Page :
Date : / /

Assignment 4

Q2-2) $P(y_i | w, x_i) = \dfrac{1}{1 + e^{-y_i w^T x_i}}$

Objective func : $\min_w \left(\dfrac{\lambda}{2} \|w\|^2 + \sum_{i=1}^{N} \ln\left(1 + e^{-y_i w^T x_i}\right)\right)$

$E(w) = \dfrac{\delta LL}{\delta w}$

$= \dfrac{\delta}{\delta w}\left(\dfrac{\lambda}{2}\|w\|^2 + \sum_{i=1}^{N} \ln\left(1 + e^{-y_i w^T x_i}\right)\right)$

Let $\alpha_i = y_i w^T x_i$
$O_i = \sigma(\alpha_i)$

$\therefore E(w) = \lambda w + \dfrac{\delta}{\delta w}\left(\sum_{i=1}^{N} \ln (O_i)^{-1}\right)$

Applying chain rule :

$\dfrac{\delta \alpha_i}{\delta w} = \dfrac{\delta}{\delta w}(y_i w^T x_i) = y_i x_i$

$\dfrac{\delta O_i}{\delta \alpha_i} = \dfrac{\delta(\sigma(\alpha_i))}{\delta \alpha_i} = \dfrac{\delta}{\delta \alpha_i}\left(\dfrac{1}{1 + e^{-\alpha_i}}\right)$

$= \dfrac{-e^{-\alpha_i}}{1 + e^{-\alpha_i}} = (1 - O_i) O_i$

[from derl/b slides]

$\dfrac{\delta LL}{\delta O_i} = \dfrac{\delta}{\delta O_i}\left(\ln(O_i)^{-1}\right) = -(O_i)^{-1}$

Using chain rule:

$$\frac{\delta LL}{\delta w} = \frac{\delta LL}{\delta O_i} \times \frac{\delta O_i}{\delta \alpha_i} \times \frac{\delta \alpha_i}{\delta w}$$

Substituting above for $E(w)$

$$\therefore E(w) = \lambda w + \sum_{i=1}^{N} \left[y_i x_i \times (1 - O_i) O_i \times \frac{-1}{O_i} \right]$$

$$= \lambda w + \sum_{i=1}^{N} y_i x_i (O_i - 1)$$

$$= \lambda w + \sum_{i=1}^{N} O_i (x_i^T y_i)^T - \sum_{i=1}^{N} x_i^T y_i$$

Taking negative LL:

$$\therefore E(w) = \sum_{i=1}^{N} x_i^T y_i - \lambda w - \sum_{i=1}^{N} O_i (x^T y_i)^T$$

$$\boxed{w_{new} = w_{old} - \eta (E(w))}$$

Substitute above Eq^n here to get updated weight vector

Date : / /

(Q2-5) $f(w) = \frac{\lambda}{2}||w||^2 + \sum_{i=1}^{N} \ln(1 + e^{-y_i w^T x_i})$

$\min_w f(w)$ s.t. $g(x) \leq 0$, where $g(x) : 1 - y_i w^T x_i \leq 0$

(taking $b=0$)

$L(w, \cancel{\beta}) = f(w) + \beta g(w)$

$= \frac{\lambda}{2}||w||^2 + \sum_{i=1}^{N} \ln(1 + e^{-y_i w^T x_i}) + \beta(1 - y_i w^T x_i)$

To eliminate x, take $\frac{\delta L(w)}{\delta x}$

$\therefore \frac{\delta L(w)}{\delta x} = \sum_{i=1}^{N} \frac{e^{-y_i w^T x_i}}{1 + e^{-y_i w^T x_i}} \times -y_i w^T - \beta \sum_{i=1} y_i w^T = 0$

$\therefore \beta = - \sum_{i=1}^{N} \frac{e^{-y_i w^T x_i}}{1 + e^{-y_i w^T x_i}}$

$= \sum_{i=1}^{N} \frac{1}{1 + e^{-y_i w^T x_i}}$

Scanned by CamScanner