Linear Regression

$x = [x_0, x_1, x_2, \ldots, x_n]^T$

$\theta = [\theta_0, \theta_1, \theta_2, \ldots, \theta_n]^T$

hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n = \theta^T x$

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

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

learning rate

vecterize: $\theta = \theta - \frac{\alpha}{m} X^T (X\theta - y)$

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

Logistic Regression

hypothesis: $h_\theta(x) = \dfrac{1}{1+e^{-\theta^T x}}$ (sigmod)

cost function: $J(\theta) = -\frac{1}{m}\sum \left(y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right)$

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

regularization cost function:

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

gradient descent for regularization

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

normal equation:

$$\theta = \Big(X^T X + \lambda \underbrace{\begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}}_{(n+1)\times(n+1)} \Big)^{-1} X^T y$$