Homework#3 ; Support Vector Machines (SVM)

By: Vikrant Bhati

1.) N pairs of (x_i, y_i) where,

$x_i \equiv$ the feature vectors

$y_i \equiv$ binary class label $(-1, 1)$

for linear SVM classifier: $y = wx_i + b$ the lossfunction is defined as:-

$$L = \sum_{i=1}^{N} \max(0, 1 - y_i(wx_i + b)) + \alpha||w||^2 \quad (1)$$

where α is regularization parameter

for gradient descent $\frac{\partial L}{\partial w \partial b} = 0$

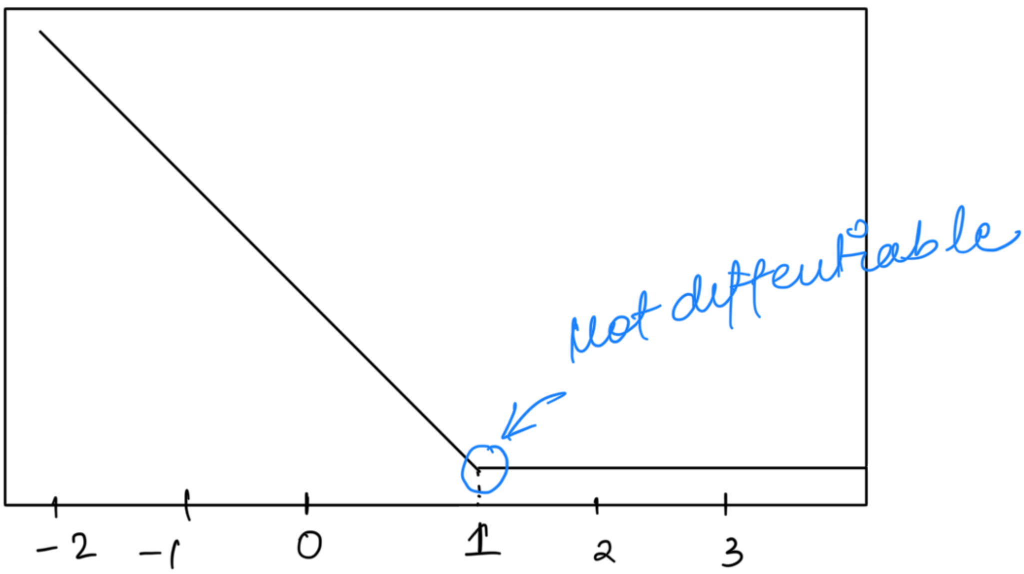
where above equation (1) can be broken into

hinge loss i.e $\sum_{i=1}^{N} \max(0, 1 - y_i(wx_i + b)) \quad (2)$

and regularization term:

$$\alpha||w||^2 \quad (3)$$

Now differentiating eq-(2) is not possible directly because hinge loss is not differentiable at ($t=1$)



Hinge loss function

$\therefore$ we will use concept of subgradient

$$H = \sum_{i=1}^{n} \max\left(0, 1 - y_i(w \cdot x_i + b)\right)$$

where $H=0$ when points are correctly classified

and

$$\sum_{i=1}^{n} 1 - y_i(w \cdot x_i + b) \text{ otherwise}$$

$\therefore \frac{\partial H}{\partial w} = -y_i x_i$ for missclassified points — (1)

$\frac{\partial H}{\partial b} = -y_i$ for miss classified point — (2)

$\frac{\partial H}{\partial w} = \frac{\partial H}{\partial b} = 0$ for correctly classified points — (3)

Simillaly

$$\frac{\partial}{\partial b \partial w}(\alpha \|w\|^2)$$

$$\frac{\partial}{\partial b}(\alpha \|w\|^2) = 0 \quad — (4)$$

$$\frac{\partial}{\partial w}\alpha \|w\|^2 = 2\alpha|w| \quad — (5)$$

Combining 1 2 3 4 and 5

$$\frac{\partial L}{\partial \vec{w}} = \begin{cases} \sum_{i=1}^{n} 2\alpha|\vec{w}| - y_i x_i & \text{for missclassified points} \\ \sum_{i=1}^{n} 2\alpha|\vec{w}| & \text{for correctly classified point} \end{cases}$$

$$\frac{\partial L}{\partial b} = \begin{cases} \sum_{i=1}^{n} -y_i & \text{for missclassified points} \\ 0 & \text{for correctly classified points} \end{cases}$$