Derivation of Stochastic Gradient Descent

Jody Shu

March 10, 2020

We want to prove:

$\nabla l(\theta) = \sum_{i=0}^{n}[y^{(i)}\frac{1}{h_\theta(x^{(i)})}(\frac{\partial}{\partial\theta_j}h_\theta(x^{(i)})) + (1-y^{(i)})\frac{-1}{(1-h_\theta(x^{(i)}))}(\frac{\partial}{\partial\theta_j}h_\theta(x^{(i)}))]$
$=\sum_{i=0}^{n}[(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}]$

we know $l(\theta) = \sum_{i=0}^{n}[y^{(i)}logh_\theta(x^{(i)}) + (1-y^{(i)})log(1-h_\theta(x^{(i)})]$

So, $\nabla l(\theta) = \sum_{i=0}^{n}[y^{(i)}\frac{1}{h_\theta(x^{(i)})}(\frac{\partial}{\partial\theta_j}h_\theta(x^{(i)}))+(1-y^{(i)})\frac{-1}{(1-h_\theta(x^{(i)}))}(\frac{\partial}{\partial\theta_j}h_\theta(x^{(i)}))]$

since $h_\theta(x) = g(\theta^T x) = g(z) = \frac{1}{1+e^{-Z}} \ldots (1)$,

$\Longrightarrow g'(z) = \frac{-e^{-z}}{(1+e^{-Z})^2} = \frac{1}{1+e^{-Z}}\frac{-e^{-z}}{1+e^{-Z}} = g(z)(1-g(z)) \ldots (2)$

Let $\frac{\partial z}{\partial\theta_j} = x_j$,

Then from equation (1) and (2), $\frac{\partial}{\partial\theta_j}h_\theta(x) = g'(z)\frac{\partial z}{\partial\theta_j} = g(z)(1-g(z))x_j$

$\Longrightarrow \nabla l(\theta) = \sum_{i=0}^{n}[y^{(i)}\frac{1}{g(z)}(g(z)(1-g(z))x_j^{(i)})+(1-y^{(i)})\frac{-1}{(1-g(z))}(g(z)(1-g(z))x_j^{(i)})]$

$\Longleftrightarrow \sum_{i=0}^{n}[y^{(i)}(1-g(z))x_j^{(i)}) - (1-y^{(i)})(g(z)x_j^{(i)})]$

$\iff \sum_{i=0}^{n}[y^{(i)}x_j^{(i)} - y^{(i)}g(z)x_j^{(i)} - g(z)x_j^{(i)} + y^{(i)}(g(z)x_j^{(i)})]$

$\iff \sum_{i=0}^{n}[y^{(i)}x_j^{(i)} - g(z)x_j^{(i)}]$ (After cancelling terms)

$\iff \sum_{i=0}^{n}[(y^{(i)} - g(z))x_j^{(i)}]$

$\iff \sum_{i=0}^{n}[(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}]$

Now, we have proved that

$\nabla l(\theta) = \sum_{i=0}^{n}[y^{(i)}\frac{1}{h_\theta(x^{(i)})}(\frac{\partial}{\partial\theta_j}h_\theta(x^{(i)})) + (1-y^{(i)})\frac{-1}{(1-h_\theta(x^{(i)}))}(\frac{\partial}{\partial\theta_j}h_\theta(x^{(i)}))]$
$=\sum_{i=0}^{n}[(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}]$