1 Week 1

1.1 Model Representation

$$h_\theta(x) = \theta_0 + \theta_1 x + \epsilon$$

1.2 Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

1.3 Gradient Descent

Want: $minJ(\theta_0, \theta_1)$
Algorithm: repeat until convergence {
$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1), j = 0, 1$
simutaneously update θ_j
}

2 Week 2

2.1 Multivariate Linear Regression Model

Hypothesis:

$$h_\theta(x) = X\Theta$$

where

$$X = \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{pmatrix}_{m \times (n+1)}, \Theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}_{(n+1) \times 1}$$

2.2 Cost function(Vectorization)

$$J(\theta) = \frac{1}{2m} (X\Theta - \vec{y})^T (X\Theta - \vec{y})$$

2.3 Normal Equation

$$\Theta = (X^T X)^{-1} X^T \vec{y}$$

Deduction:

$$\begin{aligned}
J(\theta) &= \frac{1}{2m}(X\Theta - \vec{y})^T(X\Theta - \vec{y}) \\
&= \frac{1}{2m}\|X\Theta - \vec{y}\|^2 \\
\nabla J(\Theta) &= \frac{1}{m}X^T(X\Theta - \vec{y}) \\
\nabla J(\Theta) &= 0 \\
X^T X\Theta - X^T\vec{y} &= 0 \\
\Theta &= (X^T X)^{-1}X^T\vec{y}
\end{aligned}$$

3 Week 3

3.1 Classification

Classification: $y \in \{0, 1\}$, $h_\theta(x)$ can be > 1 or < 0.
Logistic Regression: $0 \leq h_\theta(x) \leq 1$

3.2 Logistic Regression

Logistic Regression Model:

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

Interpretation:

- $h_\theta(x)$ =estimated probability that $y = 1$ on input x.
- $h_\theta(x) = P(y = 1|x; \theta)$ means probability that $y = 1$, given x, parameterized by θ.
- $P(y = 0|x; \theta) + P(y = 1|x; \theta) = 1$