Note:

$h_o(x) = g(\theta^T x) = g(x_1\theta_1 + ... + x_j\theta_j + ...x_n\theta_n)$, $(z = x_1\theta_1 + ... + x_j\theta_j + ...x_n\theta_n)$

$\frac{\partial}{\partial\theta_j} h_o(x) = g'(z)\frac{\partial z}{\partial\theta_j}$, $x_j = \frac{\partial z}{\partial\theta_j}$

$g(z) = \frac{1}{1+e^-z}$

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

$(\therefore 1 - g(z) = 1 - \frac{1}{1+e^{-z}} = \frac{e^{-z}}{1+e^{-z}})$

Stochastic Gradient Descent:

$\theta_j := \theta_j + \alpha\frac{\partial}{\partial\theta_j} l(0) = \theta_j + \alpha(y^{(i)} - h(x^{(i)})x_j^{(i)})$

There is another method which runs faster than this stochastic gradient descent method. It it called Newton's method.

Newton's Method:

Say $f : \mathbb{R} \to \mathbb{R}$

$f(x_{n+1}) = f(x_{n+1}) = f(x_n) + f'(x_n)t + \frac{1}{2}f''(x_n)t^2$

$\frac{df}{dt}(x_{n+t}) = f'(x_n) + \frac{1}{2}f''(x_n)2t = 0$

$\frac{df}{dt}(x_{n+t}) = f'(x_n) + f''(x_n)t = 0$

$t = \frac{-f'(x_k)}{f'(x_k)}$ $x_{k+1} = x_k - \frac{f'(x_k)}{f'(x_k)}$

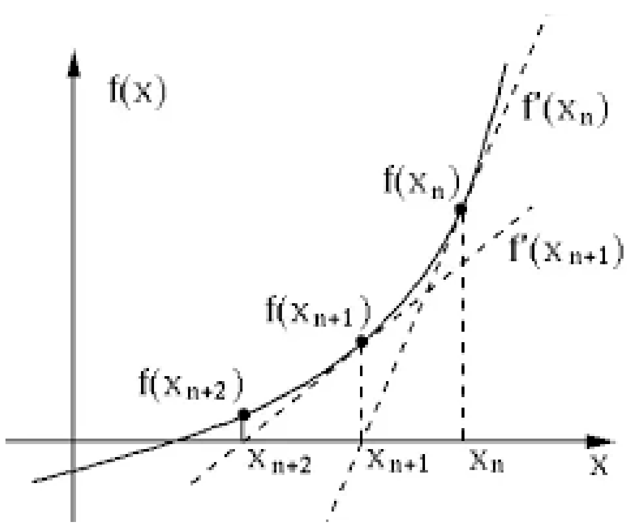

Figure 1: Newton's Method