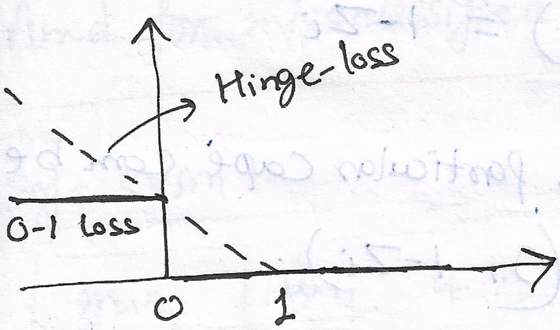
Loss Function (Hinge Loss) based interpretation



$y_i f(x_i)$
$= y_i(w^T x_i + b) = z_i$

This is what we talk about 0-1 Loss

$z_i > 0 \Rightarrow x_i$ is correctly classified
$z_i < 0 \Rightarrow x_i$ is incorrectly classified

Hingeloss is a straight line from $-\infty$ to 1 And after that it becomes $= 0$.

From, the graph we can see that it is not differentiable at ~~$z_i = 0$~~ $x = 1$

$z_i \geq 1$ hinge loss $= 0$
$z_i < 1$ hinge loss $= 1 - z_i$

Hinge loss

$\boxed{\max(0, 1 - z_i)}$

This is the way we can write Hinge loss

Case I $z_i \geq 1$

$\Rightarrow 1 - z_i < 0$ i.e. -ve value.

max 0 and $1 - z_i$ is 0.

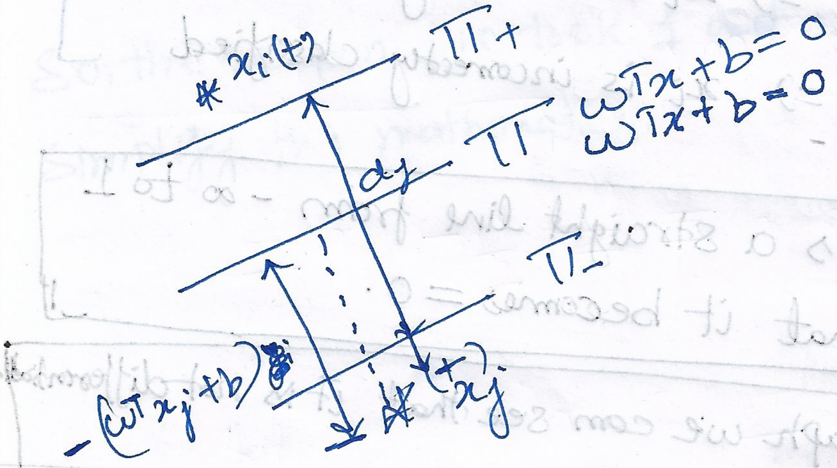
i.e. $\max(0, 1 - z_i) = 0$

Case II $Z_i < 1 \Rightarrow 1 - Z_i > 0$

$\Rightarrow \max(0, 1 - Z_i) = 1 - Z_i$

$\therefore$ Hinge loss in this particular case can be written as max of $(0, 1 - Z_i)$.

Now, let's see with the help of geometry.



$d_j \rightarrow$ Total distance of $\boxed{\text{miss classified}}$ point from the Π_+

$d_j = 1 - y_i(w^T x_j + b) = 1 - Z_j$

$\varepsilon_j \rightarrow$ Distance from x_j to the Π^+

$= d_j = 1 - Z_j$

$$\boxed{\therefore \varepsilon_j = 1 - Z_j \rightarrow \text{when } x_j \text{ is missclassified.}}$$

Hence for correctly classified point $\varepsilon_i = 0$
And for missclassified point $\varepsilon_j = 1 - z_j$

Soft SVM

$$\min_{\omega, b} \frac{\|\omega\|}{2} + C \sum_{i=1}^{n} \varepsilon_i$$

$$\text{s.t. } 1 - y_i(\omega^T x_i + b) \geq \varepsilon_i \quad \forall i$$

$$\varepsilon_i \geq 0$$

Loss-minimization

$$\min_{\omega, b} \sum_{i=1}^{n} \max(0, 1 - y_i(\omega^T x_i + b)) + \lambda \|\omega\|^2$$

$\|w\| \geq 0$, $\therefore$ minimizing $\frac{\|w\|}{2}$ is same as minimizing $\|\omega\|^2$.

From the above:
$C \uparrow \Rightarrow$ Over fit and $C \downarrow \Rightarrow$ Underfit
Similarly as $\lambda \uparrow \Rightarrow$ Underfit $\Rightarrow \lambda \downarrow \Rightarrow$ overfit