Pooja Jadhav

H·W-03

2) we know that $\sigma(\alpha) = \frac{1}{1+e^{-\alpha}}$, assume $h_w(x) = g(w^T x) = \frac{1}{1+e^{-w^T x}}$

(a) calculating $\frac{d\sigma}{d\alpha} = \sigma'(\alpha) = \frac{d}{d\alpha}\left(\frac{1}{1+e^{-\alpha}}\right)$

$$\Rightarrow \frac{1}{(1+e^{-\alpha})^2}(e^{-\alpha}) \Rightarrow \frac{1}{(1+e^{-\alpha})} \cdot \left[1 - \frac{1}{(1+e^{-\alpha})}\right]$$

$$\frac{d\sigma}{d\alpha} = \sigma(\alpha)(1-\sigma(\alpha))$$

Hence proved

(b) given that $P(y=1|x) = \sigma(w^T x)$ can be written as $h_w(x)$

$P(y=-1|x) = 1-\sigma(w^T x)$ can be written as $(1-h_w(x))$

$P(y|x) = P(y=1|x)^{\frac{y+1}{2}} \cdot P(y=-1|x)^{\frac{1-y}{2}}$ can be written as

$$P(y|x) = [h_w(x)]^{\frac{y+1}{2}} [1-h_w(x)]^{\frac{1-y}{2}}$$

on maximizing the log-likelihood we get,

$$\ell(w) = \sum_{n=1}^{N} \log P\left(x|_{n}\right)$$

$$\ell(w) = \sum_{n=1}^{N} \log\left(h_w(x)\right)^{\frac{y+1}{2}} + \log\left[1 - h_w(x)\right]^{\frac{1-y}{2}}$$

But we know that $\log a^{b} = b \log a$

$$\ell(w) = \sum_{n=1}^{N} \left(\frac{y+1}{2}\right) \log h_w(x) + \left(\frac{1-y}{2}\right) \log\left[1 - h_w(x)\right]$$

In order, to maximize we calculate the gradient ascent

Pooja Jadhav ②

$$w := w + \beta \nabla_w l(w)$$

now consider for one training sample, (x, y) we need to ~~derive~~ perform derivatives of $l(w)$ to derive the stochastic gradient ascent rule,

$$\frac{\partial}{\partial w_j} l(w) = \left[\left(\frac{y_n+1}{2}\right) \frac{1}{g(w^T x)} - \left(\frac{1-y_n}{2}\right) \frac{1}{(1-g(w^T x))} \right] \frac{\partial}{\partial w_j} g(w^T x)$$

$$\Rightarrow \left[\left(\frac{y_n+1}{2}\right) \frac{1}{g(w^T x)} - \left(\frac{1-y_n}{2}\right) \frac{1}{(1-g(w^T x))} \right] g(w^T x) \left[1 - g(w^T x) \frac{\partial}{\partial w_j}(w^T x)\right]$$

$$\Rightarrow \left[\left(\frac{y_n+1}{2}\right) \frac{1}{g(w^T x)} - \left(\frac{1-y_n}{2}\right) \frac{1}{(1-g(w^T x))} \right] g(w^T x) \left[1 - g(w^T x)\right] x_j \cdot 1$$

$$\Rightarrow \frac{\left[(y_n+1)(1-g(w^Tx)) - (1-y_n)(g(w^Tx))\right]}{2\cdot \cancel{g(w^Tx)}\,\cancel{(1-g(w^Tx))}}\ \cancel{g(w^Tx)}\,\cancel{[1-g(w^Tx)]}\,x_j$$

$$\Rightarrow \frac{x_j}{2}\left[y_n+1-\cancel{y_n g(w^Tx)}-g(w^Tx)+\cancel{y_n g(w^Tx)}-g(w^Tx)\right]$$

$$\Rightarrow \frac{x_j}{2}\left[y_n+1-2g(w^Tx)\right]$$

we know that $g(w^Tx) = h_w(x)$

$$\frac{\partial}{\partial w_j} l(w) = \frac{x_j}{2}\left[y_n+1-2h_w(x)\right]$$

stochastic gradient ascent rule is given by,

$$w_j := w_j + \beta\left(y_n + 1 - 2h_w(x)\right)\frac{x_j}{2}$$

where $h_w(x)$ is non-linear function of $w^T x$