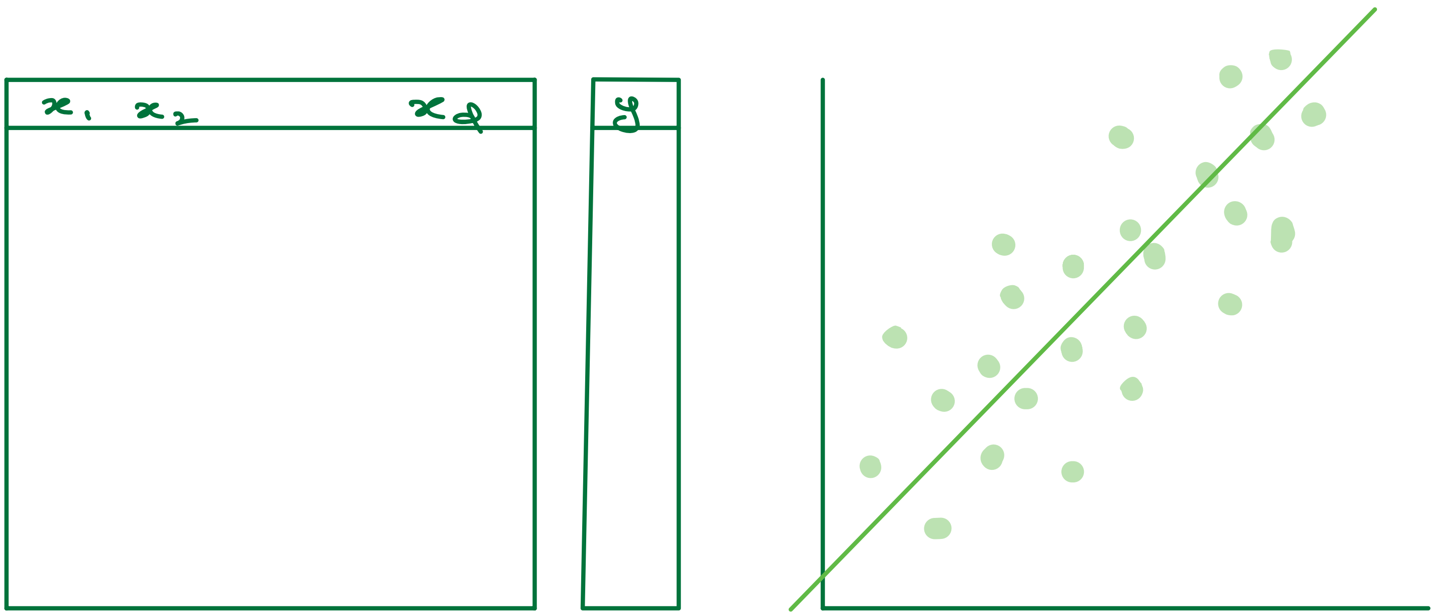
Linear Model

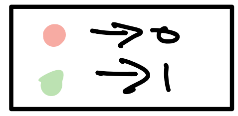
* Linear Regression

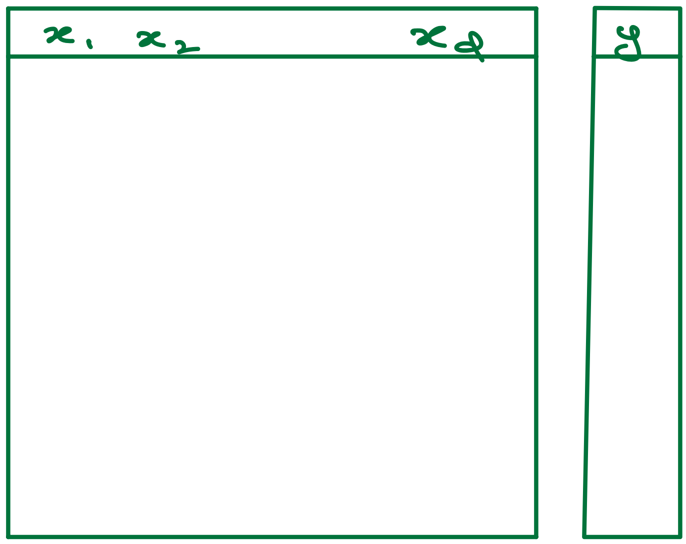


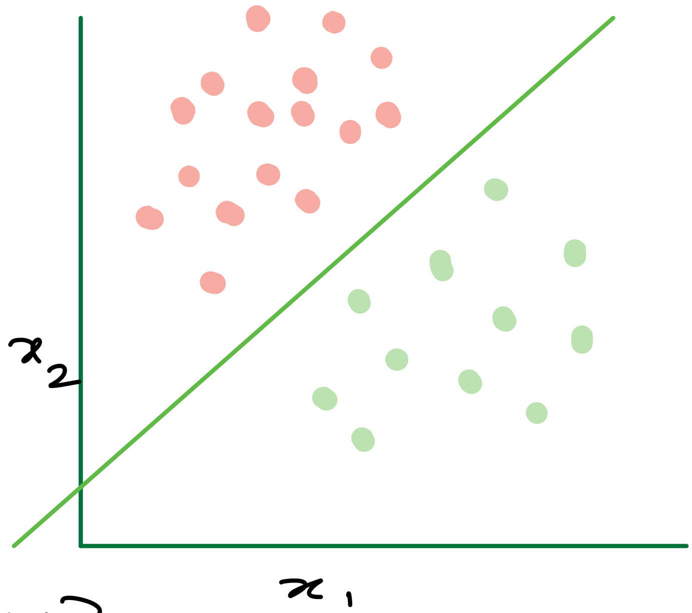
$x \in \mathbb{R}^d$ $y \in \mathbb{R}$

1. Dataset ⇒ $\{(x^{(i)}, y^{(i)})_{i=1}^{n} ; x \in \mathbb{R}^d ; y \in \mathbb{R}\}$
2. Hyperplane/Model ⇒ $w^T x + w_0$
3. Prediction ⇒ $x_i \Rightarrow w^T x_i + w_0 \Rightarrow \hat{y} \in \mathbb{R}$
4. Loss function ⇒ $\frac{1}{n}$ SQ – ⇒ $\frac{1}{n}\sum_{i=1}^{n}(\hat{y}-y)^2$

Classification







$x \in \mathbb{R}^d$ $\quad y \in \{0,1\}$ (Binary Classification

① Dataset $\Rightarrow \{(x^{(i)}, y^{(i)})_{i=1}^{n} ; x \in \mathbb{R}^d ; y \in \{0,1\}\}$

② Hyperplane/Model $\Rightarrow$ $w^T x + w_0$

③ Prediction $\Rightarrow$ $x_i \Rightarrow (w^T x_i + w_0 \Rightarrow \hat{y} \in \{0,1\}$

④ Loss function $\Rightarrow$

$$\hat{y} \Rightarrow \mathbb{R} \Rightarrow (-\infty, \infty)$$

$$\downarrow$$

$$f(\hat{y}) \longrightarrow \begin{cases} 1 : \hat{y} \geq 0 \\ 0 : \hat{y} < 0 \end{cases}$$

$$\downarrow$$

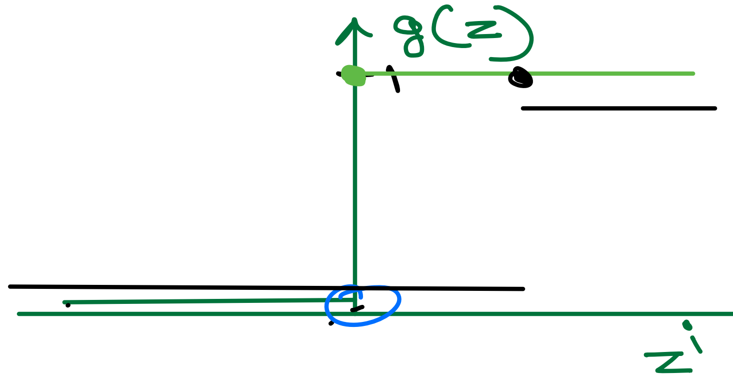
$$(0, 1)$$

Logistic Regression

$z \Rightarrow w^T x + w_0$

$z \rightarrow (-\infty, \infty)$

$g(z) \Rightarrow \{0, 1\}$



$$g(z) = \begin{cases} 1 : z \geq 5 \\ 0 : z < 5 \end{cases}$$

* threshold ⇒ 0

Sigmoid / Logistic

① $z^{(i)} \rightarrow \infty$

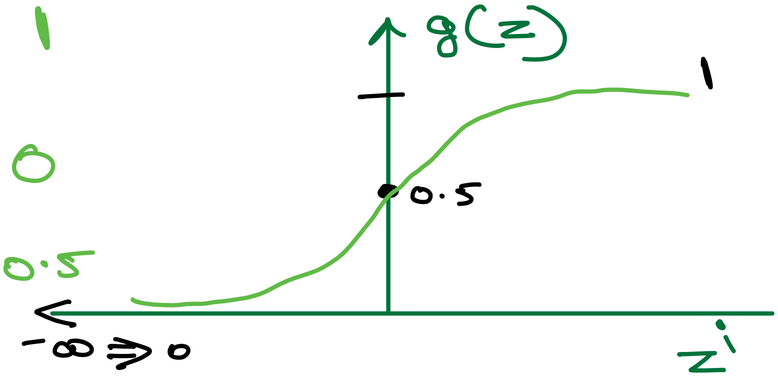
② $z^{(i)} \rightarrow -\infty$

③ $z^{(i)} \rightarrow 0$

$g(z') = 1$

$g(z') = 0$

$g(z') = 0.5$



$g(z) \Rightarrow (0, 1)$ $\not\Rightarrow$ $\{0, 1\}$

(range) (set)

$(0, 1)$

$0.3 \rightarrow 1$

$0.3 \rightarrow 0\ (0.7)$

if $g(z) > 0.5$ then class 1
else class 0

$$\text{Sigmoid} = \frac{1}{1 + e^{-z}}$$

$\hat{y}_i \Rightarrow$ Prediction

$$z = w^T x + w_0$$

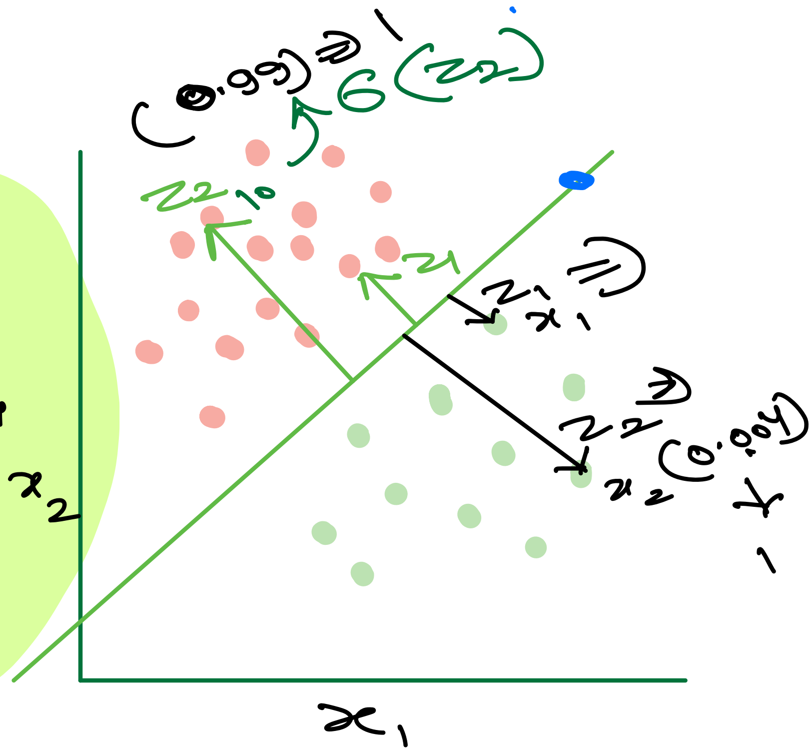
$$\hat{y} \Rightarrow g(z) \longrightarrow (0, 1)$$

$z_1 \Rightarrow -1 \Rightarrow g(z_1) \Rightarrow 0.26$

$z_2 \Rightarrow -10 \Rightarrow g(z_2) \Rightarrow 0.0004$

$z_1 \Rightarrow 1 \Rightarrow g(z_1) \Rightarrow 0.73$

$z_2 \Rightarrow 10 \Rightarrow g(z_2) \Rightarrow 0.99$



$$\hat{y}^{(i)} = P\left(y^{(i)} = 1 / x^i\right)$$

Class 0

$1 - P_{y^i = 1}$

Probality of point belonging to Class 1 given x_i feature

$$x_i \longrightarrow w^T x_i + w_0 \longrightarrow \underset{(-\infty, \infty)}{z} \longrightarrow \sigma(z)$$

$$\sigma(z) \longrightarrow \hat{y} \Rightarrow P(1 | x_i) \longrightarrow \text{apply threshold}$$

Loss Function

$$y_i \Rightarrow \sigma(w^T x + w_0)$$

$$\Rightarrow$$

$$\text{LR MSE} \Rightarrow \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (w^T x + w_0)\right)^2$$

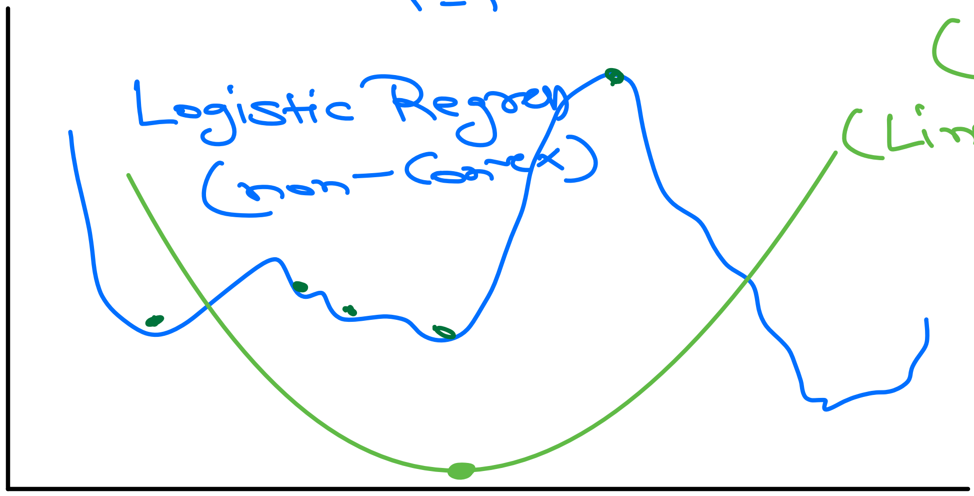
$\rightarrow y_i - \hat{y}_i$

$\cdot - \sigma(\;)$

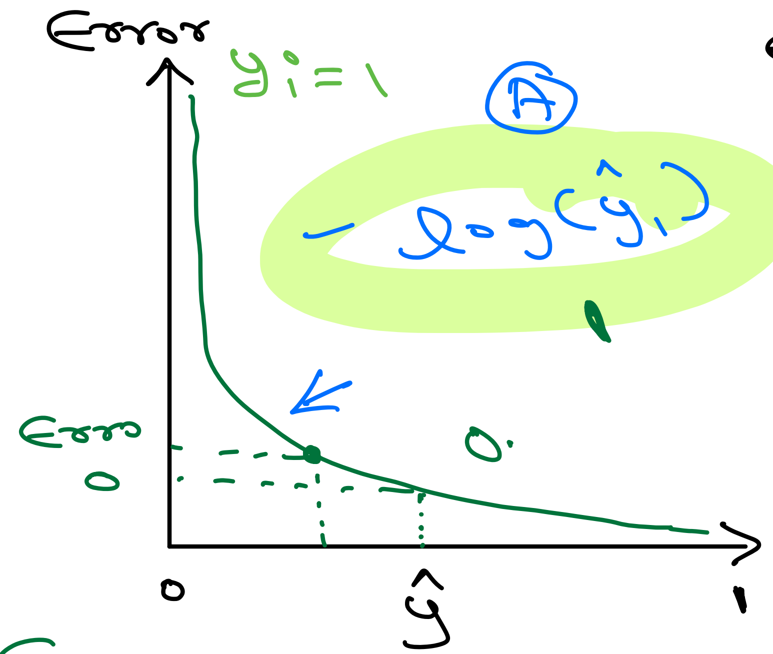
$$\hat{y} \Rightarrow \sigma(w^T x + w_0)$$

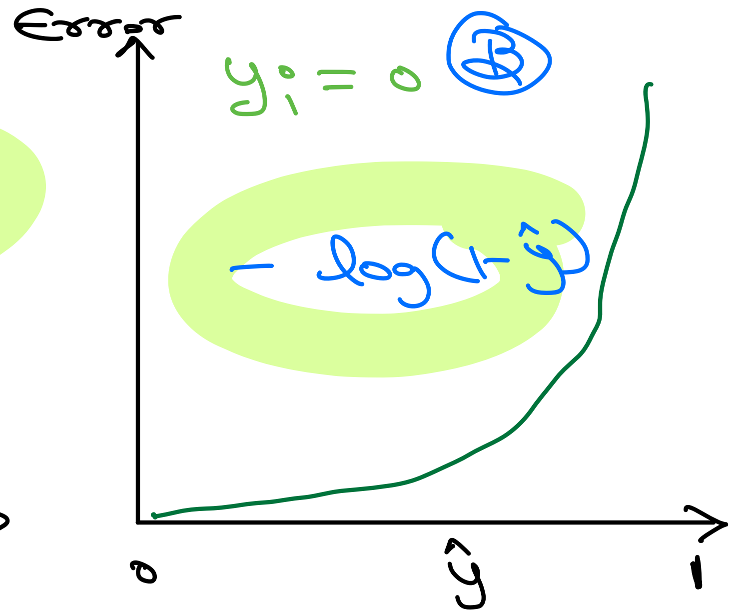
y_i

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sigma(w^T x + w_0)\right)^2$$



Log Loss Function





Case-1:

if $y_i = 1$

$\hat{y} \Rightarrow 0$ Error $\Rightarrow \infty$

Case-2:

if $y_i = 1$

$\hat{y} \Rightarrow 0.999$ Error $\Rightarrow 0$

Case 3:

if $y_i = 0$

$\hat{y} \Rightarrow 0.999$ Error $\Rightarrow \infty$

Case-4:

if $y_i \Rightarrow 0$

$\hat{y} \Rightarrow 0$ Error $\Rightarrow 0$

Log-Loss Eqn

$$-y^{(i)} \cdot \log(\hat{y}^{(i)}) - (1-y^{(i)}) \times \log(1-\hat{y}^{(i)})$$

Case-1 $\quad y^i \Rightarrow 1$

$\hat{y} \Rightarrow 0.4$

$$-1 \times \log(0.4) - \underbrace{(1-1)}_{\to 0} \times \log(1-4)$$

Case 2

$y_i \Rightarrow 0$

$$\underbrace{0 \times \log(\hat{y})}_{0} - (1-0) \times \log(1-\hat{y})$$

Optimization

$\Rightarrow L \Rightarrow -y^{(i)} \cdot \log(\hat{y}^{(i)}) - (1-y^{(i)}) \times \log(1-\hat{y}^{(i)})$

(Part A) (Part B)

$\frac{\partial L}{\partial \omega}$

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

$\sigma' z \Rightarrow \sigma z \times (1 - \sigma(z))$

Part $\Rightarrow -(y_i \times (\log \hat{y}))$

$\sigma(\omega^T x + \omega_0)$

z

$$\frac{\partial L_A}{\partial \omega_j} \Rightarrow \frac{\partial A}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial \omega_j}$$

$\frac{\partial z}{\partial \omega_j} = x_j$

$$\Rightarrow -\frac{y_i}{\cancel{\hat{y}_i}} \times \cancel{\hat{y}}(1-\hat{y}) \times x_j$$

Part B $\Rightarrow -(1-y^{(i)}) \times \log(1-\hat{y}^{(i)})$

$$\frac{\partial B}{\partial w_j} \Rightarrow \frac{\partial w_B}{\partial(1-\hat{y}_i)} \times \frac{\partial(1-\hat{y}_i)}{\hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial w}$$

$$\Rightarrow \left(\cancel{\frac{1-y_i}{1-\hat{y}_i}} \times -1 \times \hat{y} \times (\cancel{1-\hat{y}}) \times x_j \right)$$

$$\Rightarrow (1-y_i) \times \hat{y} \times x_j$$

$$\frac{\partial L}{\partial w_j} \Rightarrow \frac{-y_i}{\cancel{\hat{y}_i}} \times \cancel{\hat{y}}(1-\hat{y}) \times x_j + (1-y_i) \times \hat{y} \times x_j$$

* Gradient Descent Function

$$w_j \Rightarrow w_j - \eta\,(\partial A + \partial B)$$
