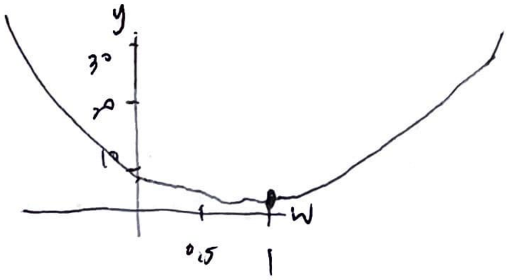
~~valid norm satisfy rules~~

Activity 20

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



No, 2 different curves connected at $w=1$

b) No, at $w=1$, we have 2 different derivative, so we use subderivative

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

2. a) $l_i(\underline{w}) = (1 - d_i \underline{x}_i^T \underline{w})_+ = \begin{cases} 1 - d_i x_i^T \underline{w}, & d_i \underline{x}_i^T \underline{w} < 1 \\ 0 & , d_i \underline{x}_i^T \underline{w} \geq 1 \end{cases}$

$\nabla l_i(\underline{w}) = \begin{cases} -d_i \underline{x}_i & , d_i \underline{x}_i^T \underline{w} < 1 \\ 0 & , d_i \underline{x}_i^T \underline{w} \geq 1 \end{cases} = -d_i \underline{x}_i I\{d_i \underline{x}_i^T \underline{w} < 1\}$

cost $\nabla f(\underline{w})|_{w^{(k)}} = \sum_{i=1}^{N} (-d_i \underline{x}_i I\{d_i \underline{x}_i^T \underline{w}^{(k)} < 1\})$

Gradient descent $\underline{w}^{(k+1)} = \underline{w}^{(k)} - \tau \nabla f(\underline{w})|_{w^{(k)}}$

b) If we have substaintial margin the derivative is 0, converge stop.

3. a) cyclic SGD $\lambda = 2$

$i_k = 1, 2, 3, 4, 1, 2, 3, 4 \dots$, $\tau = 1$, $w^{(0)} = 0$

$w^{(k+1)} = w^{(k)} + \tau (d_{i_k} - \underline{x}_{i_k}^T w^{(k)}) \underline{x}_{i_k} - \frac{\lambda \tau}{2N} \operatorname{sign}(\underline{w}^{(k)})$

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

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

② $w^2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + (2 - [1\ -2]\begin{bmatrix} 1 \\ -1 \end{bmatrix})\begin{bmatrix} 1 \\ -2 \end{bmatrix} - \frac{2}{8}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$(1+2) - 1$

$= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{4} \\ -\frac{1}{4} \end{bmatrix} = \begin{bmatrix} -\frac{1}{4} \\ \frac{5}{4} \end{bmatrix}$

data use for the 1st 6 updates

① x_1, y_1 ④ x_4, y_4

② x_2, y_2 ⑤ x_5, y_5

③ x_3, y_3 ⑥ x_6, y_6

$\begin{bmatrix} 1 \\ -2 \end{bmatrix}$

↓

$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$