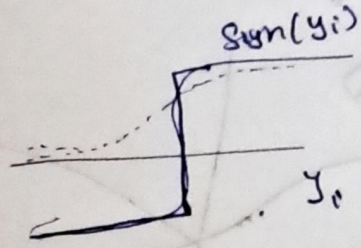
Gradient Descent for linear Classifn

Regression: loss = MSE
MAE
Regularization → L2 norm of W
→ L1 norm of W

linear classifiers:

(i) miss classification rate

$$\sum_{i}^{N} \cdot \mathbb{1}_{\text{Sign}(y_i) \neq t_i}$$



$y_i = w^T x_i + b$; $\frac{1}{N} \sum \mathbb{1}_{\text{sign}(w^T x_i + b) \neq t_i}$

$t_i \in \{-1, 1\}$

Replace $\text{Sign}(y_i)$ with $\sigma(y_i) = \frac{1}{1+e^{-y_i}}$

lets, $y_i = \sigma(w^T x_i + b)$ interpreted it as $p(t_i = 1 | x_i)$

$t_i \in \{0, 1\}$ $y_i \in (0, 1)$

$w^T x_i + b \in (-\infty, \infty)$

(ii) Loss $= \frac{1}{2N} \sum (y_i - t_i)^2$ can work but its error is Gaussian dist.

(iii) BCE : KL divergence b/w 't' and 'y'

{Binary cross entropy} → Binary cross entropy

$$\text{BCE} = -\sum_i \sum_c t_c \log y_c = -\sum_i t_i \log y_i + (1 - t_i) \log (1 - y_i)$$

↓ assume $t_i = 0$ and $y_i = 0$

$t_i = 1$, $y_i = 0$

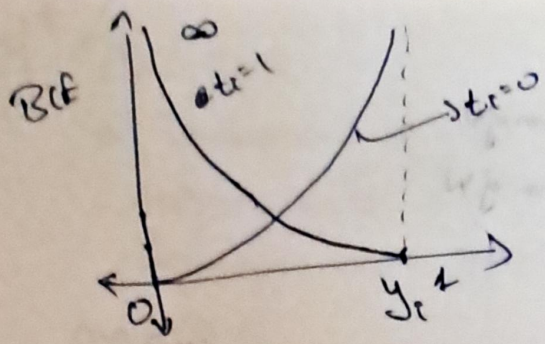
$-(1 \log 0 + 0 \log 1)$

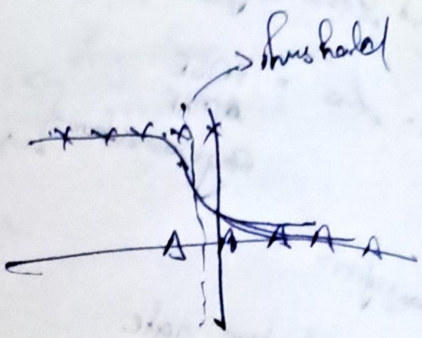
$= \infty$

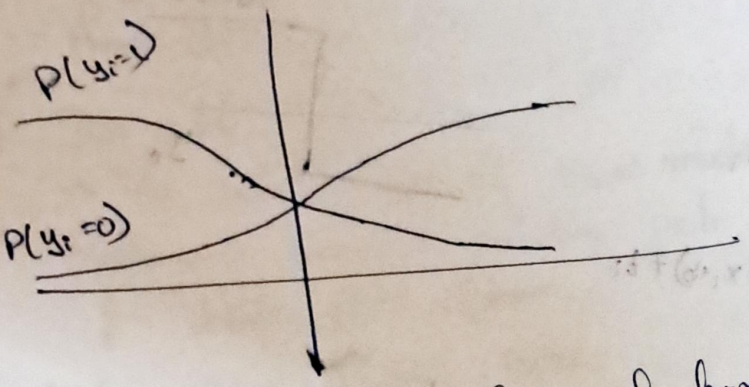
$-(0 + 1 \log 1)$

$t_i = 1$ and $y_i = 1$

$-1 \log 1 - 0$







→ By rescaling we can sharpen the boundary.

like: $\sigma(\alpha(wx+b))$

$$BCE = -t_i \log y_i - (1-t_i)\log(1-y_i)$$

$$L = +t_i \log(\sigma(wx_i)) - (1-t_i)\log(1-\sigma(wx_i))$$

$$\frac{\partial L}{\partial w} = \left[-\frac{t_i}{y} + \frac{1-t_i}{1-y_i}\right]\left[y_i(1-y_i)\right]x_i$$