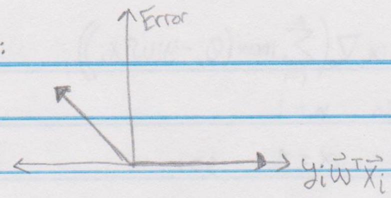
Briana Zhao
U93987617
EC 414, HW 4

4 (a) From lecture:



If $-y_i w^T x < 0$ then error $= 0$
If $-y_i w^T x > 0$ then error $= -y_i w^T x$

'tve' = positive
'-ve' = negative

If $y_i = sgn(w^T x_i)$ [error $= 0$] then $y_i w^T x_i > 0$
and $y_i w^T x_i$ is either $(+1)\cdot(+ve) =$ +ve value OR $(-1)\cdot(-ve) =$ +ve value } Thus $(-y_i w^T x_i)$ is always $\underline{-ve}$

If $y_i \neq sgn(w^T x_i)$ [error $= |w^T x|$] then $y_i w^T x_i < 0$
and $y_i w^T x_i$ is either $(-1)\cdot(+ve) =$ -ve value OR $(+1)\cdot(-ve) =$ -ve value } Thus $(-y_i w^T x_i)$ is always $\underline{+ve}$

So: $L_i(w) = \max(0, -y_i w^T x_i)$

→ if we classify correctly $(y_i = sgn(w^T x_i))$, then we choose the max between 0 and $-y_i w^T x_i$, which is a negative value. So we choose 0, which is what error should be when $y_i = sgn(w^T x_i)$.

→ If we classify incorrectly $(y_i \neq sgn(w^T x_i))$, then we choose the max between 0 and $-y_i w^T x_i$, which is a positive value. So then we choose $-y_i w^T x_i$, which is the same as $|w^T x_i|$. This is what the error should be when $y_i \neq sgn(w^T x_i)$

Therefore $L_i(w) = \max(0, -y_i w^T x_i)$

(b) Loss is always either 0 (correct classification), or $-y_i w^T x_i$ (incorrect classification)

If $h_w(x)$ classifies X correctly, then $L(w) = \max(0, -y_i w^T x_i) = 0$

so $\nabla_w L_i = \nabla_w(0) = 0$

If $h_w(x)$ classifies X incorrectly, then $L(w) = \max(0, -y_i w^T x_i) = -y_i w^T x_i$

so $\nabla_w L_i = \nabla_w(-y_i w^T x_i) = -y_i x_i$

$$\text{Same as: } \nabla L(w) = \nabla \sum_{i=1}^{n} \max(0, -y_i w^T x_i) = \sum_{i=1}^{n} \nabla \max(0, -y_i w^T x_i)$$

$$= \sum_{i=1}^{n} \nabla \begin{cases} 0 & h_w \text{ classifies correctly} \\ -y_i w^T x_i & h_w \text{ classifies incorrectly} \end{cases} \quad \begin{matrix} \leftarrow \nabla_w(0) = 0 \\ \leftarrow \nabla_w(-y_i w^T x_i) = -y_i x_i \end{matrix}$$

$$= \sum_{i=1}^{n} \begin{cases} 0 & h_w \text{ classifies } x_i \text{ correctly} \\ -y_i x_i & h_w \text{ classifies } x_i \text{ incorrectly} \end{cases}$$

Therefore, the subgradient is $\nabla_w L_i = \begin{cases} 0 & h_w \text{ classifies } x_i \text{ correctly} \\ -y_i x_i & h_w \text{ classifies } x_i \text{ incorrectly} \end{cases}$

ⓒ $w_{i+1} = w_i - \alpha \nabla_w L(w, x_i) = w_i - \alpha \nabla_w \left(\sum_{i=1}^{n} \max(0, -y_i w^T x_i) \right)$

batch size $= 1$ so $n = 1$

Step size $= 1$ so $\alpha = 1$ so:

$$= w_i - \nabla_w \left(\max(0, -y_i w^T x_i) \right)$$

Substitute in the subgradient from ⓑ:

$$\boxed{w_{i+1} = w_i - \begin{bmatrix} 0 & \text{if } h_w(x) \text{ classifies } x_i \text{ correctly} \\ -y_i x_i & \text{if } h_w(x) \text{ classifies } x_i \text{ incorrectly} \end{bmatrix}}$$