$$(AB)^T = B^T A^T$$

$$J(\theta) = (X\theta - Y)^2$$

row vector   col vector

$$= (X\theta - Y)^T (X\theta - Y)$$

$5^T = 5$

$$= (\theta^T X^T - Y^T)(X\theta - Y)$$

$Y^T X \theta$

$[1\ 2\ 3]\begin{bmatrix}1\\2\\3\end{bmatrix}$

$$= (\theta^T X^T X\theta - \underbrace{Y^T X\theta}_{Y^T X} - \underbrace{\theta^T X^T Y}_{1\times(n+1)\ (n+1\times m)\ (m\times 1)} + Y^T Y)$$

$= [1\times 1]$

$\nabla_\theta J(\theta) = 0$ ← Minimize Loss

$$\frac{\partial\, \theta^T A \theta}{\partial \theta} = 2A\theta$$

$$= \cancel{2}X^T X\theta - \cancel{2}Y^T X = 0$$

$$\Rightarrow (X^T X)^{-1} X^T X \theta = (X^T X)^{-1} Y^T X$$

$$\Rightarrow 1\ \theta = (X^T X)^{-1} Y^T X \ ✓$$

$$= (X^T X)^{-1} X^T Y \ ✓$$

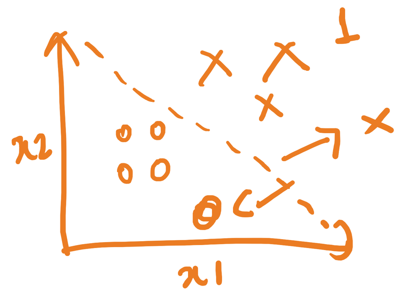
$\boxed{n+1\times m}$ $m\times n+1$ $n+1\times m \times m\times\boxed{1}$

Closed Form solution for Regression (Normal equations)

$$= \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} \begin{array}{l} \}1 \\ \\ \}n \end{array}$$

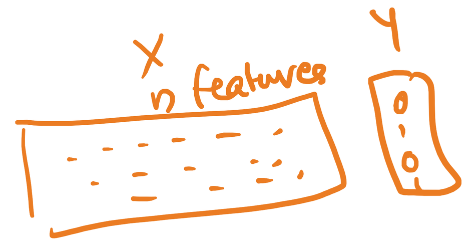
Logistic Regression

$\hookrightarrow$ X Binary Classification Algorithm



$X \in \mathbb{R}^n$

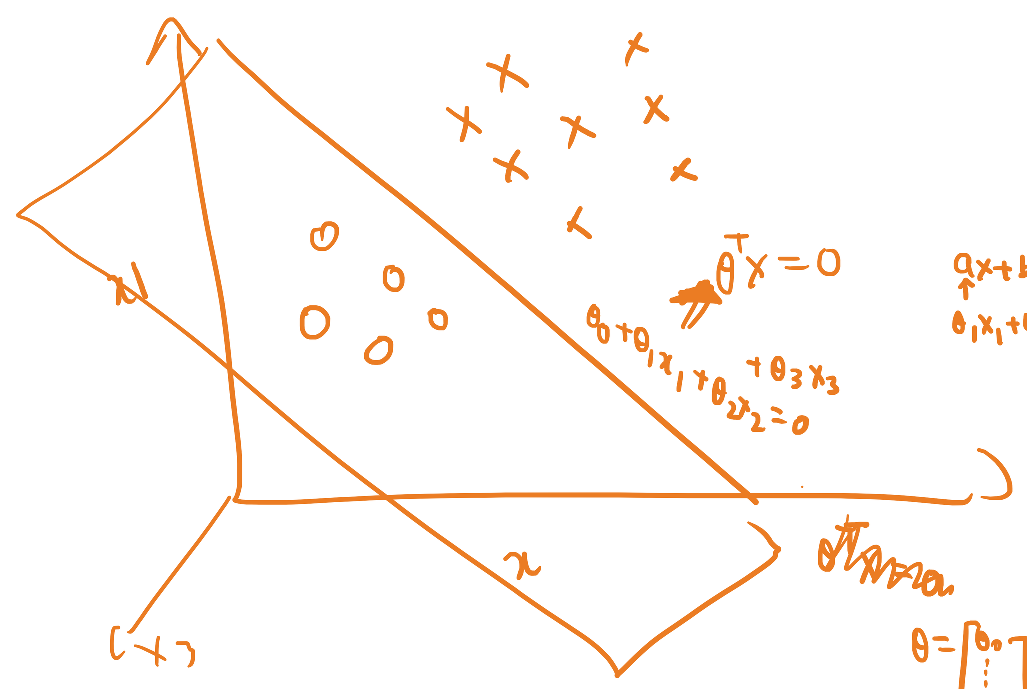
$y \in (0,1)$



$X_{new} \in \mathbb{R}^n$

$\downarrow$

$Y_{pred} = ?$



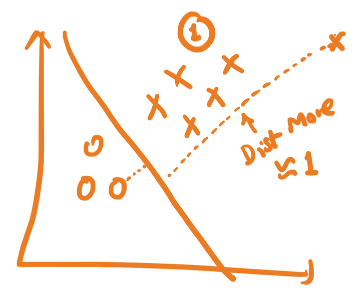
$\theta^T x = 0$

$\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 = 0$

$ax + by + c = 0$

$\theta_1 x_1 + \theta_2 x_2 + \theta_0 = 0$

$$\theta = \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_n \end{bmatrix}_{n+1 \times 1} \qquad X^{(i)} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}_{n+1 \times 1}$$



Goal:

$P(y|x)$

$\hookrightarrow P(y=1|x) = 0.70$

$P(y=0|x) = 0.30$

$\frac{ax_1 + by_1 + c}{\sqrt{a^2+b^2}}$ $\quad \frac{\sum \theta_i x_i}{\cancel{\sqrt{\ }}}$

Dist $= \alpha$

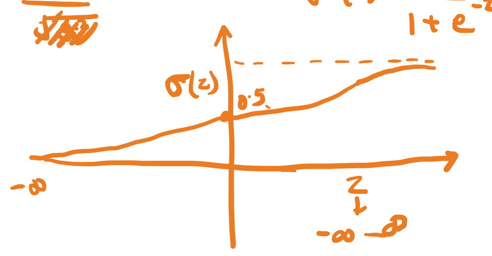
$(-\infty \text{ — } \infty)$

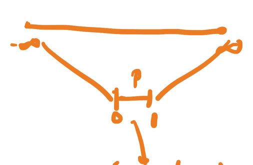
$\downarrow f$

$(0,1)$

$\theta_0 + \sum_{i=1}^{n} \theta_i x_i$

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





$\hat{y} = \cancel{\ \ }$ where $z = \theta^T x$

$P(y=1|x) = \sigma(\theta^T x) = \frac{1}{1+e^{-\theta^T x}} \in (0,1)$

$\downarrow$

$\boxed{\theta}$ Confidence that given x is +ve.

where $h_\theta(x) = \sigma(\theta^T x)$

$$\left[\begin{array}{l} P(y=1|x) = \cancel{h_\theta(x) = \sigma(\theta^T x)} = h_\theta(x) \quad 1 \\ P(y=0|x) = 1 - h_\theta(x) \quad 2 \end{array} \right.$$

$P(y|x) = [h_\theta(x)]^y [1 - h_\theta(x)]^{1-y}$ Bernoulli Dist.

$y=1$ $y=0$

0.01 0.99

Likelihood Compute.

m examples in training set.

$L(\theta) = \prod_{i=1}^{m} P(y^{(i)}|x^{(i)})$ should be high.

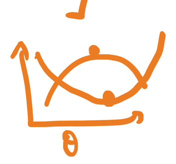
$\text{Log } L(\theta) = \sum_{i=1}^{m} \text{Log } P(y^{(i)}|x^{(i)})$

$\|$

$$= \sum_{i=1}^{m} \left[\underbrace{y^i \text{Log}(h_\theta(x^i)) + (1-y^i) \text{Log}(1-h_\theta(x^i))}_{\text{Log Likelihood}} \right]$$

Loss $= -LL(\theta)$

$\hookrightarrow$ Binary Cross Entropy.



$\frac{\partial L}{\partial \theta}$ = Loss $L \rightarrow = -\sum_{i=1}^{m} \left[y^{(i)} \log\left(h_\theta(x^{(i)})\right) + (1-y^{(i)}) \log(1-h_\theta(x^{(i)})) \right]$

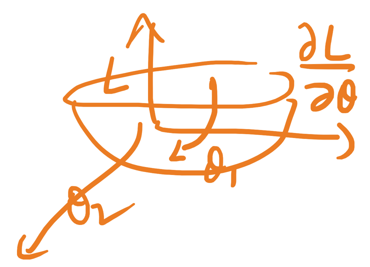
↳ Goal is to minizme Loss wrt θ.

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

$\frac{\partial \sigma}{\partial z} = \frac{+1}{(1+e^{-z})^2} (e^{-z}) = \left(\frac{1}{1+e^{-z}}\right)\left(1 - \frac{1}{1+e^{-z}}\right)$

$\sigma'(z) = \sigma(z)\{1 - \sigma(z)]$

$= -\sum_{i=1}^{m}$ Solvi



$\text{grad}_j = \frac{\partial L}{\partial \theta_j} = \sum_{i=1}^{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)}$

Gradient Descent

$$\begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} = \begin{bmatrix} \theta \end{bmatrix} - \eta \cdot \begin{bmatrix} \frac{\partial L}{\partial \theta} \end{bmatrix} \rightarrow \begin{bmatrix} \frac{\partial L}{\partial \theta_0} \\ \frac{\partial L}{\partial \theta_1} \\ \vdots \\ \frac{\partial L}{\partial \theta_n} \end{bmatrix}$$