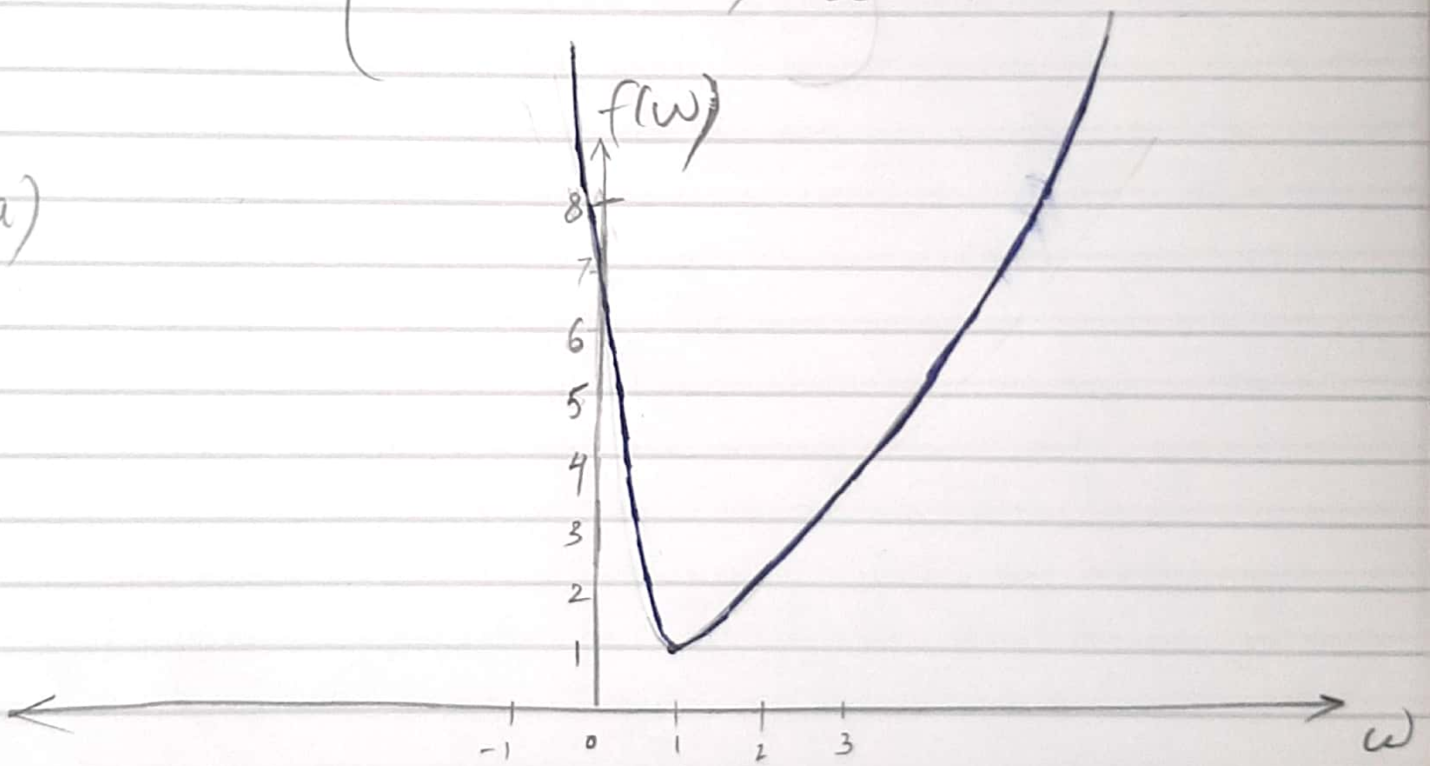
Activity 20 ECE 932

Oyar Beep Werzu

1. $f(\omega) = \begin{cases} e^{-2(\omega-1)}, & \omega < 1 \\ e^{\omega-1}, & \omega \geq 1 \end{cases}$

a)



Thus the function is convex

b) The function is not differentiable at $\omega = 1$.

The function is continuous but does not have the same LHD & RHD.

c) $$\partial f(w) = \begin{cases} -2e^{-2(w-1)} & , w<1 \\ e^{w-1} & , w>1 \\ [-2, 1] & , w=1 \end{cases}$$

Scanned with CamScanner

2. a) We have gradient descent for hinge loss as,

$$\nabla f(w)\Big|_{w^{(k)}} = \sum_{i=1}^{N} \left(-d_i x_i I_{\{d_i x_i^T w^{(k)} < 1\}}\right) + 2\lambda w^{(k)}$$

Thus for this problem, we have

$$\nabla f(w)\Big|_{w^{(k)}} = \sum_{i=1}^{m} \left(-b_i x_i I_{\{b_i x_i^T w^{(k)} < 1\}}\right) + 2\lambda w^{(k)}$$

where b_i contains the outcome of the experimental conditions (x_i) for some $i \in [1, m]$

We have to sum $-b_i x_i I$ over all the indices and then add $2\lambda w^{(k)}$ to it to calculate gradient descent.

Scanned with CamScanner

b) $w^{(k+1)} = w^{(k)} - \tau \nabla I(w)_{w^{(k)}}$

This when,

$b_i \cdot x_i^T * w > 1$,

$I(w) = 0$ by definition

Thus, $\nabla I(w) = 0$ logically.

(Derivative of a constant is 0)

Thus

$w^{(k+1)} = w^{(k)}$

So you would no longer have newer values should you continue the process.

Scanned with CamScanner

3. As $\min_{w} \|y - Xw\|_2^2 + 2\|w\|_1$

we have $\lambda = 2$.

$\tau = 1 \quad w^{(0)} = 0$.

$N = 4$ as there are 4 samples.

We know,

$$w^{(k+1)} = w^{(k)} + \tau\left(d_{i_k} - x_{i_k}^T w^{(k)}\right)x_{i_k} - \frac{\lambda\tau}{2N}\,\text{sign}\left(w^{(k)}\right)$$

for $i = 1$

* for $y_1 = 1$, $x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$$w^{(1)} = w^{(0)} + 1\left(1 - [1 \ -1]w^{(0)}\right)\begin{bmatrix} 1 \\ -1 \end{bmatrix} - \frac{2(1)}{2(4)}(0)$$

$$w^{(1)} = 0 + 1(1-0)\begin{bmatrix} 1 \\ -1 \end{bmatrix} - 0$$

$$w^{(1)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Scanned with CamScanner

$$w^{(2)} = w^{(1)} + \tau\left(d_{i_k} - x_{i_k}^{T} w^{(1)}\right) x_{i_k}$$

$$- \frac{\lambda\tau}{2N} \operatorname{sign}\left(w^{(k)}\right)$$

$$= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 1\left(1 - \begin{bmatrix} 1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$- \frac{\cancel{2}(1)}{\cancel{2}(4)} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 1(1-2)\begin{bmatrix} 1 \\ -1 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \cancel{\begin{bmatrix} 1 \\ -1 \end{bmatrix}} + \cancel{\begin{bmatrix} -1 \\ 1 \end{bmatrix}} - \frac{1}{4}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.25 \\ +0.25 \end{bmatrix}$$

for $i=2$:

for $y_2 = 2$, $x_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

$$w^{(1)} = \cancel{w^{(0)}}^{\,0} + \tau \left(d_{i_k} - x_{i_k}^{T} w^{(0)} \right) x_{i_k} - \cancel{\frac{\lambda \tau}{2N} \operatorname{sign}(w^{(0)})}^{\,0}$$

$$= 1 \left(2 - \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \begin{bmatrix} 1 \\ -2 \end{bmatrix} + 0$$

$$= 1 \cdot (2) \cdot \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$$

$$w^{(2)} = \begin{bmatrix} 2 \\ -4 \end{bmatrix} + 1 \left(2 - \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ -4 \end{bmatrix} \right) \begin{bmatrix} 1 \\ -2 \end{bmatrix} - \frac{\cancel{2}(1)}{\cancel{2}(4)} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ -4 \end{bmatrix} + 1(-8) \begin{bmatrix} 1 \\ -2 \end{bmatrix} - \begin{bmatrix} 0.25 \\ -0.25 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ -4 \end{bmatrix} + \begin{bmatrix} -8 \\ 16 \end{bmatrix} + \begin{bmatrix} -0.25 \\ 0.25 \end{bmatrix} = \begin{bmatrix} -6.25 \\ 12.25 \end{bmatrix}$$

Scanned with CamScanner

for $i = 3$

for $y_3 = -1 \quad x_3 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

$$w^{(1)} = \cancel{w^{(0)}} \to 0 + 1\left(-1 - x_{1,n}^{i\,T}(0)\right)\begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$- \cancel{\frac{\lambda 2}{2N}(0)} \to 0$$

$$w^{(1)} = 1(-1)\begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$w^{(2)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1\left(-1 - [-1 \;\; 0]\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)\begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$- \frac{2(1)}{2(4)}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \cancel{(-1 - (-1))\begin{bmatrix} -1 \\ 0 \end{bmatrix}} \to 0 - \begin{bmatrix} 0.25 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.25 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.75 \\ 0 \end{bmatrix}$$

Scanned with CamScanner

for $i = 4$

for $y_4 = -2$ $\quad x_4 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

$$w^{(1)} = \cancel{w^{(0)}}^{\,0} + 1\left(-2 - \cancel{x_{i_k}^{T}(\,)}^{\,0}\right)\begin{bmatrix} -2 \\ 1 \end{bmatrix} - \cancel{\frac{\lambda \tau}{2N}(w^{(0)})}^{\,0}$$

$$w^{(1)} = -2\left(\begin{bmatrix} -2 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$$

$$w^{(2)} = w^{(1)} + \tau\left(d_{i_k} - x_{i_k}^{T} w^{(k)}\right) x_{i_k} - \frac{1\cancel{(2)}}{\cancel{2}(4)} \operatorname{sign}\begin{pmatrix} 4 \\ -2 \end{pmatrix}$$

$$= \begin{bmatrix} 4 \\ -2 \end{bmatrix} + 1\left(-2 - [-2 \;\; 1]\begin{bmatrix} 4 \\ -2 \end{bmatrix}\right)\begin{bmatrix} -2 \\ 1 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 4 \\ -2 \end{bmatrix} + 8\begin{bmatrix} -2 \\ 1 \end{bmatrix} + \begin{bmatrix} -0.25 \\ 0.25 \end{bmatrix}$$

$$= \begin{bmatrix} -12.25 \\ 6.25 \end{bmatrix}$$

Scanned with CamScanner

Data used for first B updates:

$$\boxed{\begin{matrix} \tau, w^{(0)}, w^{(1)}, w^{(2)}, w^{(3)}, w^{(4)}, \\ w^{(5)}, \lambda, N, x_{i_k}, y_i \end{matrix}}$$

for every value of i in $\{1, 2, 3, 4\}$ as there are 4 training samples.