} w\text{: weight} \\ b\text{: bias} \end{array}\right\} \text{Adjustable parameters}$$

Cost Function

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \Big(\underbrace{F(x^{(i)})}_{\hat{y}^{(i)}} - y^{(i)}\Big)^2$$

m: batch size

Aim is to minimize cost function J by adjusting w,b.

In order to do that we use a algorithm called gradient descent.

Gradient Descent

- Take derivatives of J respect to w, b.
- Adjust w and b respect to their derivatives.

Adjusting part must be done simultaneously.

$$w = w - \alpha \frac{\partial J(w,b)}{\partial w}$$

$$b = b - \alpha \frac{\partial J(w,b)}{\partial b}$$

$$\frac{\partial J}{\partial w} = \frac{1}{m} \sum_{i=1}^{m} \left(F(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

$$\frac{\partial J}{\partial b} = \frac{1}{m} \sum_{i=1}^{m} \left(F(x^{(i)}) - y^{(i)}\right)$$