$$\alpha = 0.4$$

$$\begin{cases} \Sigma_0^{RDA} = \underbrace{(1-\alpha)}_{0.6} \cdot \Sigma_p + \underbrace{\alpha}_{0.4} \cdot \hat{\Sigma}_0 \\ \Sigma_0^{RDA} = \underbrace{\alpha}_{0.4} \cdot \Sigma_p + \underbrace{(1-\alpha)}_{0.6} \cdot \hat{\Sigma}_0 \end{cases}$$

$$\boxed{0 \leq \alpha \leq 1}$$

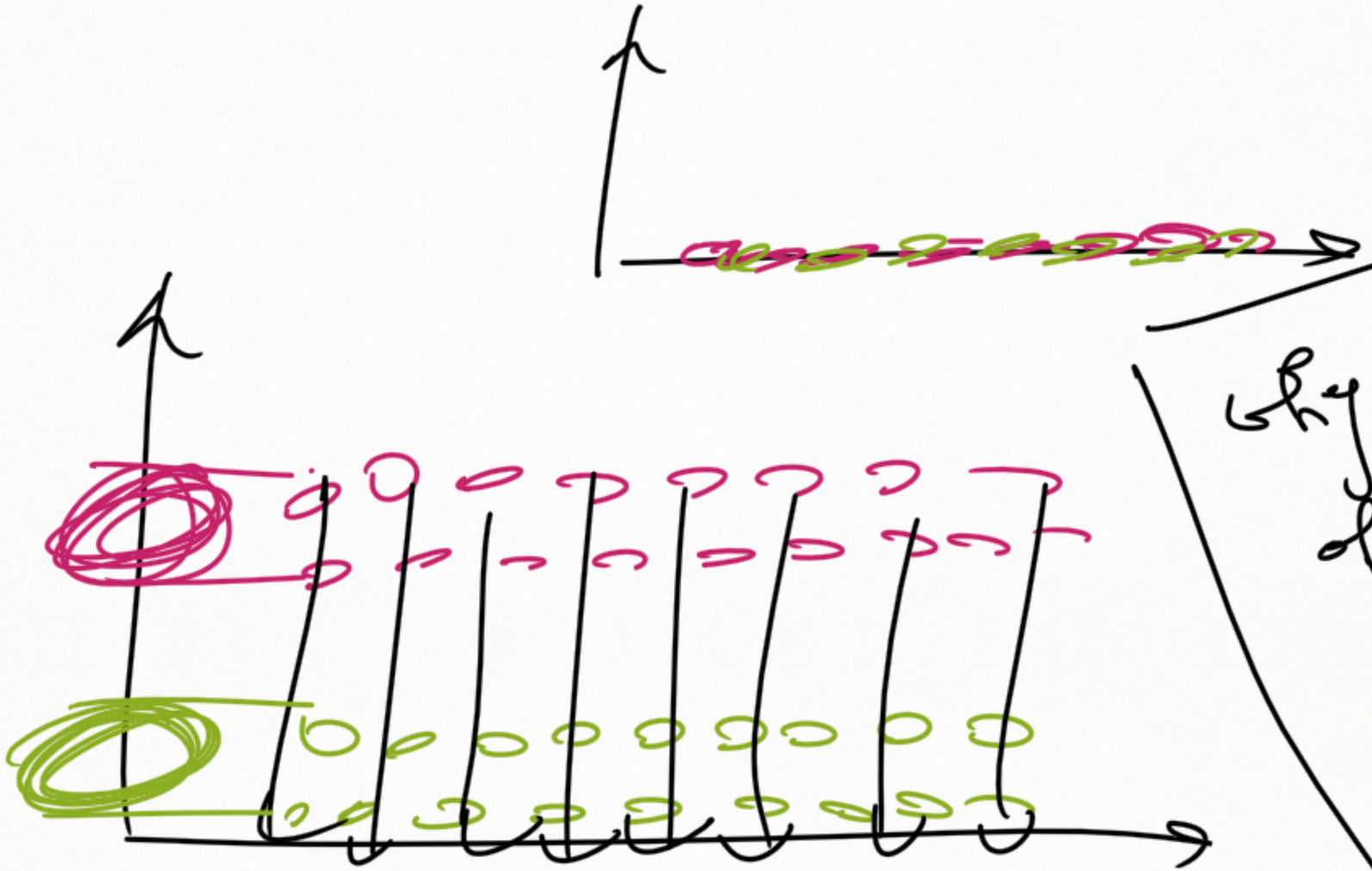
$$\Sigma_0^{RDA} = \alpha \cdot \Sigma_p + (1-\alpha) \cdot \hat{\Sigma}_0$$

LDA

QDA

$$\Sigma_0^{RDA} = \underbrace{\alpha}_{1} \cdot \Sigma_p + \underbrace{(1-\alpha)}_{0} \hat{\Sigma}_0 = \hat{\Sigma}_0$$

Σ_p



Why we should apply PCA after PCA Scaling

Binary Logistic Regression :-

Logistic Regression is a parametric classification algorithm

Logistic Regression is a Discriminative Classifier,

Classification algorithms
- Discriminative Classification algos
- Generative Classification algos

$$\underbrace{P\left(C=0 \,\middle|\, X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}$$ $\longrightarrow$ In generative classifiers, we can't compute this probability directly. First, we have to compute $\boxed{P\left(X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \,\middle|\, C=0\right)}$ from

Bayes Classifier is a perfect example

Only we can compute $P\left(C=0 \,\middle|\, X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)$

In discriminative classifiers we can directly calculate

$$\boxed{P\left(C=0 \,\middle|\, X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}$$

without having to need to evaluate $\cancel{P\left(X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \,\middle|\, C=0\right)}$. For this we will

represent $P\left(C=c_i \,\middle|\, X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)$ as a PDF having <u>some</u> <u>parameters</u> and try to

find their best estimates by running MLE

$$P(X=k) = \boxed{\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(k-\mu)^2}{2\sigma^2}}} = \boxed{P(X=k)}$$

$$P\left(c=1 \,\middle|\, X_i = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \frac{P\left(X_i = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \,\middle|\, c=1\right) P(c=1)}{\sum}$$

Generative Classification

$$\log_e \frac{P\left(C=1 \mid X_i=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{P\left(C=0 \mid X_i=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = \left(\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2\right)$$

Best Estimates

$$\boxed{y = f(X) = f(x_1, x_2) = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 \begin{matrix} > 0 \\ < 0 \end{matrix}}$$

$$y = f(X) = f(x_1, x_2) = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2$$

'4' '2'

Best Estimates

$\hat{\theta}_0$ = Longley answer $\hat{\theta}_1$

Best Estimates

$\hat{\theta}_2$

Best Estimates

Maximum Likelihood Estimation.

$$\boxed{y = f(X) = f(x_1, x_2) = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2}$$

It is not a PDF $> +1$ < 0

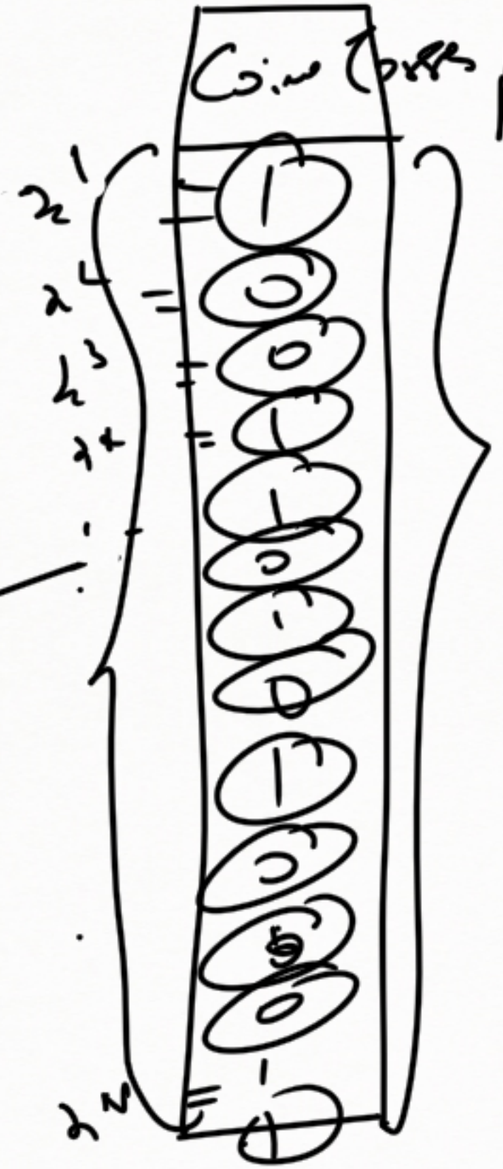
$$\frac{P\left(C=1 \,\middle|\, x_i = \begin{bmatrix} x_1^i \\ \vdots \\ x_L^i \end{bmatrix}\right)}{P\left(C=0 \,\middle|\, x_i = \begin{bmatrix} x_1^i \\ \vdots \\ x_L^i \end{bmatrix}\right)} = e^{\left(\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2\right)}$$

Sigmoid Function

$$\boxed{P\left(\underline{\underline{C=1}} \,\middle|\, \underline{\underline{x_i}} = \begin{bmatrix} x_1^i \\ \vdots \\ x_L^i \end{bmatrix}\right) = \left\{ \frac{1}{1 + e^{-\left(\hat{\theta}_0 + \hat{\theta}_1 \underline{x_1}^i + \hat{\theta}_L \underline{x_L}^i\right)}} \right\}}$$

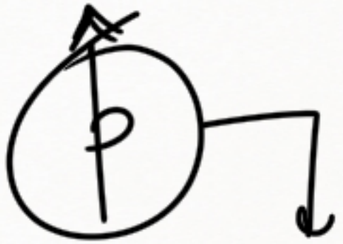
$$P\left(C=0 \,\middle|\, X_i = \begin{bmatrix} x_1^i \\ x_2^i \end{bmatrix}\right) = 1 - P\left(C=1 \,\middle|\, x_i = \begin{bmatrix} x_1^i \\ x_2^i \end{bmatrix}\right)$$

$$= \left[1 - \frac{1}{1+e^{-(\hat{\theta}_0 + \hat{\theta}_1 x_1^i + \hat{\theta}_2 x_2^i)}}\right]$$



N

Binomial Random Variable



In fair coin

Likelihood function

task: best estimate of probability of occurrence of 1

In any single, i^{th} coin toss, I can write -

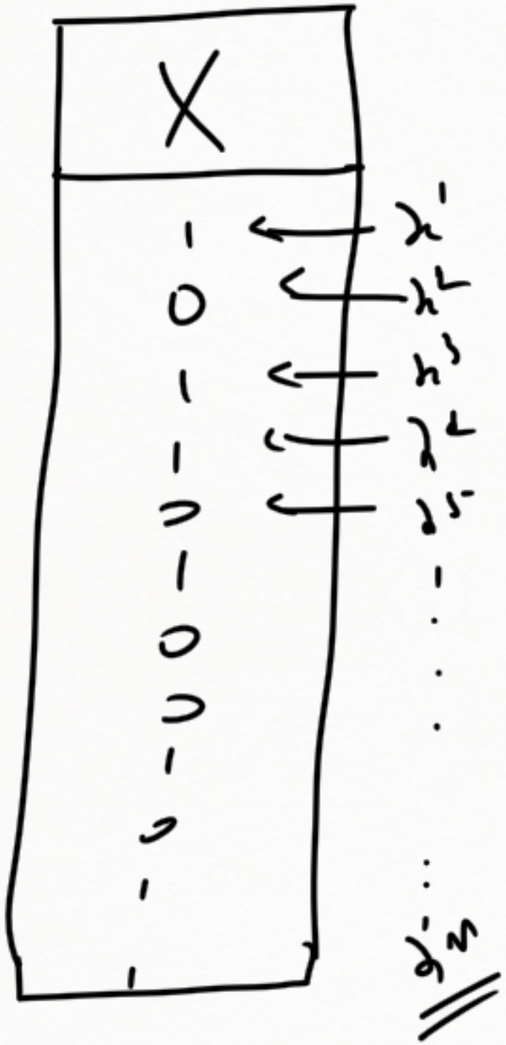
$$\boxed{P(X = x^i) = (\hat{p})^{x^i} \cdot (1-\hat{p})^{(1-x^i)}}$$

$$\{0, 1\}$$

$$P(X = 0) = (\hat{p})^0 (1-\hat{p})^{(1-0)}$$

$$= (1-\hat{p})$$

$$\boxed{P(X = 1) = (\hat{p})^1 \cdot (1-\hat{p})^{(1-1)} = \hat{p}}$$



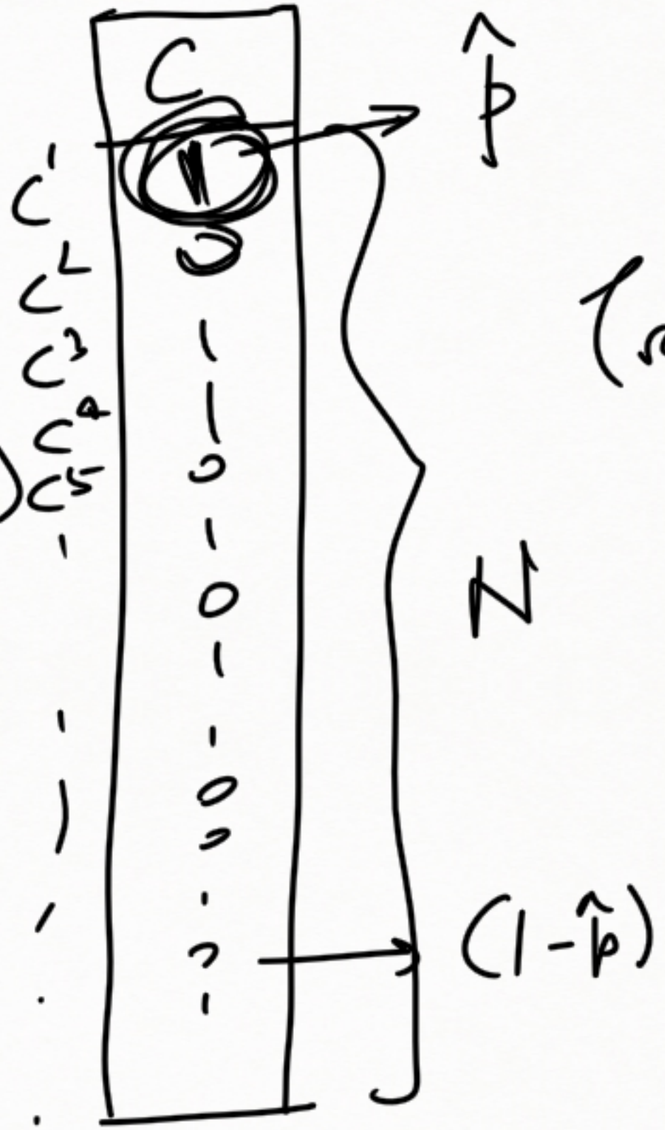
$$\underbrace{P\left(X = x^1 \cap X = x^2 \cap X = x^3 \cap \ldots X = x^N\right)}$$

$$\underbrace{P\left(X = x^1\right)} \cdot \underline{P\left(X = x^2\right)} \cdot P\left(X = x^3\right) \ldots P\left(X = x^N\right)$$

$$\left[\left(\hat{p}\right)^{x^1} \cdot \left(1 - \hat{p}\right)^{(1 - x^1)}\right] \left[\left(\hat{p}\right)^{x^2} \cdot \left(1 - \hat{p}\right)^{(1 - x^2)}\right]$$

$$\underbrace{P(X=x^1 \cap X=x^2 \cap \ldots X=x^N)}_{L(\hat{p})} = \underbrace{\prod_{i=1}^{N} (\hat{p})^{x^i} (1-\hat{p})^{(1-x^i)}}_{\text{Likelihood function}}$$

$$L(\hat{p}) = \prod_{i=1}^{N} (\hat{p}_i)^{c^i} (1-\hat{p}_i)^{(1-c^i)}$$

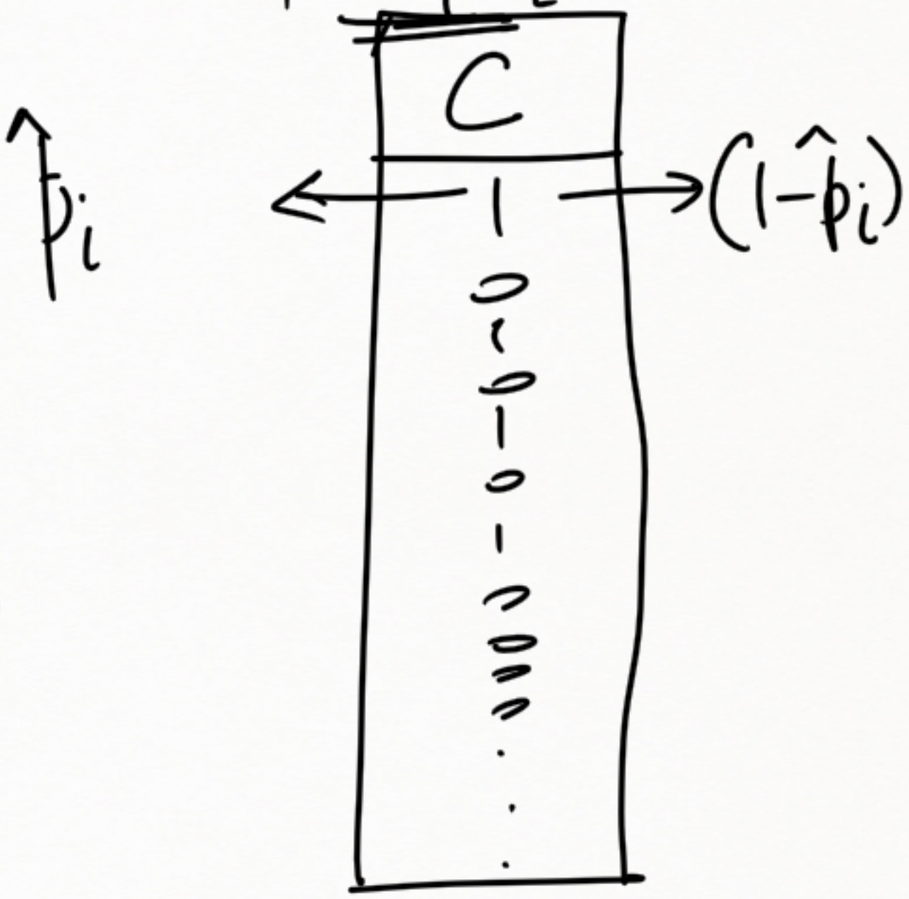


Training Data

$$\hat{p}_i = \frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}_1 x_1^i+\hat{\theta}_L x_L^i)}} = P\left(c^i=1 \,\middle|\, X_i = \begin{bmatrix} x_1^i \\ \vdots \\ x_L^i \end{bmatrix}\right)$$

$$L(\hat{p}_1, \hat{p}_L, \ldots \hat{p}_N) =$$

$$\prod_{i=1}^{N} (\hat{p}_i)^{c^i} (1-\hat{p}_i)^{(1-c^i)}$$



$$L(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_N) = \prod_{i=1}^{N} (\hat{p}_i)^{c^i} (1-\hat{p}_i)^{(1-c^i)}$$

$$= \underbrace{\prod_{i=1}^{N} \left[P\left(c^i = 1 \,\middle|\, X_i = \begin{bmatrix} x_1^i \\ x_2^i \end{bmatrix}\right) \right]^{c^i} \left[1 - P\left(c^i = 1 \,\middle|\, X_i = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) \right]^{(1-c^i)}}$$

Likelihood function for logistic Regression

$$\log_e L(\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2) = \sum_{i=1}^{N} \log_e \left[P\left(c^i = 1 \,\middle|\, x_i = \begin{bmatrix} x_1^i \\ x_2^i \end{bmatrix}\right)^{c^i} \cdot P\left(c^i = 0 \,\middle|\, x_i = \begin{bmatrix} x_1^i \\ x_2^i \end{bmatrix}\right)^{(1-c^i)} \right]$$

$$= \sum_{i=1}^{N} \left[\log_e \left[P(c^i = 1 \mid x_i) \right]^{c^i} + \log_e \left[1 - P(c^i = 1 \mid x_i) \right]^{(1-c^i)} \right]$$

$$\sum_{i=1}^{N}\left[c^i \log_e \underbrace{P(c^i=1 \mid x_i)} + (1-c^i)\log_e \underbrace{\left[1-P(c^i=1 \mid x_i)\right]}\right]$$

$$= \underbrace{\sum_{i=1}^{N}\left[c^i \log_e \left[\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}\right] + (1-c^i)\log_e \left[1-\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}\right]\right]}_{= -\log_e L(\hat{\underline{\theta}}_0, \hat{\underline{\theta}}_1, \hat{\underline{\theta}}_2)}$$

So, now our optimization problem becomes —

$$\max_{\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2} \log_e L(\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2)$$

→ CONCAVE FUNCTION

$$\min_{\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2} -\log_e L(\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2)$$

→ CONVEX FUNCTION

→ GRADIENT DESCENT ALGORITHM

PSEUDOCODE FOR GRADIENT DESCENT ALGORITHM :-

$\hat{\theta}_{0\,initial}$, $\hat{\theta}_{initial} = \begin{bmatrix} \hat{\theta}_{1\,initial} \\ \hat{\theta}_{2\,initial} \end{bmatrix}$, $\epsilon = 10^{-1}$ to 10^{-5}, $tol = 10^{-5}$

while(True):

$$\hat{\theta}_{0\,final} = \hat{\theta}_{0\,initial} + \epsilon \cdot \frac{\partial}{\partial \hat{\theta}_0}\left[\log_e L(\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2)\right]$$

$$\begin{bmatrix} \hat{\theta}_{1\,final} \\ \hat{\theta}_{2\,final} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_{1\,initial} \\ \hat{\theta}_{2\,initial} \end{bmatrix} + \epsilon \cdot \nabla \log_e L(\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2)$$

$\hat{\theta}_{final} \leftarrow$

$\hat{\theta}_0 = \hat{\theta}_{0\,initial}$
$\hat{\theta}_1 = \hat{\theta}_{1\,initial}$
$\hat{\theta}_2 = \hat{\theta}_{2\,initial}$

$\hat{\theta}_0 = \hat{\theta}_{0\,final}$
$\hat{\theta}_1 = \hat{\theta}_{1\,final}$
$\hat{\theta}_2 = \hat{\theta}_{2\,final}$

if $\left| -\log_c L(\hat{\theta}_{0\,final}, \hat{\theta}_{1\,final}, \hat{\theta}_{2\,final}) + \log_c L(\hat{\theta}_{0\,initial}, \hat{\theta}_{1\,initial}, \hat{\theta}_{2\,initial}) \right| < tol$:

<u>break</u>

$\hat{\theta}_{0\,initial} = \boxed{\hat{\theta}_{0\,final}} \rightarrow \hat{\theta}_0^*$

$\hat{\theta}_{initial} = \boxed{\hat{\theta}_{final}} \rightarrow \hat{\theta}^*$

GRADIENT DESCENT ALGO for BLR

$\boxed{P(X = x^i)}$ $\longrightarrow$ This is the probability of occurrence of $\{1, 0\}$ in only i^{th} toss

Result of ith toss $= \{1, 0\}$

Let $\hat{p}$ be the probability of occurrence of 1 in a single toss

$(1-\hat{p})$ be the probability of occurrence of 0 in a single toss

$$P(X = x_i) = (\hat{p})^{x^i} \cdot (1-\hat{p})^{(1-x^i)}$$

$$\downarrow$$

$$1 \quad = (\hat{p})^{1} \cdot (1-\hat{p})^{0}$$

$$= \hat{p}$$

$$P(X = x^i) = (\hat{p})^{0} \cdot (1-\hat{p})^{(1-0)}$$

$$\downarrow$$

$$0 \quad = \underline{\underline{(1-\hat{p})}}$$

$$\longrightarrow P(X = x_i) = (\hat{p})^{x^i} \cdot \underline{\underline{(1-\hat{p})^{(1-x^i)}}}$$

1st toss $\rightarrow x^1 \rightarrow \{0,1\}$

2nd toss $\rightarrow x^2 \rightarrow \{0,1\}$

$\vdots$

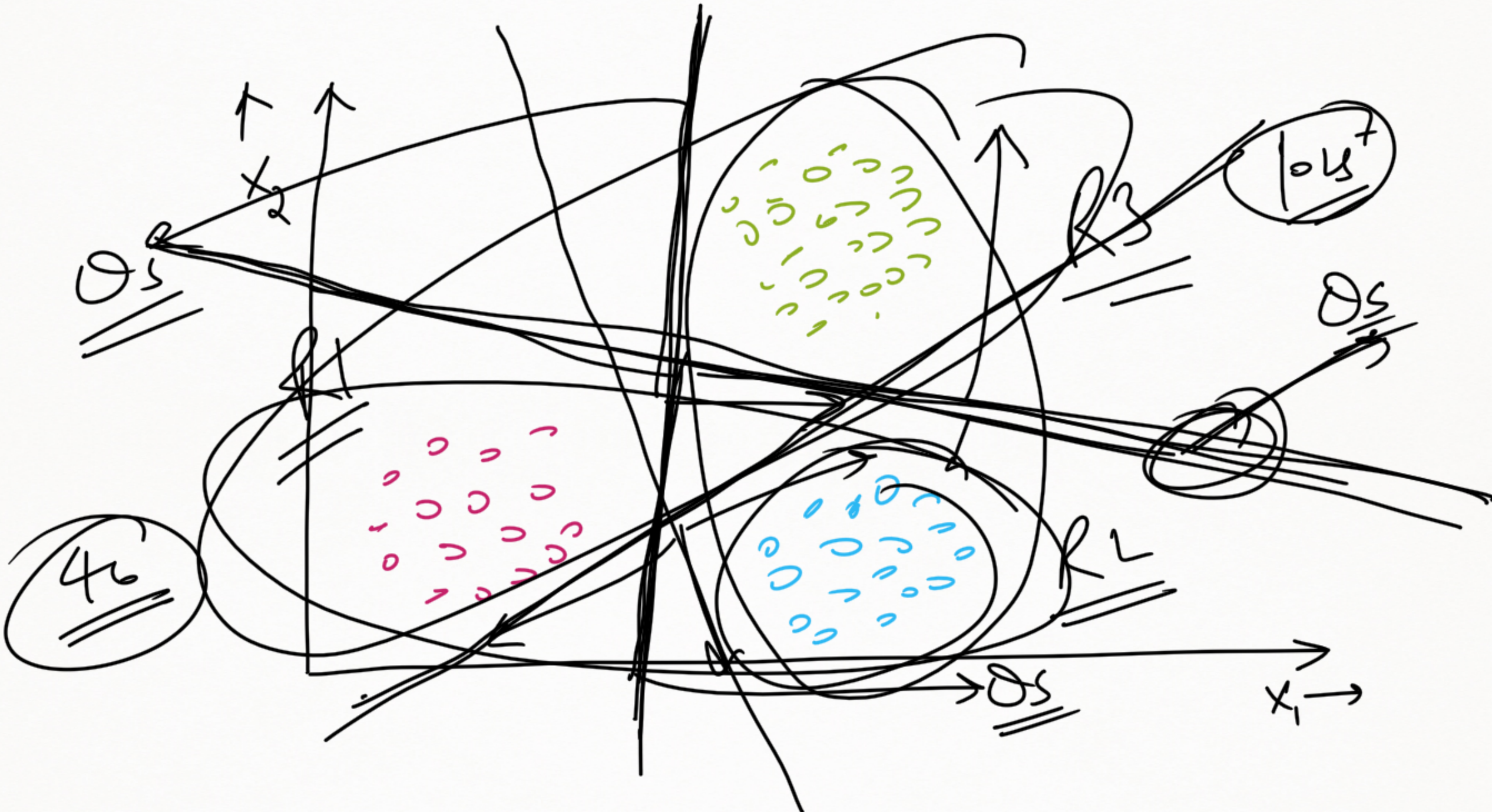
$x^N \rightarrow \{0,1\}$

$$\prod_{i=1}^{N} (\hat{p}_i)^{c^i} (1-\hat{p}_i)^{(1-c^i)}$$

$$\hat{p}_i \longrightarrow P(C=1 \mid \underline{x_i}) = \frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}_1 x_1^i+\hat{\theta}_2 x_2^i)}}$$

$$P(X=x^1 \cap X=x^1 \cap \ldots X=x^N) = \underbrace{P(X=x^1)}\cdot P(X=x^2) \ldots P(X=x^N)$$

$$\prod_{i=1}^{N}\left[(\hat{p})^{x^i}(1-\hat{p})^{(1-x^i)}\right] \Leftarrow \left[(\hat{p})^{x^1}\cdot(1-\hat{p})^{(1-x^1)}\right] \ldots \left[(\hat{p})^{x^N}\cdot(1-\hat{p})^{(1-x^N)}\right]$$


x_2
x_1
R_1
R_2
R_3
θ_s
θ_s
$\theta_s

$$\boxed{P(c^i = 1 \mid \underline{x_i}) = \frac{1}{1 + e^{-(\hat{\theta}_0 + \hat{\theta}^T x_i)}}} \longrightarrow \text{Sigmoid}$$

$$\underline{h \times 1} \qquad \hat{\theta} = \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_n \end{bmatrix} \; h \times 1 \qquad \underline{(n+1)}$$

$$\boxed{\text{Softmax function}}$$

X_i

$$P(C^i=1 \mid X_i) = \frac{1}{1+e^{-(\hat{\theta}_0^* + \hat{\theta}^{*T} \cdot X_i)}}$$

$> 0.5 \longrightarrow$ Tumor belongs 1 class

$< 0.5 \longrightarrow$ Tumor belongs to 0 class

$e^{+0.5}$ $\quad -0.5$

If $\boxed{f(X_i) = \hat{\theta}_0^* + \hat{\theta}^{*T} X_i}$

$> 0 \longrightarrow$ Tumor belongs to 1 class

$= 0$

$< 0 \longrightarrow$ Tumor belongs to 0 class $\rightarrow \boxed{-0.5}$

$e = 2.__ = 1.__$ $\quad 2.$

$$-\log_e L(\hat{\theta}_0, \hat{\theta}) = -\sum_{i=1}^{N} c^i \log_e \left[\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}\right] + (1-c^i)\log_e\left[1-\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}\right]$$

$$\hat{\theta}^T x_i = (\hat{\theta}_1 x_1^i + \hat{\theta}_2 x_2^i)$$

$$\begin{bmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{bmatrix}$$

$$f_i(\hat{\theta}_0, \hat{\theta}) = \underbrace{\boxed{c^i \log_e \left[\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}\right]}}_{1} + \underbrace{\boxed{(1-c^i)\log_e\left[1-\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}\right]}}_{2}$$

① $\frac{\partial}{\partial \hat{\theta}_1} c^i \log\left[\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}_1 x_1^i+\hat{\theta}_2 x_2^i)}}\right] \Rightarrow \frac{\partial}{\partial \hat{\theta}_2}$

$$c^i \cancel{\left(1+e^{-(\hat{\theta}_0+\hat{\theta}_1 x_1^i+\hat{\theta}_2 x_2^i)}\right)} \cdot \frac{1}{\left(1+e^{-(\hat{\theta}_0+\hat{\theta}_1 x_1^i+\hat{\theta}_2 x_2^i)}\right)^{\cancel{2}}} \cdot e^{-(\hat{\theta}_0+\hat{\theta}_1 x_1^i+\hat{\theta}_2 x_2^i)} \cdot x_1^i$$

$$\frac{c^i \cdot x_1^i \cdot e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}} = \boxed{c^i \cdot x_1^i \cdot \left[1-\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}\right]}$$

$$\frac{\partial}{\partial \hat{\theta}_1}\left[c^i \log_e\left\{\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}x_i)}}\right\}\right] = c^i x_1^i \left[1-\underbrace{\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}}_{g_i(\hat{\theta}_0,\hat{\theta})}\right]$$

$$\frac{\partial}{\partial \hat{\theta}_2} = c^i x_2^i \left[1-\underbrace{\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}}_{\rightarrow g_i(\hat{\theta}_0,\hat{\theta})}\right]$$

$$\frac{\partial}{\partial \hat{\theta}_0} = c^i \left[1-\underbrace{\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T \underline{x_i})}}}_{\rightarrow g_i(\hat{\theta}_0,\hat{\theta})}\right]$$

② $\frac{\partial}{\partial \hat{\theta}_1} (1-c^i) \log_e \left[1 - \frac{1}{1+e^{-(\hat{\theta}_0 + \hat{\theta}^T x_i)}} \right] \Rightarrow$

$$\frac{-(1-c^i)}{\left[1 - \frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}} \right]} \cdot \boxed{\frac{1}{\left(1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}\right)^2} e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}} \cdot x_1^i$$

$$\frac{e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}{\left(1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}\right)} \cdot \frac{1}{\left(1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}\right)} \cdot x_1^i$$

$$= \frac{-(1-c^i)\cancel{\left[1-\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}\right]}\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}\cdot x_1^i}{\cancel{\left[1-\frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}\right]}}$$

$$= \frac{-(1-c^i)\cdot x_1^i}{1+e^{-(\hat{\theta}_0+\hat{\theta}^T x_i)}}$$

$$\frac{\partial}{\partial \hat{\theta}_0} = \boxed{-(1-c^i)\cdot g_i(\hat{\theta}_0, \hat{\theta})}$$

$$\frac{\partial}{\partial \hat{\theta}_1} = -(1-c^i)\cdot x_1^i \cdot g_i(\hat{\theta}_0, \hat{\theta})$$

$$\frac{\partial}{\partial \hat{\theta}_2} = -(1-c^i)\, x_2^i \cdot g_i(\hat{\theta}_0, \hat{\theta})$$

$$\frac{\partial f_i(\hat{\theta}_0, \hat{\theta})}{\partial \hat{\theta}_0} = \underbrace{\frac{\partial}{\partial \hat{\theta}_1}\left[c^i \log_e g_i(\hat{\theta}_0, \hat{\theta})\right]} + \underbrace{\frac{\partial}{\partial \hat{\theta}_0}\left[(1-c^i) \log_e (1 - g_i(\hat{\theta}_0, \hat{\theta}))\right]}$$

$$= c^i\left(1 - g_i(\hat{\theta}_0, \hat{\theta})\right) - (1-c^i) \cdot g_i(\hat{\theta}_0, \hat{\theta})$$

$$= c^i - \cancel{c^i \cdot g_i(\hat{\theta}_0, \hat{\theta})} - g_i(\hat{\theta}_0, \hat{\theta}) + \cancel{c^i g_i(\hat{\theta}_0, \hat{\theta})} = \left[c^i - g_i(\hat{\theta}_0, \hat{\theta})\right] \cdot 1$$

$$\frac{\partial f_i}{\partial \hat{\theta}_1} = \left[c^i - g_i(\hat{\theta}_0, \hat{\theta})\right] \cdot x_1^i \quad , \quad \frac{\partial f}{\partial \hat{\theta}_2} = \left[c^i - g_i(\hat{\theta}_0, \hat{\theta}_1)\right] \cdot x_2^i$$

$$\frac{\partial}{\partial \hat{\theta}_0}\left[-\log_c L(\hat{\theta}_0, \hat{\theta})\right] = -\sum_{i=1}^{N} \frac{\partial f_i(\hat{\theta}_0, \hat{\theta})}{\partial \hat{\theta}_0}$$

$$= -\sum_{i=1}^{N}\left[c^i - g_i(\hat{\theta}_0, \hat{\theta})\right]$$

$$\frac{\partial}{\partial \hat{\theta}_1}\left[-\log_e L(\hat{\theta}_0, \hat{\theta})\right] = -\sum_{i=1}^{N}\left[c^i - g_i(\hat{\theta}_0, \hat{\theta})\right] \cdot x_1^i \longrightarrow \underline{\underline{\text{scalar}}}$$

$$\frac{\partial}{\partial \hat{\theta}_2}\left[-\log_e L(\hat{\theta}_0, \hat{\theta})\right] = -\sum_{i=1}^{N}\left[c_i - \underbrace{g_i(\hat{\theta}_0, \hat{\theta})}_{\frac{1}{1+e^{-(\hat{\theta}_0 + \hat{\theta}^T x_i)}}}\right] \cdot x_2^i \longrightarrow \underline{\underline{\text{scalar}}}$$

$$\nabla\left[-\log L(\hat{\theta}_0, \hat{\theta})\right] = \begin{bmatrix} \frac{\partial}{\partial \hat{\theta}_1}\left[-\log L(\hat{\theta}_0, \hat{\theta})\right] \\ \\ \frac{\partial}{\partial \hat{\theta}_2}\left[-\log L(\hat{\theta}_0, \hat{\theta})\right] \end{bmatrix}$$

$$\nabla\left[-\log_c L(\hat{\theta}_0, \hat{\theta})\right] = \begin{bmatrix} -\sum_{i=1}^{N}\left[c^i - g_i(\hat{\theta}_0, \hat{\theta})\right] \cdot x_1^i \\ \\ -\sum_{i=1}^{N}\left[c^i - g_i(\hat{\theta}_0, \hat{\theta})\right] \cdot x_2^i \end{bmatrix} \quad 2 \times 1$$

vector