GRADIENT DESCENT FOR LOGISTIC REGRESSION

TRAINING DATA: $(x_i, y_i),\ i=1,2,\dots,N;\ y_i \in \{0,1\}$

$$h_\theta(x) = g(\theta^T x)$$

$$g(z) = \frac{1}{1+e^{-z}} \qquad g'(z) = +\frac{1}{(1+e^{-z})^2}\cdot e^{-z} = \frac{1}{1+e^{-z}}\cdot\frac{e^{-z}}{1+e^{-z}} = g(z)(1-g(z))$$

$$J(\theta) = -\sum_{i=1}^{N} \underbrace{y_i \log h_\theta(x_i) + (1-y_i)\log(1-h_\theta(x_i))}_{C_i}$$

$$J(\theta) = -\sum_{i=1}^{N} C_i(\theta)$$

DERIVATIVES

$$\left(\frac{1}{x}\right)' = -\frac{1}{x^2}$$

$$(\log x)' = \frac{1}{x}$$

$$(e^x)' = e^x$$

$$(e^{-x})' = -e^{-x}$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\sum_{i=1}^{N} \frac{\partial C_i(\theta)}{\partial \theta_j}$$

$$\frac{\partial h_\theta(x_i)}{\partial \theta_j} = \frac{\partial g(\theta^T x_i)}{\partial \theta_j}$$

$$= \frac{\partial g\left(\sum_{j=0}^{d} \theta_j x_{ij}\right)}{\partial \theta_j} = g(\theta^T x_i)(1-g(\theta^T x_i))\cdot \frac{\partial\left(\sum_{j=0}^{d}\theta_j x_{ij}\right)}{\partial \theta_j}$$

$$= g(\theta^T x_i)(1-g(\theta^T x_i))\cdot x_{ij}$$

$$C_i = \underbrace{y_i \log h_\theta(x_i)}+\underbrace{(1-y_i)\log(1-h_\theta(x_i))}$$

$$\frac{\partial C_i}{\partial \theta_j} = y_i\cdot\frac{1}{\boxed{h_\theta(x_i)}}\boxed{g(\theta^T x_i)}(1-g(\theta^T x_i))\,x_{ij}$$

$$-(1-y_i)\frac{1}{\boxed{1-h_\theta(x_i)}}\cdot g(\theta^T x_i)\boxed{(1-g(\theta^T x_i))}\,x_{ij}$$

$$\frac{\partial c_i}{\partial \theta_j} = y_i\left(1 - g(\theta^T x_i)\right) x_{ij} - (1-y_i)\, g(\theta^T x_i)\, x_{ij}$$

$$= \left[y_i - \cancel{y_i g(\theta^T x_i)} - g(\theta^T x_i) + \cancel{y_i g(\theta^T x_i)}\right] x_{ij}$$

$$= \left[y_i - g(\theta^T x_i)\right] x_{ij}$$

$$= \boxed{\left(y_i - h_\theta(x_i)\right) x_{ij}}$$

$$J(\theta) = -\sum_{i=1}^{N} c_i$$

$$\boxed{\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{N} \left(h_\theta(x_i) - y_i\right) x_{ij}}$$

LINEAR REGRESSION : $h_\theta(x_i) = \theta^T x_i$

LOGISTIC $h_\theta(x_i) = g(\theta^T x_i) = \frac{1}{1+e^{-\theta^T x_i}}$