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 desent algorithm

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

learning rate

Simultaneous update

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

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

$$\theta_0 := temp0$$

$$\theta_1 := temp1$$

error

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

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

$$h(x) = \theta^T x = \theta_0 x_0 + \cdots \theta_n x_n \qquad x_0 = 1$$

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

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

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

range of x_i (max - min)

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

Normal Equation

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

Logistic Regression Model

$$h(x) = \frac{1}{1+e^{-\theta^T x}} = p(y=1 \mid x;\theta)$$

given x, θ, the probability of $y=1$

$$\begin{cases} h(x) \geq 0.5 & y=1 \quad (\theta^T x \geq 0) \\ h(x) < 0.5 & y=0 \quad (\theta^T x < 0) \end{cases}$$

Cost function

$$\mathrm{Cost}(h(x), y) = \begin{cases} -\log(h(x)) & y=1 \\ -\log(1-h(x)) & y=0 \end{cases}$$

$$- y\log(h(x)) - (1-y)\log(1-h(x))$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h(x), y)$$

Gradient Descent

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{m} (h(x^i) - y^i)\, x_j^i$$

Regularization

$$J(\theta) = \frac{1}{2m} \left[\sum_{i=1}^{m} (h(x^i) - y^i)^2 + \lambda \sum_{j=1}^{n} \theta_j^2 \right]$$

regularization parameter

Normal Equation

$$\theta = \left(X^T X + \lambda \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & \ddots \; 1 \end{bmatrix} \right)^{-1} X^T y$$

n+1 x n+1