

Linear Regression Model

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Mean squared error cost function.

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

Gradient descent algorithm

$$\theta_j := \theta_j - \alpha \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_j}$$

↑
learning rate

Simultaneous update

$$\text{temp0} = \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\text{temp1} = \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_0 := \text{temp0}$$

$$\theta_1 := \text{temp1}$$

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{i1} & x_{i2} \end{pmatrix} \cdot \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{pmatrix} = \begin{pmatrix} x_{11}\theta_0 + x_{12}\theta_1 - y_1 \\ \vdots \\ x_{i1}\theta_0 + x_{i2}\theta_1 - y_i \end{pmatrix}$$

error

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_0} = \frac{1}{N} \sum_{i=1}^N (x_{i1}\theta_0 + x_{i2}\theta_1 - y_i) x_{i1}$$

multiply

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_1} = \frac{1}{N} \sum_{i=1}^N (x_{i1}\theta_0 + x_{i2}\theta_1 - y_i) x_{i2}$$

Multivariate Linear Regression

$$h(x) = \theta^T x = \theta_0 x_0 + \dots + \theta_n x_n$$

$\nearrow x_0 = 1$

cost function $J = \frac{1}{2m} \sum_{i=1}^m (h(x^i) - y^i)^2$

Gradient descent:

$$\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}$$

$$= \theta_j - \alpha \cdot \frac{1}{m} \sum_{i=1}^m (h(x^i) - y^i) x_{j^i}$$

Mean normalization

$$x_i = \frac{x_i - \mu_i}{s_i}$$

\leftarrow mean

range of x_i (max-min)

feature scaling $x_i = \frac{x_i - \mu_i}{s_i}$

Normal Equation

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