Linear Regression

gradient descent : $\theta_j = \theta_j - \alpha \cdot \frac{d}{d\theta_j} J(\theta_j)$

No need to decrease α over time since gradient descent will automatically take smaller steps as approaching a local minimum.

Cost Function : $J(\theta) = \frac{1}{2m} \cdot \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$

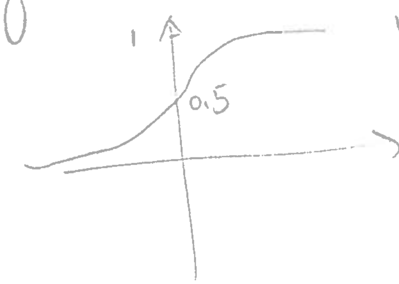
Normal Equation : Method to solve θ for analytically.

$$\theta = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

Feature Scaling : make sure features are on a similar scale.

Logistic Regression.

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T \cdot x}}$$



when $\theta^T x >= 0$,
$h_\theta(x) \geq 0.5$.

Cost function : $J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \cdot \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))\right]$

This cost function is convex.

Gradient descent :

$$\theta_j = \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$