Logistic Regression

- → Introduction
- → Business Case
- → Step Function
- → Sigmoid Function
- → Geometric Intuition
- → Log Loss
- → Optimization → H/W

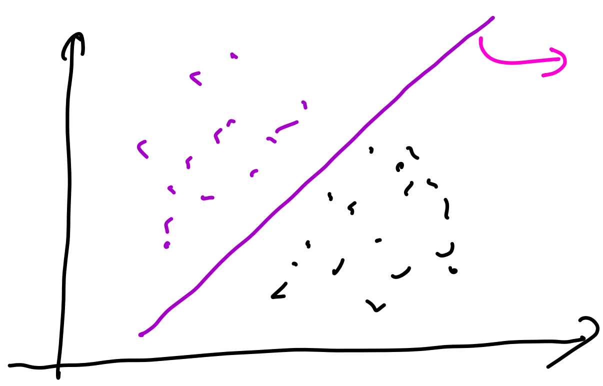
$(-\infty, \infty)$

Linear Model

$$\hat{y}_i = w^T x_0 + w_0 \rightarrow [-\infty, \underline{\infty}]$$

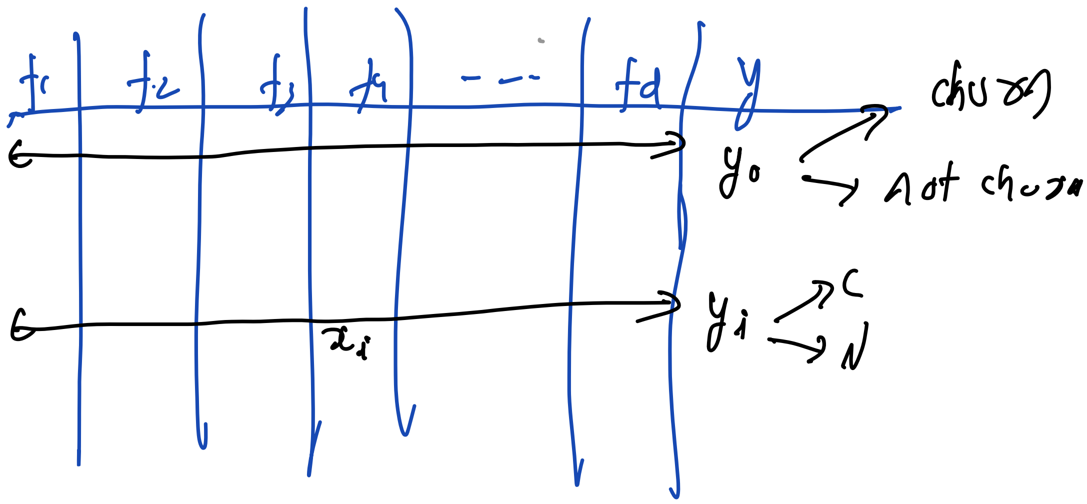
$$\text{Loss} = (y - \hat{y})^2$$

→ Telecom



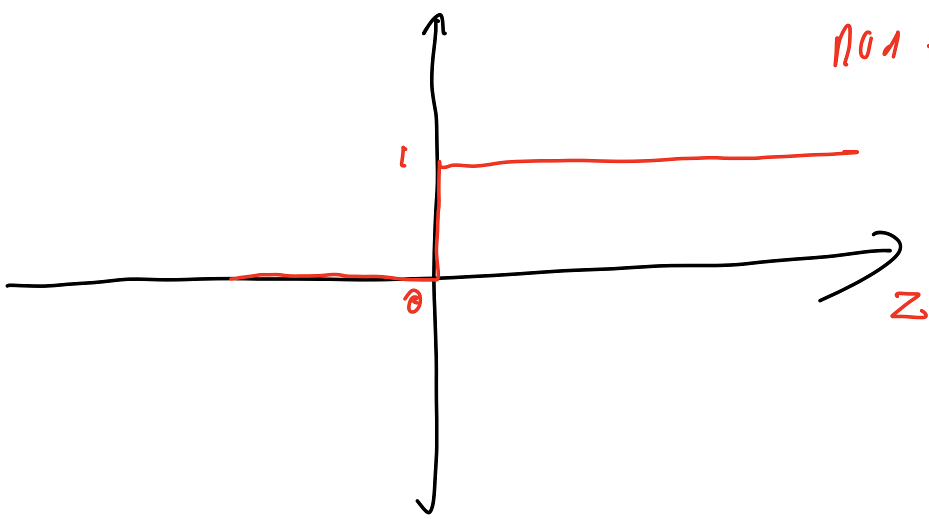
$\rightarrow w^T x_i + w_0$

$(-\infty, \underline{\infty})$

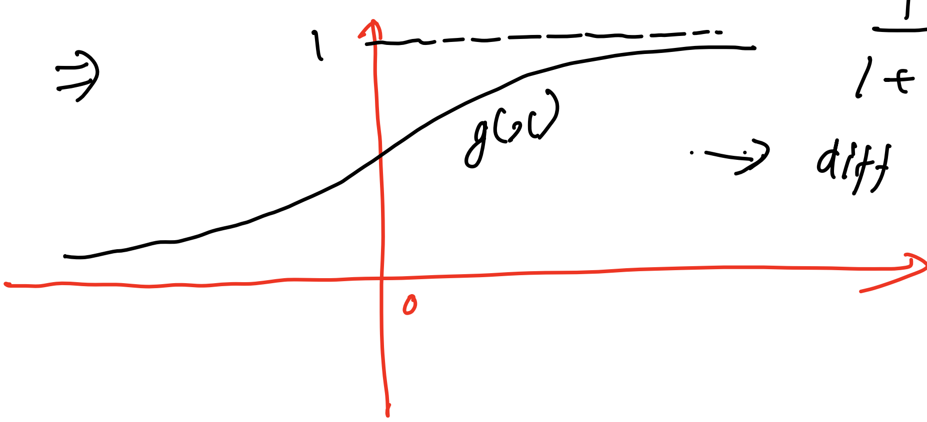


$$z = w^T x_i + w_0 \xrightarrow{g()} \{0, 1\}$$

non-differentiable



$\sigma \Rightarrow$



$$\frac{1}{1+e^{-z}}$$

$\rightarrow$ diff

$$g(w^T x + w_0) \rightarrow [0, 1]$$

Rain

.60%

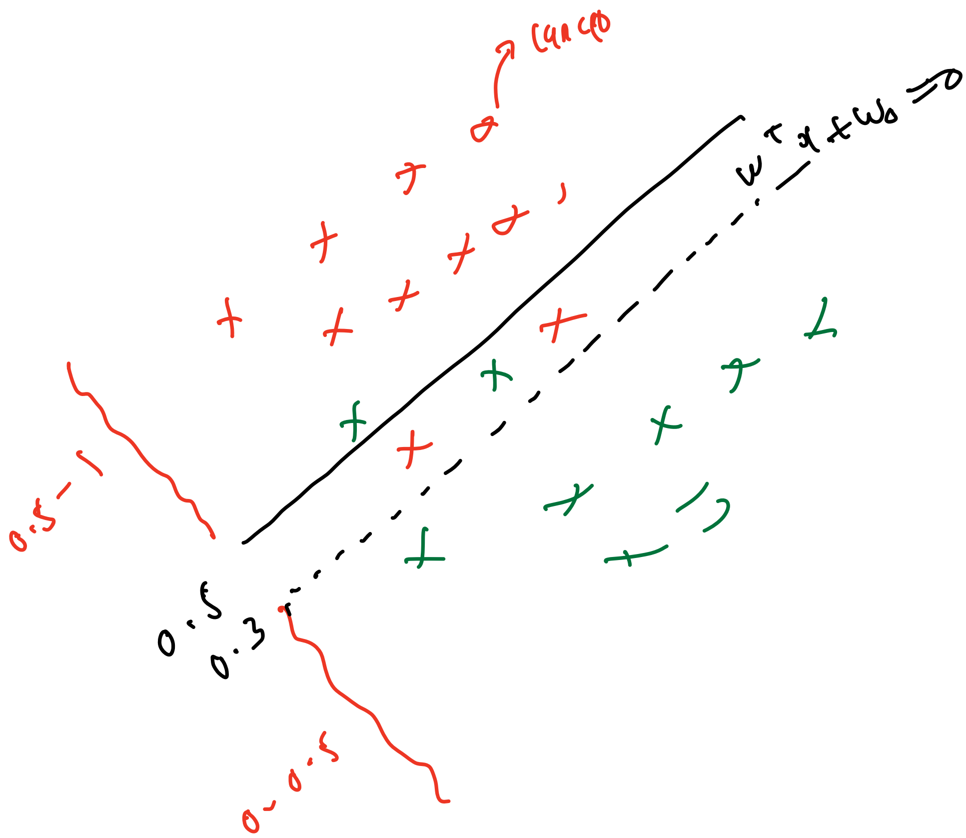
.50%

.90

No Rain

.25

$$g(z_i) = \text{sigmoid}(z_i) = \frac{1}{1 + \exp(-z_i)}$$



.1

$$D = \{x_i, y_i\}_{i=1}^{N}$$

$$y_i \in \{0, 1\}$$
$$x_i \in R^d$$

1. Linear seperator $\pi_d : w^t x + w_0 = 0$

2. $w^T x + w_0 = z_i \in (-\infty, \infty)$

$\sigma(z_i) \Downarrow$ squash

$(0, 1)$

prob

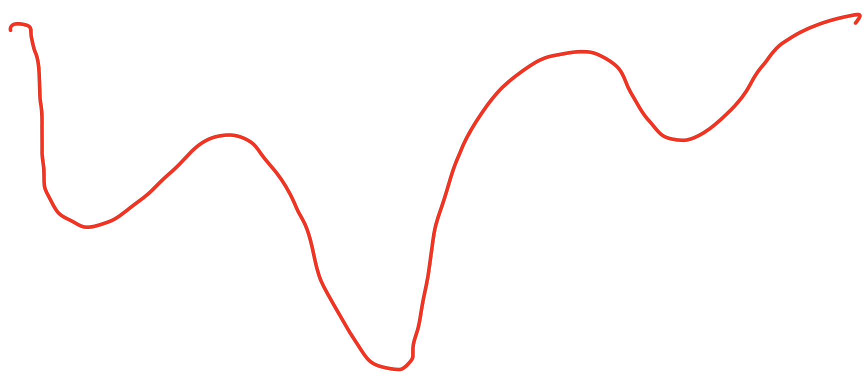
3. $y_i \in$ 0 or 1

$\hat{y}_i \in (0, 1)$ $\Rightarrow$ compare

y_i vs $\hat{y}_i$

1   0.6

$$\frac{1}{1+e^{-z}}$$



$M_1 \Rightarrow 0.81$

x	y_i	$\hat{y}_i$
—	1	0.6
—	0	0.4
—	1	0.55
—	0	0.45

$\Downarrow$

100 %

$M2 \Rightarrow 0.10$ M_2

x	y_i	$\hat{y}_i$
—	1	0.90
—	0	0.02
—	1	0.90
—	0	0.1

$\Downarrow$

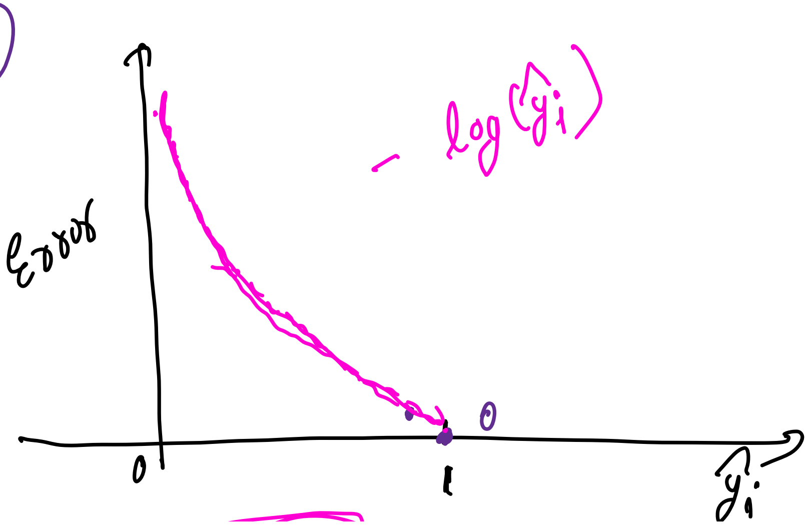
100%

$$\text{log-loss} \Rightarrow \sum -y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i)$$

$$y_i = 1 \Rightarrow -\cancel{y_i}^{1} \log(\hat{y}_i)$$

$$y_i = 0 \Rightarrow -\log(1-\hat{y}_i)$$

$y_i = 1$



$\hat{y} = 0.01$

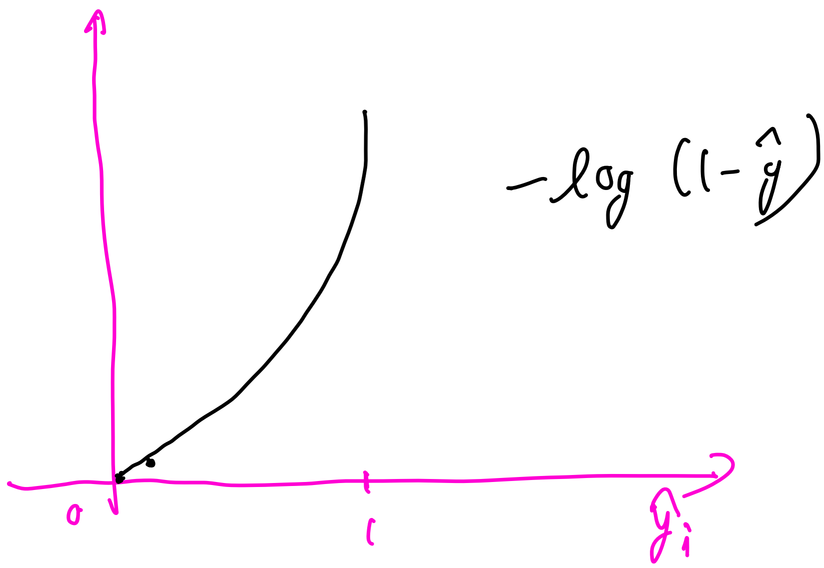
$y_i = 1$

$y_i = 0$

$\hat{y}_i = 0.0001$

$\hat{y}_i = 0.1$

$\hat{y}_i = 1$



$$\min_{w_1} \sum_{i=1}^{n} \left[-y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i) \right] + \lambda \sum w_j^2$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

$$\sigma'(z) = \sigma(z) \cdot [1 - \sigma(z)]$$

$$L = \min_{w_1} \sum_{i=1}^{n} \left[-y_i \log(\hat{y}_i) - (1-y_i) \log(1-\hat{y}_i) \right]$$

1 2

$$\frac{\partial L}{\partial w} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i) x_{ij}$$