Basic Example of Tumor Classification :-

Lets imagine that the dataset has only one feature and another column for label.

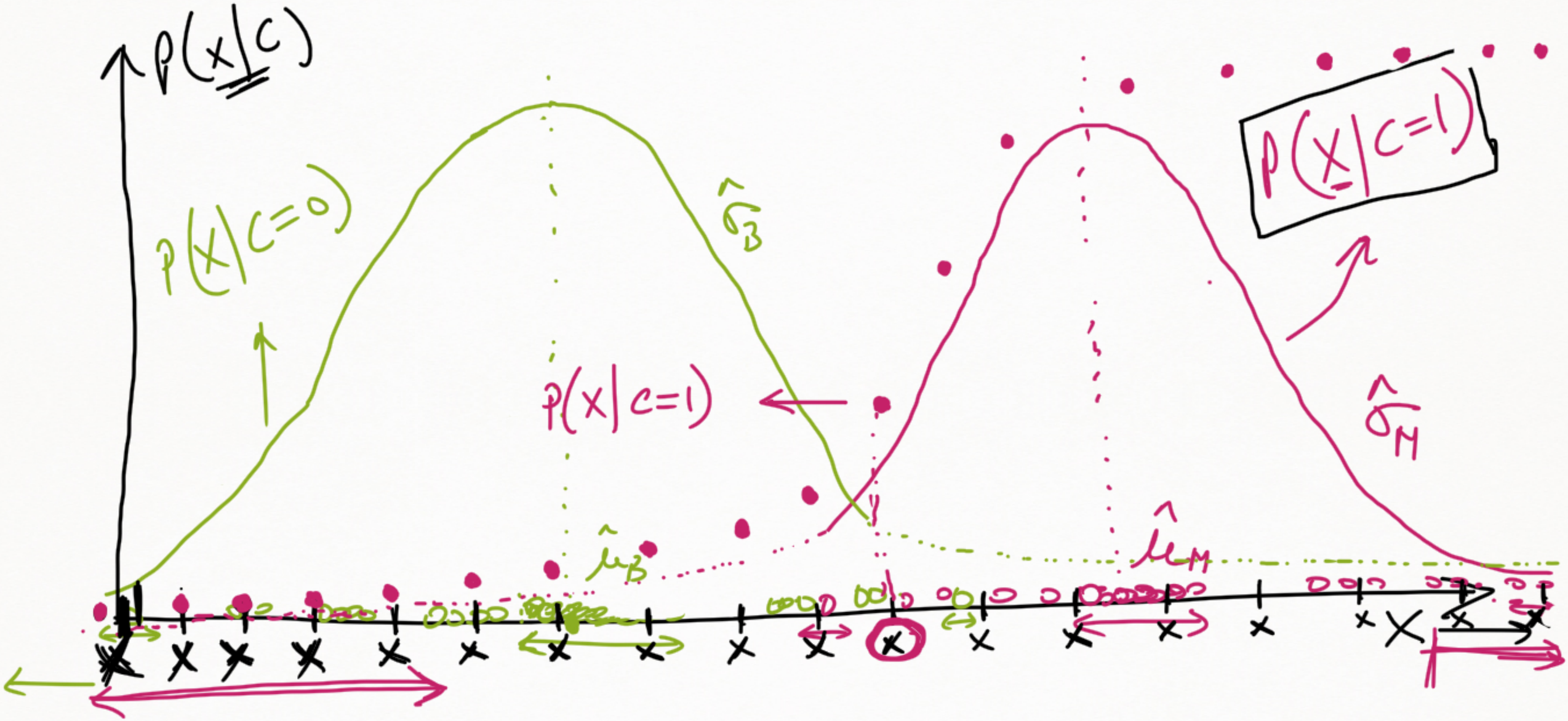
2 columns

N rows

X	C
X_1	M
X_2	B
X_3	M
X_4	M
X_5	B
X_6	B
X_7	M
⋮	⋮
X_N	B

M → Malignant

B → Benign


p(x|c)
p(x|c=0)
p(x|c=1)
p(x|c=1)
$\hat{\sigma}_B$
$\hat{\sigma}_M$
$\hat{\mu}_B$
$\hat{\mu}_M$

$$P(C=1|x) = \frac{P(x|C=1)\cdot \overbrace{\cancel{P(C=1)}}^{0.5}}{P(x|C=0)\cdot \underbrace{\cancel{P(C=0)}}_{0.5} + P(x|C=1)\cdot \underbrace{\cancel{P(C=1)}}_{0.5}}$$

$$P(C=0|x) = \frac{\boxed{P(x|C=0)}}{\underbrace{P(x|C=0) + P(x|C=1)}}$$

$$\boxed{P(C=1|x) = \frac{\boxed{P(x|C=1)}}{P(x|C=0) + P(x|C=1)}}$$

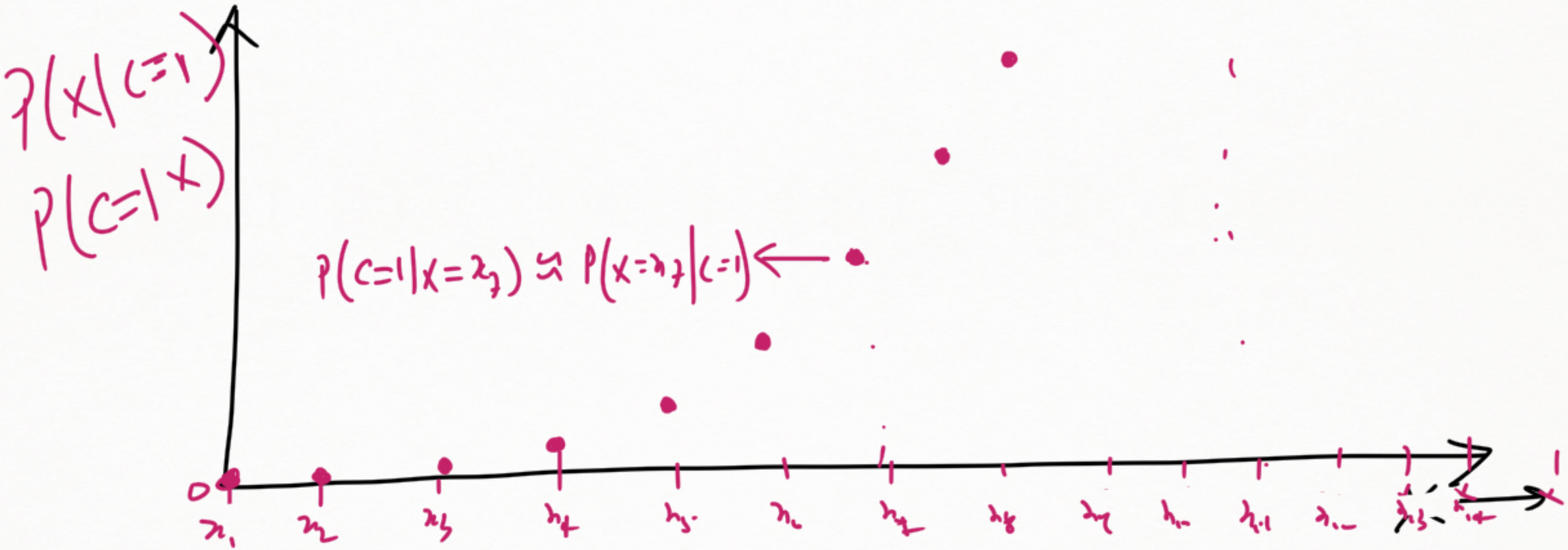
$\boxed{0.57}$

$$\frac{\approx 0}{\boxed{0.57}} \approx 0$$

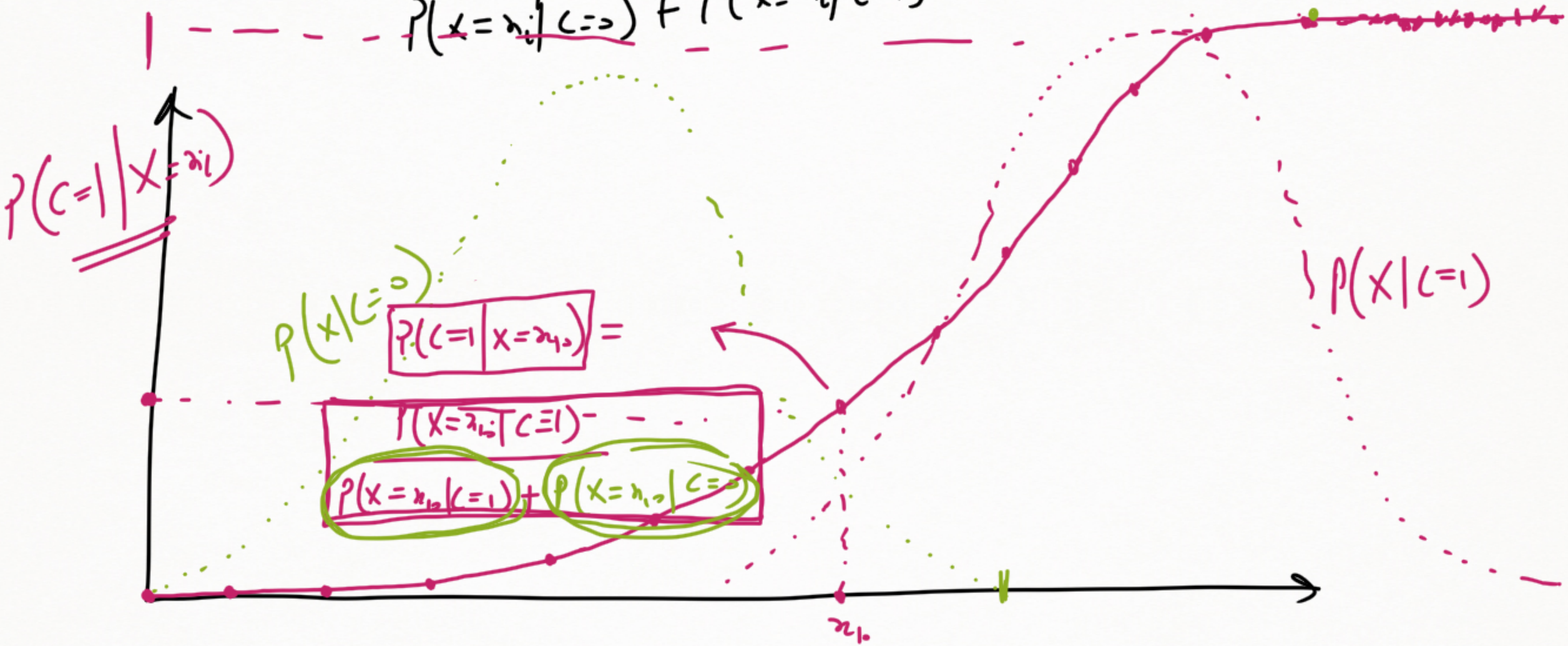
Denominator which remains constant cut for both $\underbrace{P(C=1|x)}$ and $\underbrace{P(C=0|x)}$

$$\boxed{P(c=1|x)} \propto \boxed{P(x|c=1)}$$

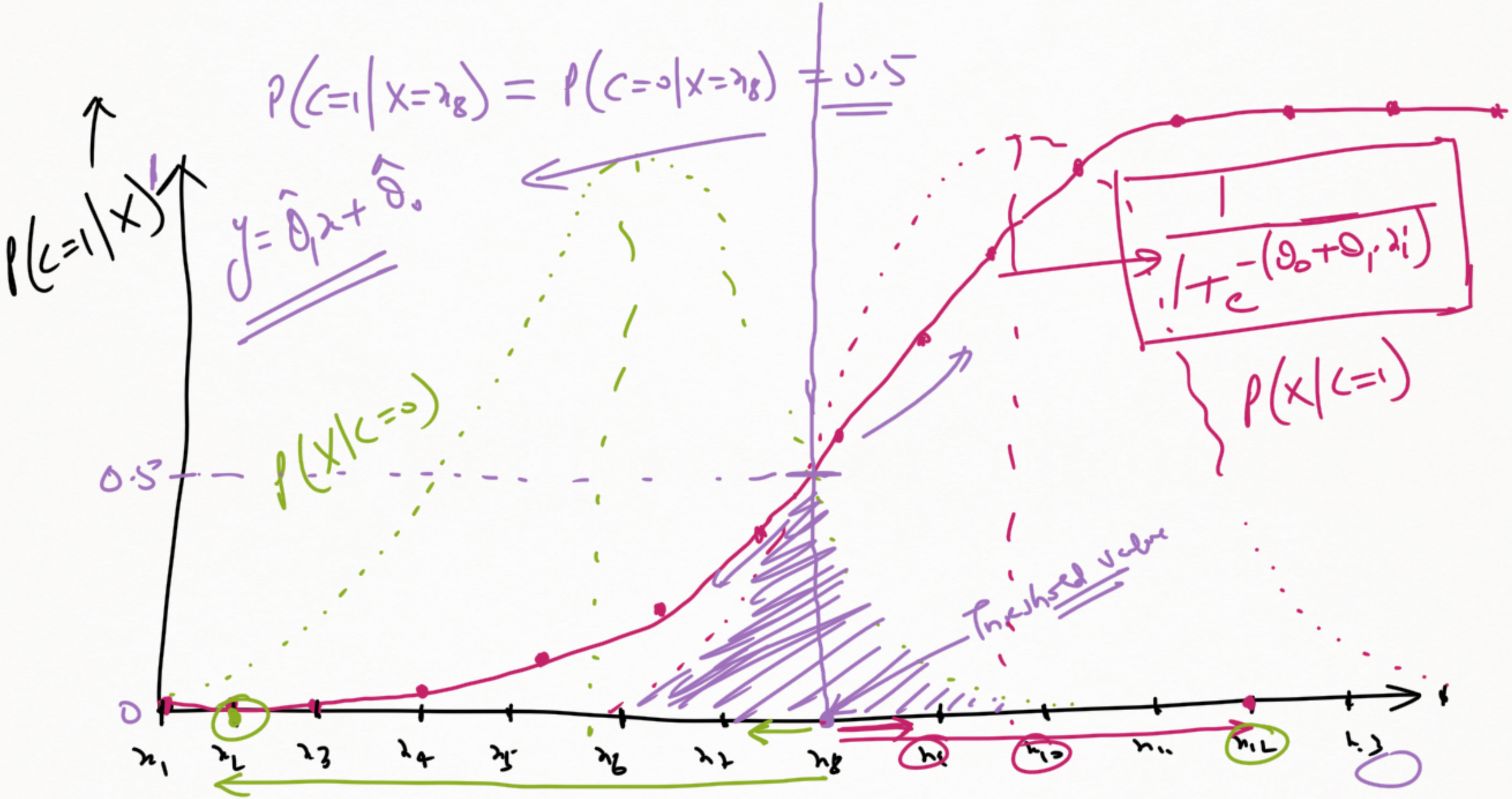
$$P(c=1|x=x_n) \approx P(x=x_n|c=1)$$



$$P(c=1 \mid x=x_i) = \frac{P(x=x_i \mid c=1)}{P(x=x_i \mid c=0) + P(x=x_i \mid c=1)}$$



$$P(c=1 \mid x=x_{10}) = \frac{P(x=x_{10} \mid c=1)}{P(x=x_{10} \mid c=1) + P(x=x_{10} \mid c=0)}$$


$P(C=1|X=x_8) = P(C=0|X=x_8) = 0.5$
$P(C=1|X)$
$y = \hat{\theta}_1 x + \hat{\theta}_0$
$P(X|C=0)$
$\frac{1}{1+e^{-(\theta_0+\theta_1 x_i)}}$
$P(X|C=1)$
0.5
0
Threshold value
x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10} x_{11} x_{12} x_{13}

$$P(C=1 \mid X=x_i) = \frac{P(X=x_i \mid C=1)}{P(X=x_i \mid C=1) + P(X=x_i \mid C=0)} \longrightarrow (1)$$

$$P(C=0 \mid X=x_i) = \frac{P(X=x_i \mid C=0)}{P(X=x_i \mid C=0) + P(X=x_i \mid C=1)} \longrightarrow (2)$$

$$\frac{2}{1} \Rightarrow \frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)} = \frac{P(X=x_i \mid C=0)}{P(X=x_i \mid C=1)}$$

$$\frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)} = \frac{\frac{1}{\sqrt{2\pi}\,\hat{\sigma}_B} e^{-\frac{1}{2}\frac{(x_i-\hat{\mu}_B)^2}{2\hat{\sigma}_B^2}}}{\frac{1}{\sqrt{2\pi}\,\hat{\sigma}_M} e^{-\frac{1}{2}\frac{(x_i-\hat{\mu}_M)^2}{2\hat{\sigma}_M^2}}}$$

We are taking an assumption that -

$$\hat{\sigma}_B^2 = \hat{\sigma}_M^2 = \sigma_p^2$$

$$\hat{\sigma}_B^2 \simeq \hat{\sigma}_M^2 \simeq \sigma_p^2$$

$$\frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)} = \frac{\cancel{\frac{1}{\sqrt{2\pi}\,\sigma_p}}\, e^{-\frac{(x_i-\hat{\mu}_B)^2}{2\sigma_p^2}}}{\cancel{\frac{1}{\sqrt{2\pi}\,\sigma_p}}\, e^{-\frac{(x_i-\hat{\mu}_M)^2}{2\sigma_p^2}}}$$

$$\frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)} = \frac{e^{-\frac{(x_i-\hat{\mu}_B)^2}{2\sigma_p^2}}}{e^{-\frac{(x_i-\hat{\mu}_M)^2}{2\sigma_p^2}}}$$

$$\frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)} = e^{-\frac{(x_i-\hat{\mu}_B)^2}{2\sigma_p^2} + \frac{(x_i-\hat{\mu}_M)^2}{2\sigma_p^2}}$$

$$= e^{-\frac{(x_i-\hat{\mu}_M)^2}{2\sigma_p^2} - \frac{(x_i-\hat{\mu}_B)^2}{2\sigma_p^2}}$$

$$= e^{\frac{(x_i-\hat{\mu}_M+x_i-\hat{\mu}_B)(\cancel{x_i}-\hat{\mu}_M-\cancel{x_i}+\hat{\mu}_B)}{2\sigma_p^2}} = e^{\frac{(2x_i-\hat{\mu}_M-\hat{\mu}_B)(\hat{\mu}_B-\hat{\mu}_M)}{2\sigma_p^2}}$$

$$= e^{\frac{\left[2x_i - (\hat{\mu}_B + \hat{\mu}_M)\right]\left[(\hat{\mu}_B - \hat{\mu}_M)\right]}{2\sigma_p^2}} \qquad \underline{\underline{\theta_1}} = \frac{(\hat{\mu}_B - \hat{\mu}_M)}{\sigma_p^2}$$

$$\theta_1$$

$$= e^{\frac{2x_i(\hat{\mu}_B - \hat{\mu}_M) - (\hat{\mu}_B^2 - \hat{\mu}_M^2)}{2\sigma_p^2}}$$

$$= e^{\left[\frac{x_i(\hat{\mu}_B - \hat{\mu}_M)}{\sigma_p^2} + \underbrace{-\frac{(\hat{\mu}_B^2 - \hat{\mu}_M^2)}{2\sigma_p^2}}\right]} = e^{\left[x_i \frac{(\hat{\mu}_B - \hat{\mu}_M)}{\sigma_p^2} + \theta_0\right]}$$

$$\underline{\underline{\theta_0}} = -\frac{(\hat{\mu}_B^2 - \hat{\mu}_M^2)}{2\sigma_p^2}$$

$$\frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)} = e^{(\hat{\theta}_0 + \hat{\theta}_1 \cdot x_i)}$$

$$\frac{1 - P(C=1 \mid X=x_i)}{P(C=1 \mid X=x_i)} = e^{(\hat{\theta}_0 + \hat{\theta}_1 x_i)}$$

$$\frac{1}{P(C=1 \mid X=x_i)} - 1 = e^{(\hat{\theta}_0 + \hat{\theta}_1 x_i)}$$

$$\frac{1}{P(C=1 \mid X=x_i)} = 1 + e^{(\hat{\theta}_0 + \hat{\theta}_1 x_i)}$$

$$P(C=1 \mid X=x_i) = \frac{1}{1 + e^{(\hat{\theta}_0 + \hat{\theta}_1 x_i)}}$$

Let $\theta_0 = -\hat{\theta}_0$, $\theta_1 = -\hat{\theta}_1$

$$\theta_0 = -\hat{\theta}_0, \quad \boxed{\hat{\delta}_0 = -\frac{(\hat{\mu}_B^2 - \hat{\mu}_M^2)}{2\sigma_p^2}} \quad \hat{\theta}_1 = \frac{(\hat{\mu}_B - \hat{\mu}_M)}{\sigma_p^2}$$

$$P(C=1 \mid X = x_i) = \frac{1}{1 + e^{-(-\hat{\theta}_0 + -\hat{\theta}_1 x_i)}}$$

$$\theta_1 = -\hat{\theta}_1$$

$$\boxed{\underbrace{P(C=1 \mid X = x_i)}_{} = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_i)}}}$$

Sigmoid function

$$\underbrace{f(x) = \frac{1}{1+e^{-x}}}$$

$$x = (\theta_0 + \theta_1 x_i)$$

$$f(-\infty) = \frac{1}{1+e^{-(-\infty)}} = \frac{1}{1+e^{\infty}}$$

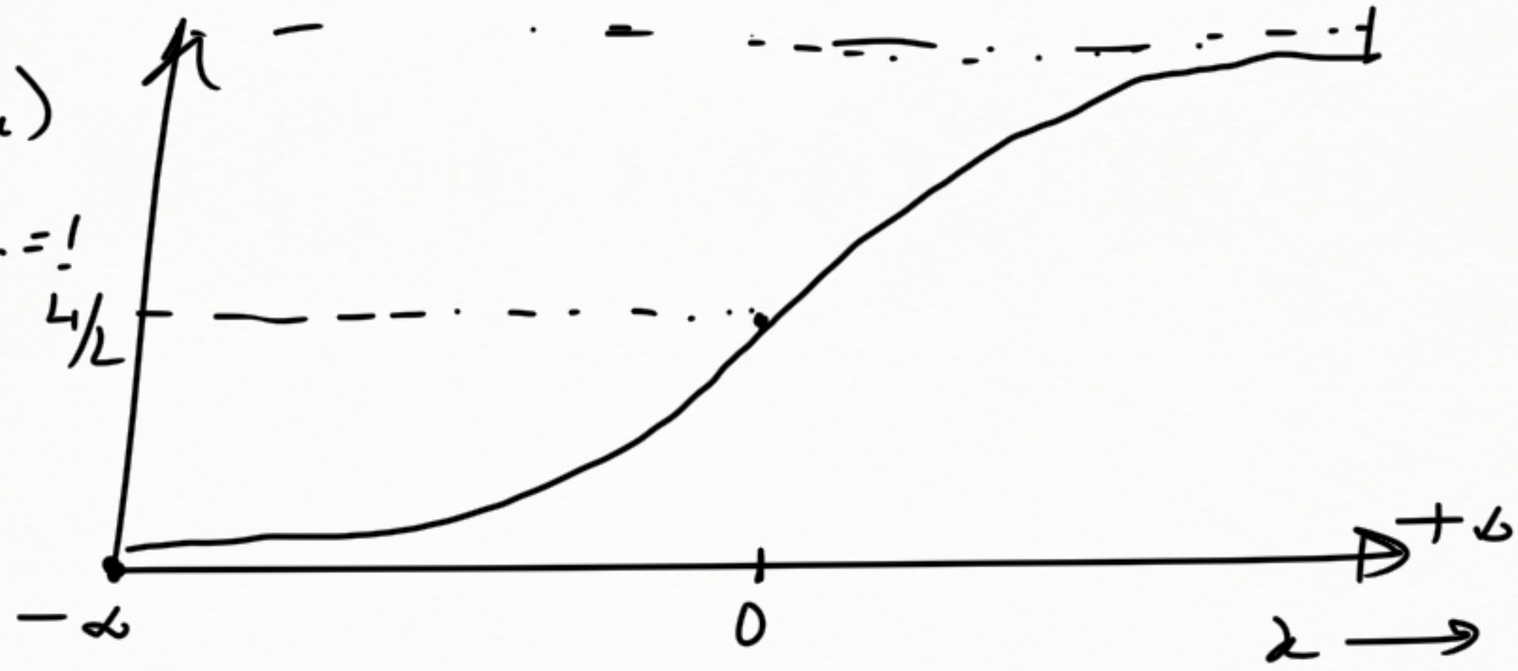
$$\frac{1}{1+\infty} = \frac{1}{\infty} = \underline{\underline{0}}$$

$x=0$

$$f(0) = \frac{1}{1+e^{-0}} = \frac{1}{1+1} = \frac{1}{2}$$

$x=+\infty$

$$f(+\infty) = \frac{1}{1+e^{-(+\infty)}} = \frac{1}{1+e^{-\infty}} = \frac{1}{1+0} = 1$$



$$\log_e \frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)} = \log_e e^{(\hat{\theta}_0 + \hat{\theta}_1 x_i)}$$

$$\log_e 1 = (\hat{\theta}_0 + \hat{\theta}_1 x_8)$$

$$\log_e \frac{\cancel{P(C=0 \mid X=x_8)}}{\cancel{P(C=1 \mid X=x_8)}} = (\hat{\theta}_0 + \hat{\theta}_1 x_8)$$

$$0 = (\hat{\theta}_0 + \hat{\theta}_1 x_8)$$

$$\boxed{x_8 = -\frac{\hat{\theta}_0}{\hat{\theta}_1}} \Rightarrow \boxed{x_{threshold} = -\frac{\hat{\theta}_0}{\hat{\theta}_1}}$$

$$\boxed{x_{threshold} = \frac{\dfrac{\left(\hat{\mu}_B^L - \hat{\mu}_M^L\right)}{2\cancel{\sigma_p^L}}}{\dfrac{\left(\hat{\mu}_B - \hat{\mu}_M\right)}{\cancel{\sigma_p^L}}} = \frac{\hat{\mu}_B + \hat{\mu}_M}{2}}$$

Inflection value

$(0,1)$

$x_i > x_8$

$$\log_e \left[\frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)} \right] = \underbrace{\hat{\theta}_0 + \theta_1 x_i}$$

$$\log_e \frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)}$$

$$\tilde{\theta}_0 + \hat{\theta}_1 x_i < 0$$

$$\boxed{P(C=0 \mid X=x_i) < P(C=1 \mid X=x_i)}$$

If $x > x_8$ then —

$$\hat{\theta}_0 + \hat{\theta}_1 x_i < 0$$

Linear Discriminant function

$\hat{\sigma}_B^2 = .7$

$\hat{\sigma}_W^2 = .8 = \sigma_p^2$

$$\boxed{\hat{\sigma}_B^2 = \hat{\sigma}_{WG}^2 = \sigma_p^2}$$

If $x_i < x_8$ then —

$$\hat{\theta}_0 + \hat{\theta}_1 x_i > 0$$

$$\boxed{\log_2 \frac{P(c=0 \mid x=x_i)}{P(c=1 \mid x=x_i)}}$$

$$F = \frac{\text{Between Group Variance}}{\text{Within Group Variance}}$$

$$\underbrace{\log_e \frac{P(C=0 \mid X=x_i)}{P(C=0 \mid X=x_i)}} = (\hat{\theta}_0 + \hat{\theta}_1 x_i) = \text{Linear Discriminant Function}$$

$$y_i = \log_e \frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)}$$

$$\boxed{y_i = \hat{\theta}_1 x_i + \hat{\theta}_0}$$

$$\boxed{y = \ln x + C}$$

$$\Rightarrow y = f(x) = (\hat{\theta}_1 x + \hat{\theta}_0)$$

$$\boxed{\log_c \frac{P(C=0 \mid X=x)}{P(C=1 \mid X=x)}} > 0$$

QDA

$$\log_e \frac{P(C=0 \mid X=x_i)}{P(C=1 \mid X=x_i)} = \underbrace{\left(\hat{\theta}_0 + \hat{\theta}_1 x_i + \hat{\theta}_2 x_i^2\right)}_{\text{Quadratic Discriminant Function}}$$

$$\frac{P(C=0 \mid X=x)}{P(C=1 \mid X=x)} = \frac{\frac{1}{\sqrt{2\pi}\,\hat{\sigma}_B^2} e^{-\frac{(x-\hat{\mu}_B)^2}{2\sigma_B^2}}}{\frac{1}{\sqrt{2\pi}\,\hat{\sigma}_M^2} e^{-\frac{(x-\hat{\mu}_M)^2}{2\hat{\sigma}_M^2}}} = \hat{\theta}_0 + \hat{\theta}_1 x + \hat{\theta}_2 x^2$$

$$\boxed{\sigma_p^2 = \frac{N_B \cdot \hat{\sigma}_B^2 + N_M \cdot \hat{\sigma}_M^2}{N}} =$$

$$\boxed{N = N_B + N_M}$$

$$\hat{\sigma}_B^2 = \hat{\sigma}_M^2 = \sigma_p^2$$

x_i

Pf for any x_i

$$\hat{\theta}_0 + \hat{\theta}_1 x_i + \hat{\theta}_2 x_i^2 < 0 \longrightarrow i^{th} \text{ tumor} \in C = 1$$

$$> 0 \longrightarrow C = 0$$

$= 0$

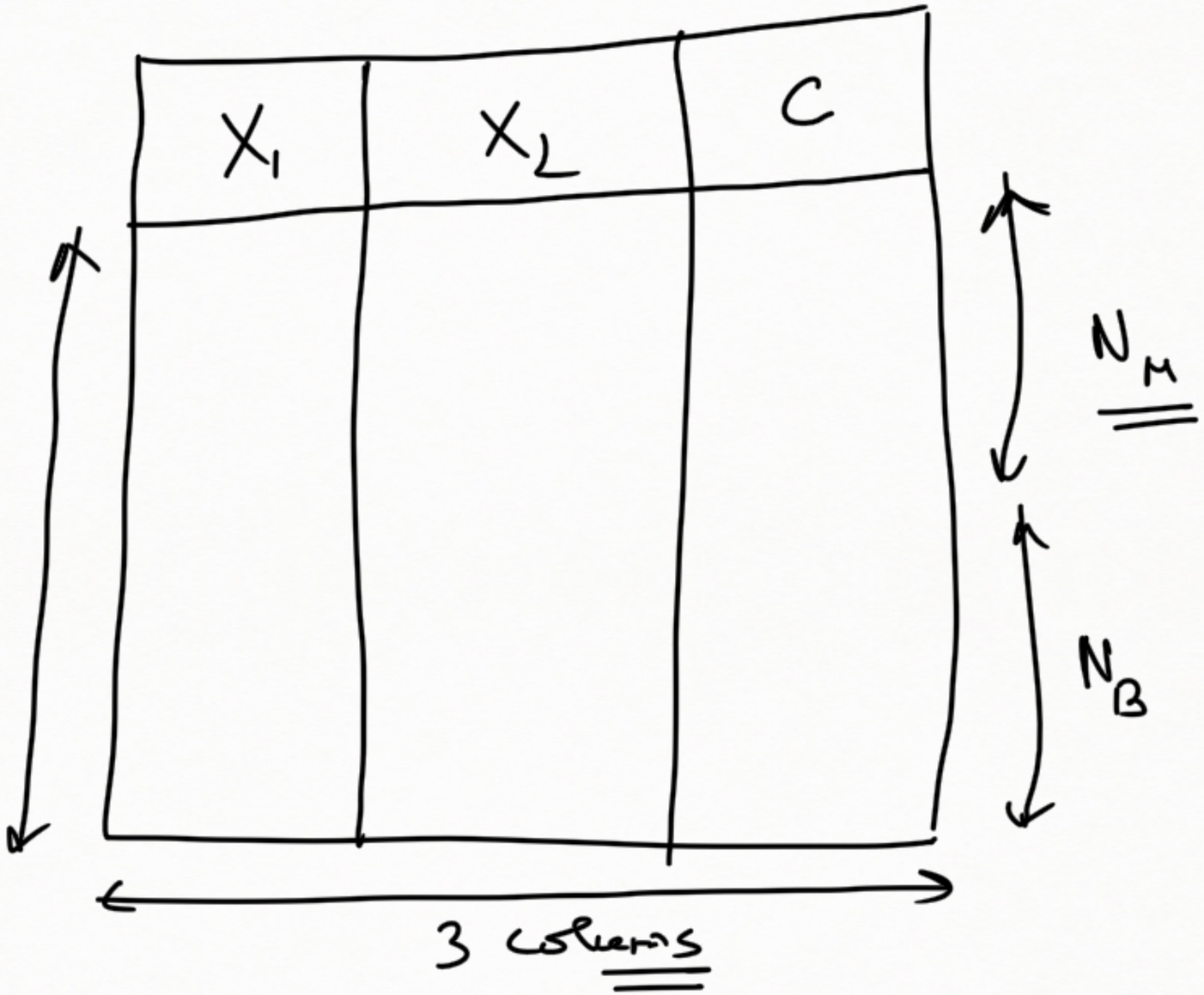
$x_1 \; x_2$

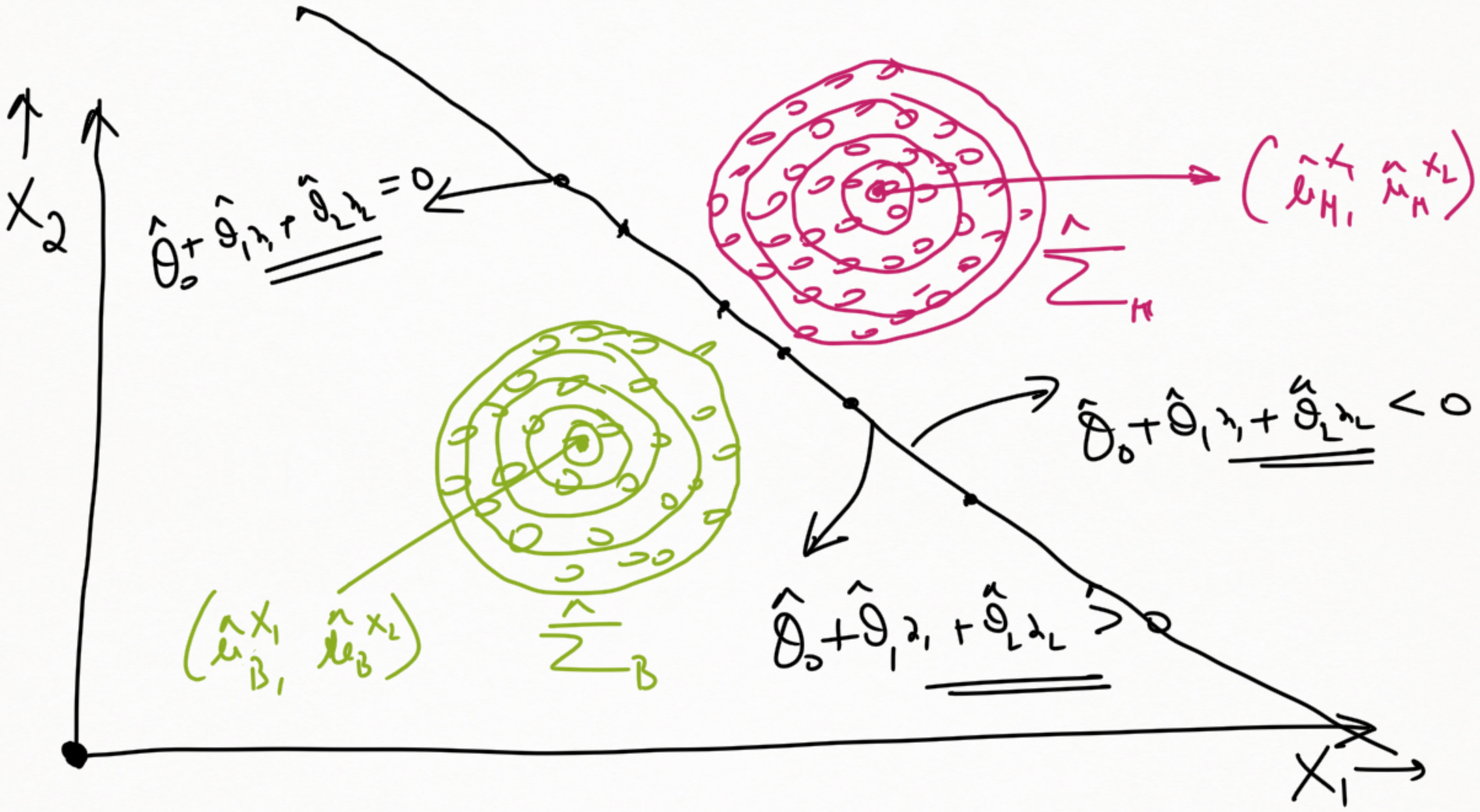
$$\boxed{y = f(x) = \hat{\theta}_0 + \hat{\theta}_1 x + \hat{\theta}_2 \cdot x^2}$$

Training Data

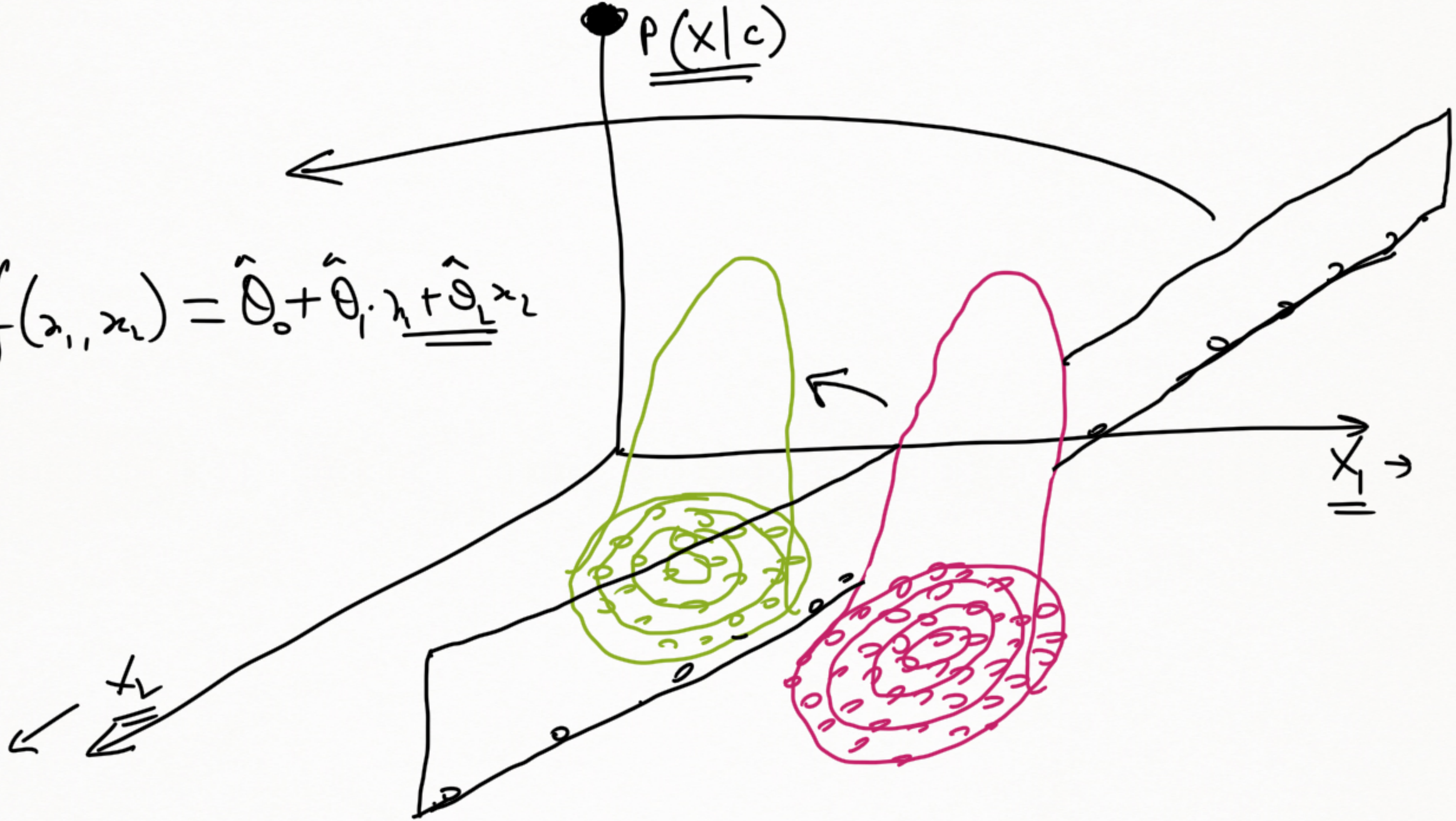
N rows

$N = N_M + N_B$




X_2
X_1
$\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 = 0$
$(\hat{\mu}_{H_1}^{x_1}, \hat{\mu}_H^{x_2})$
$\hat{\Sigma}_H$
$\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 < 0$
$(\hat{\mu}_{B_1}^{x_1}, \hat{\mu}_B^{x_2})$
$\hat{\Sigma}_B$
$\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 > 0$

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



$$\frac{P\left(C=0 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{P\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = \frac{\cancel{\frac{1}{\sqrt{2\pi}\sqrt{|\hat{\Sigma}_B|}}} \, e^{-\frac{1}{2}\begin{bmatrix} x_1 - \hat{\mu}_B^{x_1} \\ x_2 - \hat{\mu}_B^{x_2} \end{bmatrix}^T \underbrace{\hat{\Sigma}_B^{-1}}_{\Sigma_p^{-1}} \begin{bmatrix} x_1 - \hat{\mu}_B^{x_1} \\ x_2 - \hat{\mu}_B^{x_2} \end{bmatrix}}}{\cancel{\frac{1}{\sqrt{2\pi}\sqrt{|\hat{\Sigma}_M|}}} \, e^{-\frac{1}{2}\begin{bmatrix} x_1 - \hat{\mu}_M^{x_1} \\ x_2 - \hat{\mu}_M^{x_2} \end{bmatrix}^T \underbrace{\hat{\Sigma}_M^{-1}}_{\Sigma_p^{-1}} \begin{bmatrix} x_1 - \hat{\mu}_M^{x_1} \\ x_2 - \hat{\mu}_M^{x_2} \end{bmatrix}}}$$

Let $\hat{\Sigma}_B = \hat{\Sigma}_M = \Sigma_p$

$$\boxed{\Sigma_p = \frac{N_M \cdot \hat{\Sigma}_M + N_B \cdot \hat{\Sigma}_B}{N}}$$

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

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

$$\boxed{\log_e \frac{P\left(C=0 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{P\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}} = \underbrace{\left(\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2\right)}_{\tilde{\theta}_\phi \underline{\underline{x}} + \theta_0}$$

$$y = f(x_1, x_2) = \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}^T \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$$

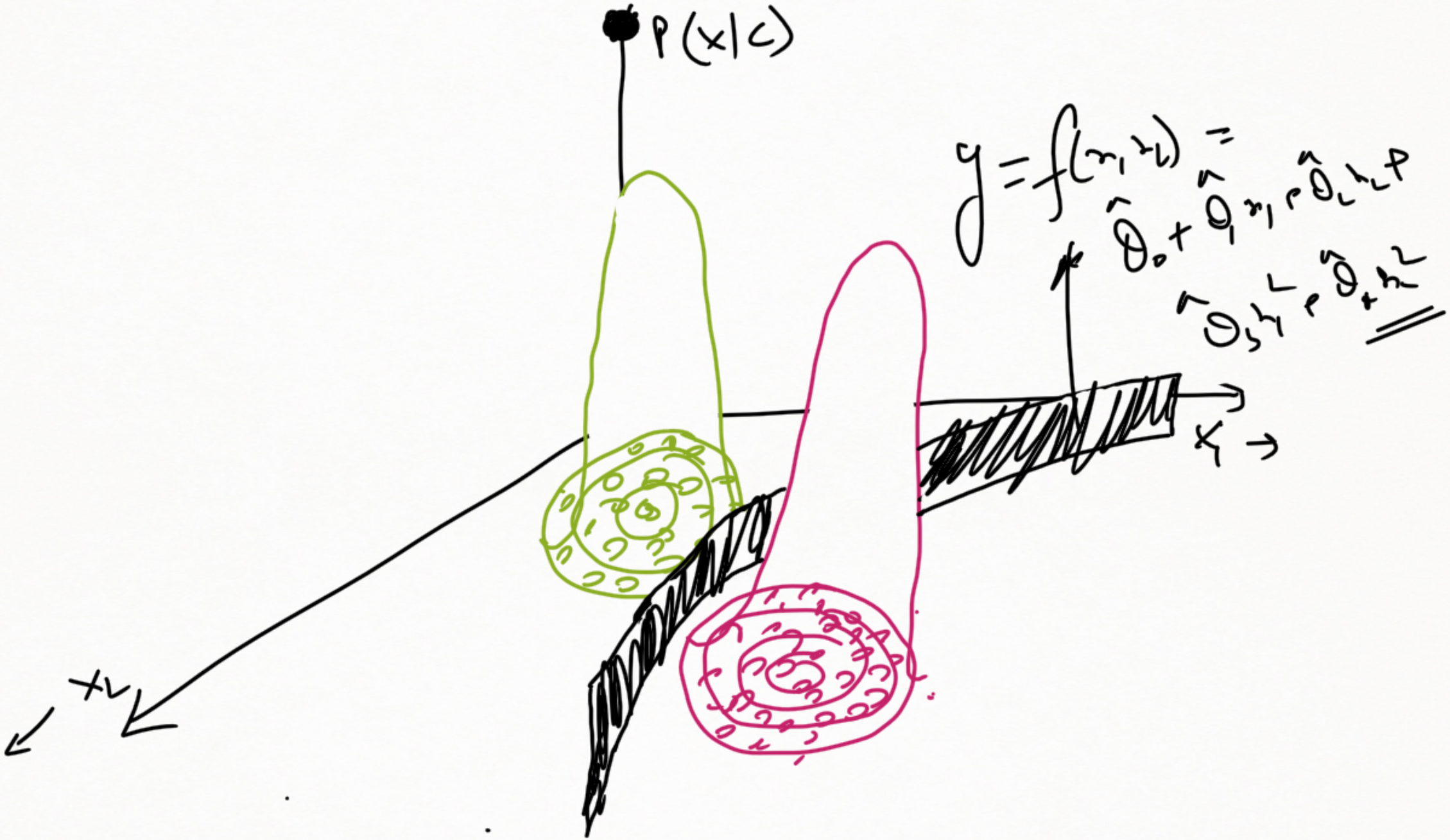
$$= \underbrace{\begin{bmatrix} \hat{\theta}_0 & \hat{\theta}_1 & \hat{\theta}_2 \end{bmatrix}} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \underline{\underline{\hat{\theta}_2}} x_2$$

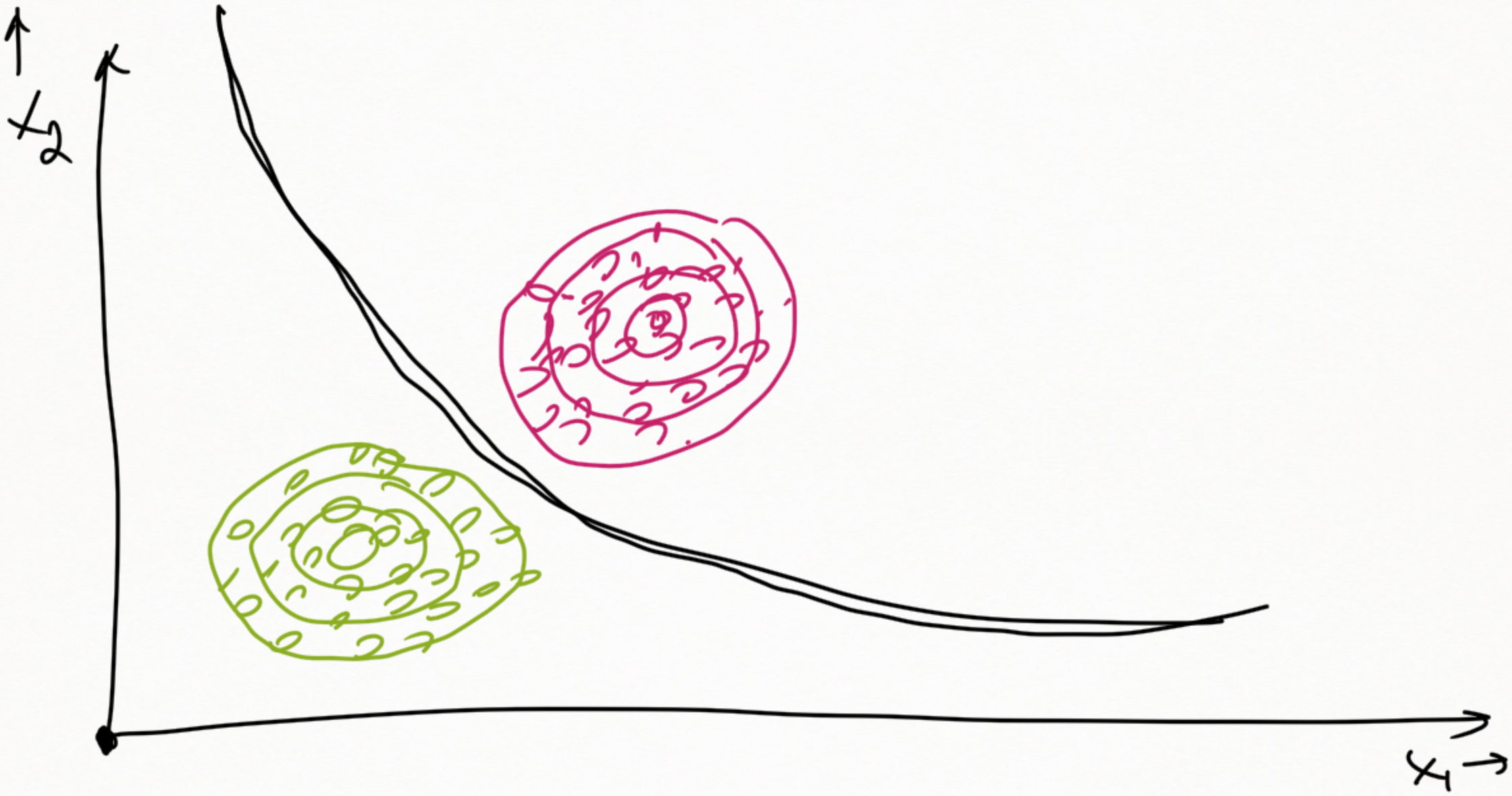
This is a unit vector on which you are projecting your.

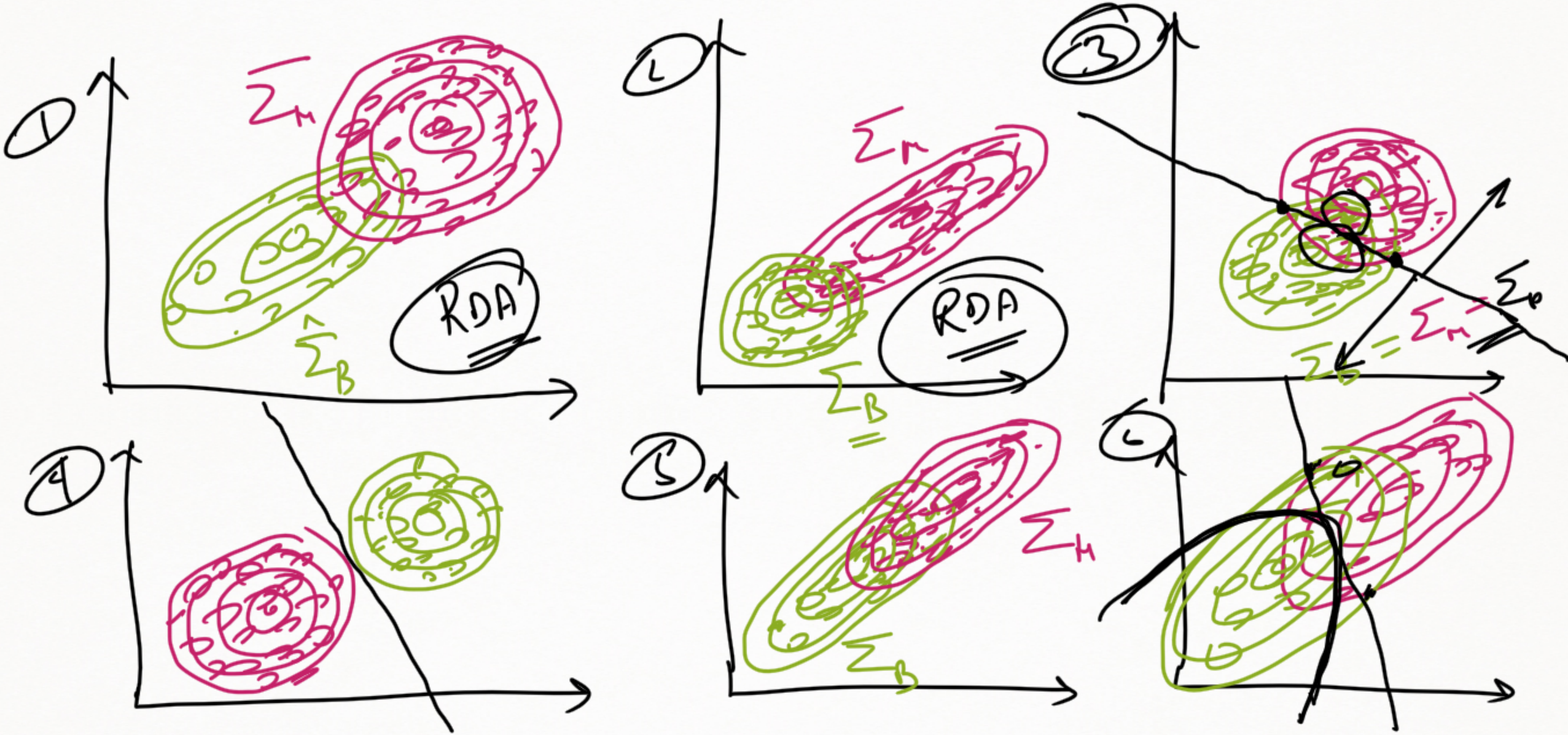
<u>Data</u>

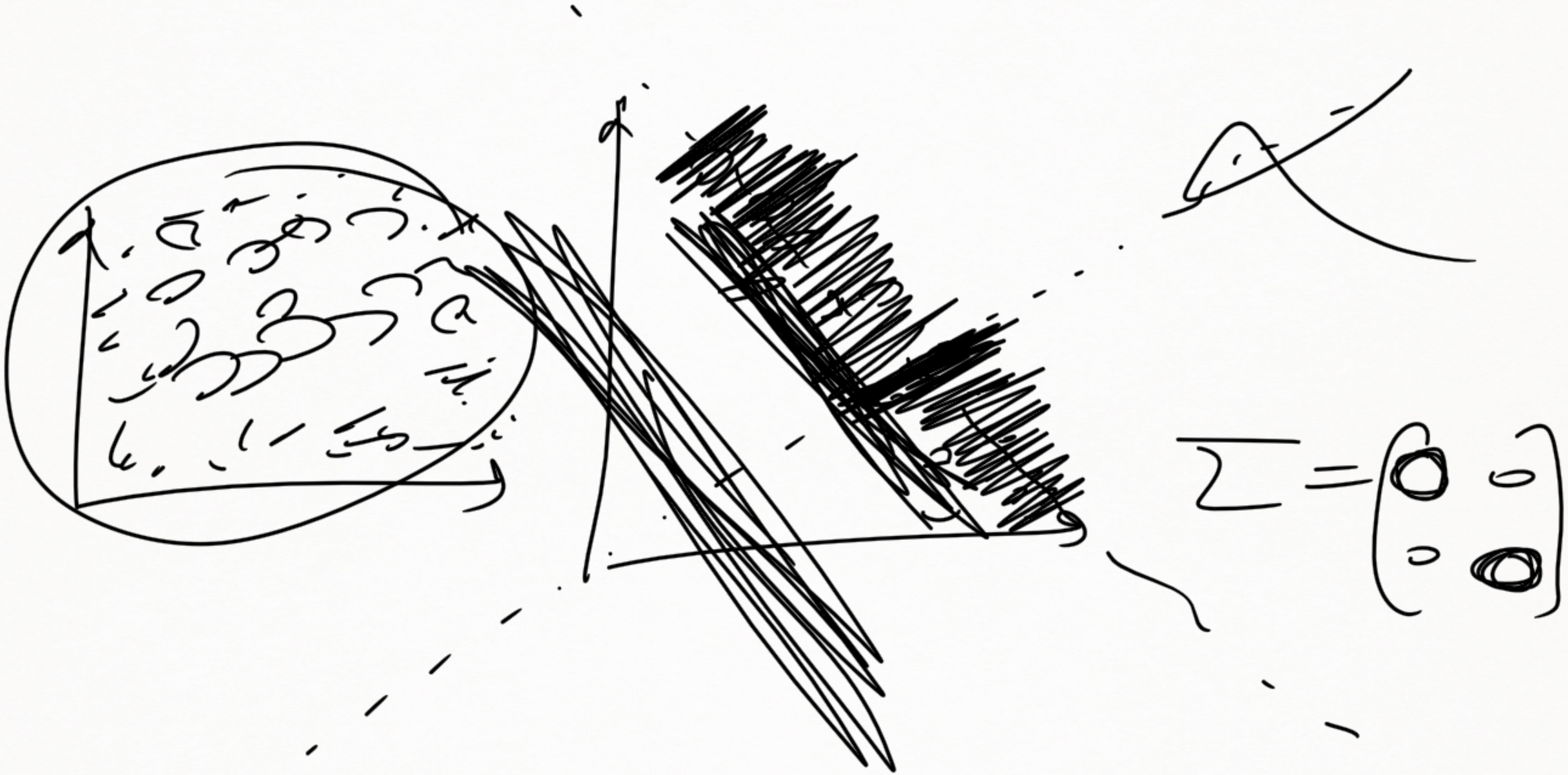
$$\frac{P\left(C=0 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{P\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = e^{\left(\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 + \hat{\theta}_3 x_1^2 + \hat{\theta}_4 x_2^2\right)}$$

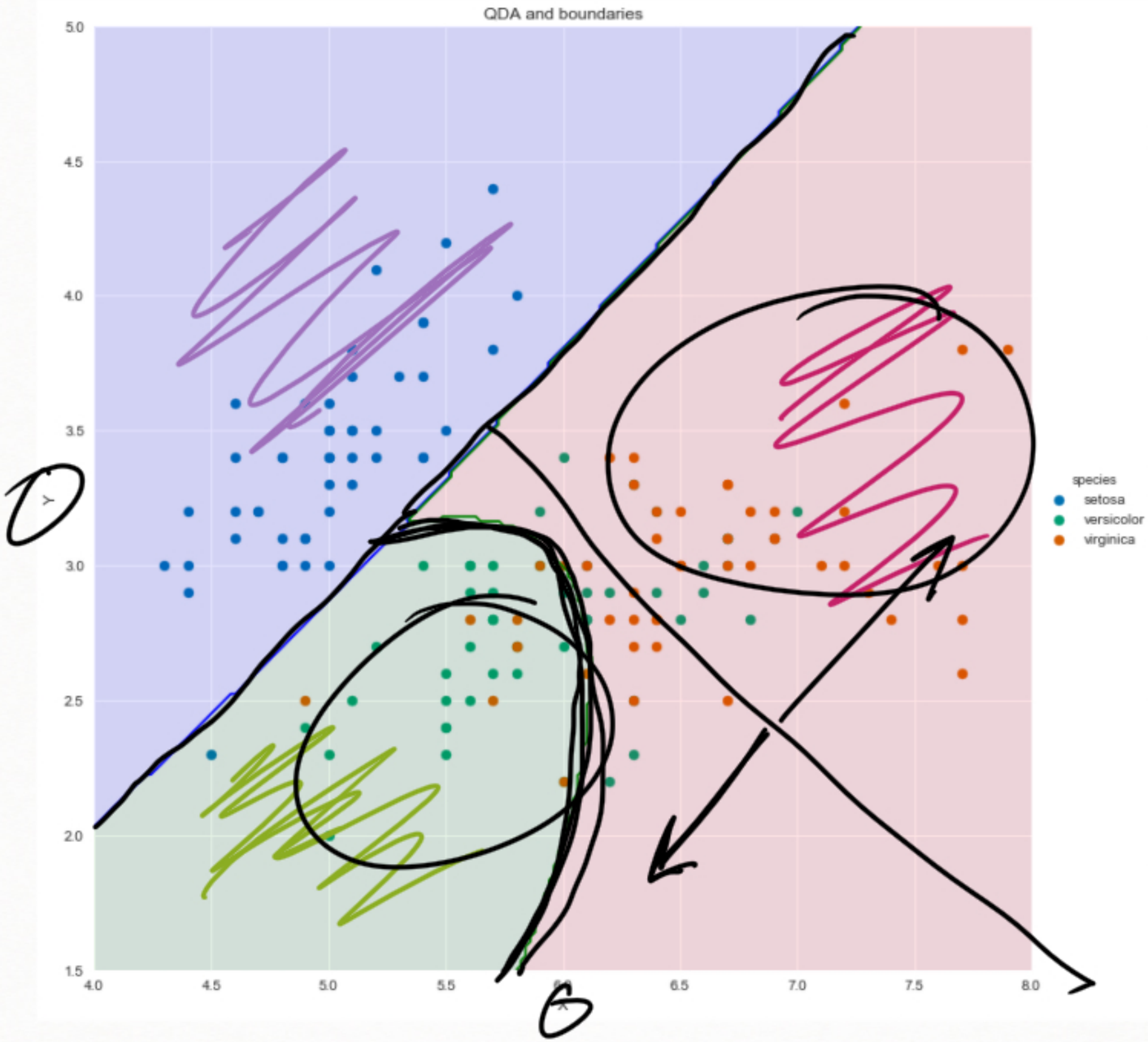
$$y = \log_e \frac{P\left(C=0 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{P\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = f(x_1, x_2) = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 + \hat{\theta}_3 x_1^2 + \underline{\underline{\hat{\theta}_4 x_2^2}}$$

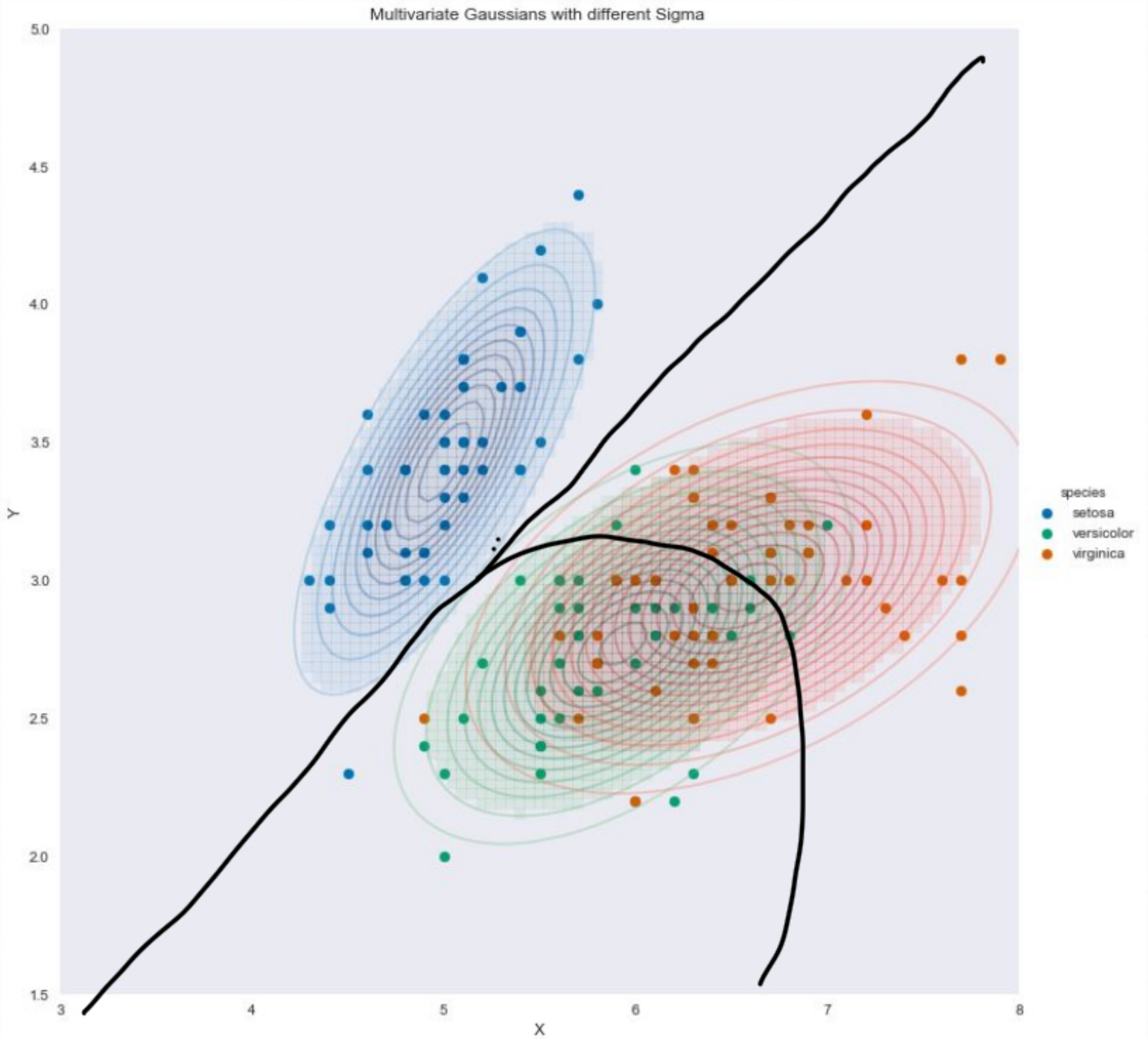

P(x|c)
$y = f(x_1, x_2) = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 + \hat{\theta}_3 x_1^2 + \hat{\theta}_4 x_2^2$
x_1
x_2

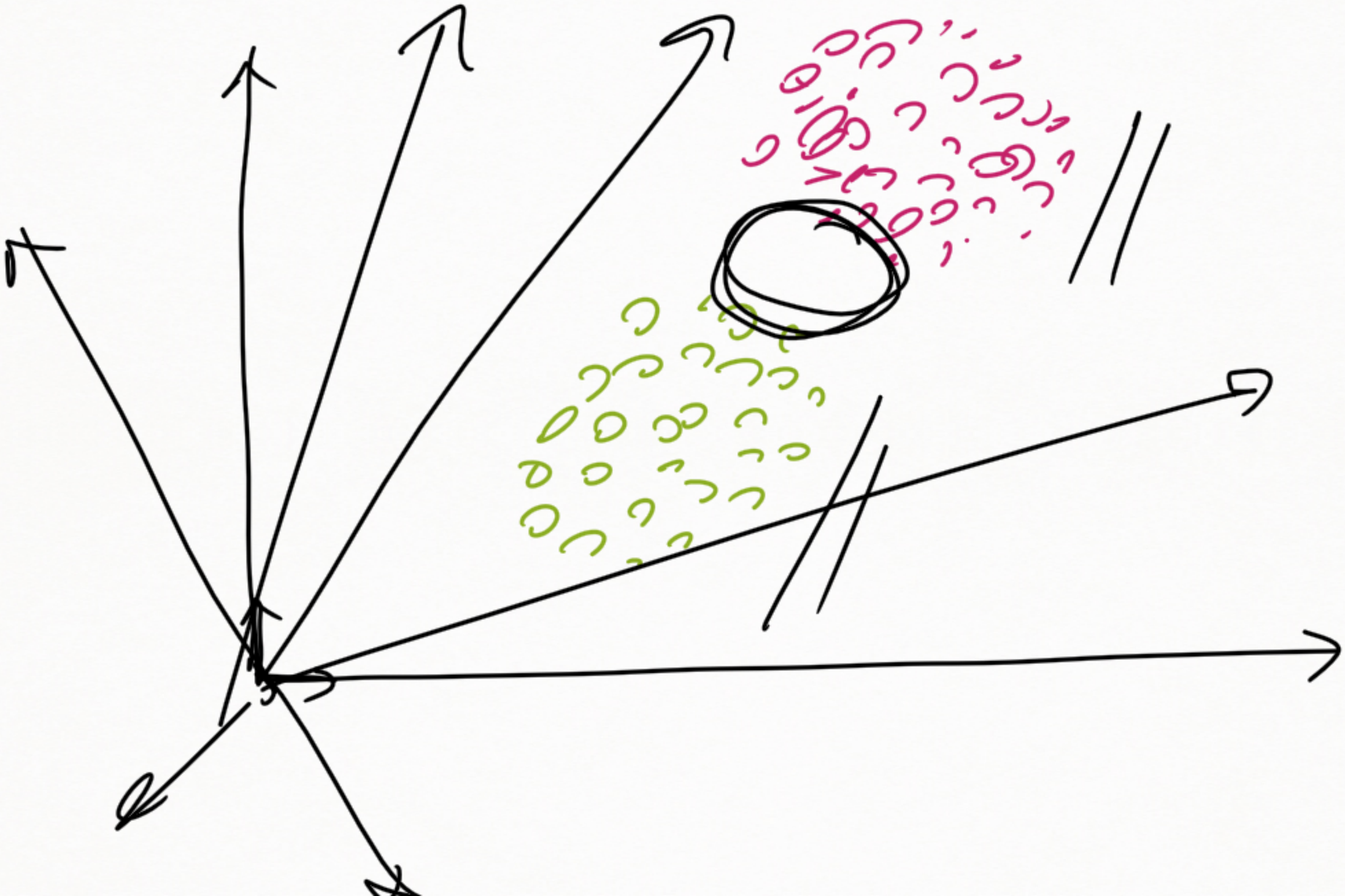

x_2
x_1

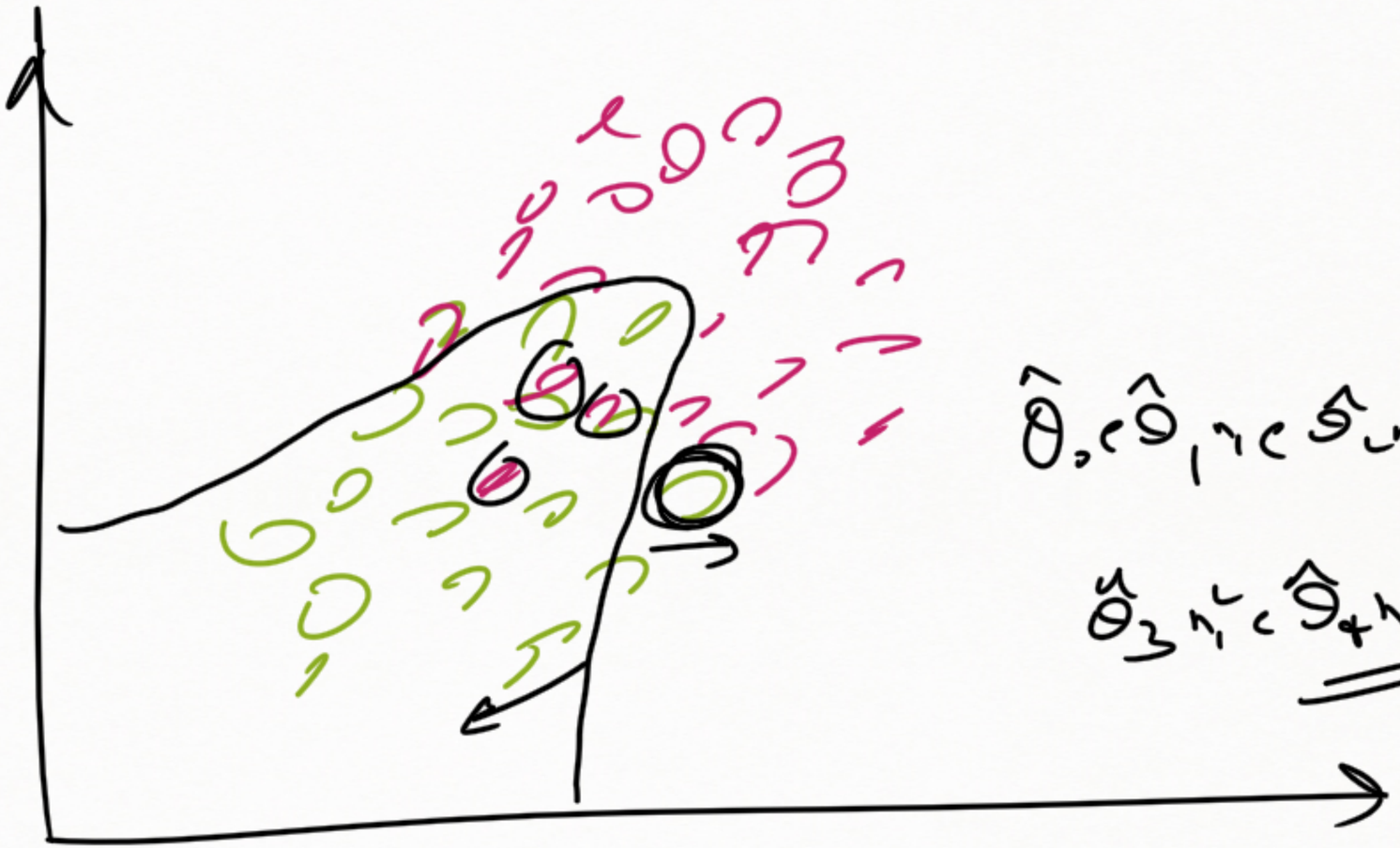

1
Σ_M
$\hat{\Sigma}_B$
RDA
2
Σ_M
Σ_B
RDA
3
Σ_M
Σ_B
4
5
Σ_M
Σ_B
6


$\Sigma = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$


QDA and boundaries
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
Y
4.0
4.5
5.0
5.5
6.5
7.0
7.5
8.0
species
setosa
versicolor
virginica


Multivariate Gaussians with different Sigma
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
Y
3
4
5
6
7
8
X
species
setosa
versicolor
virginica





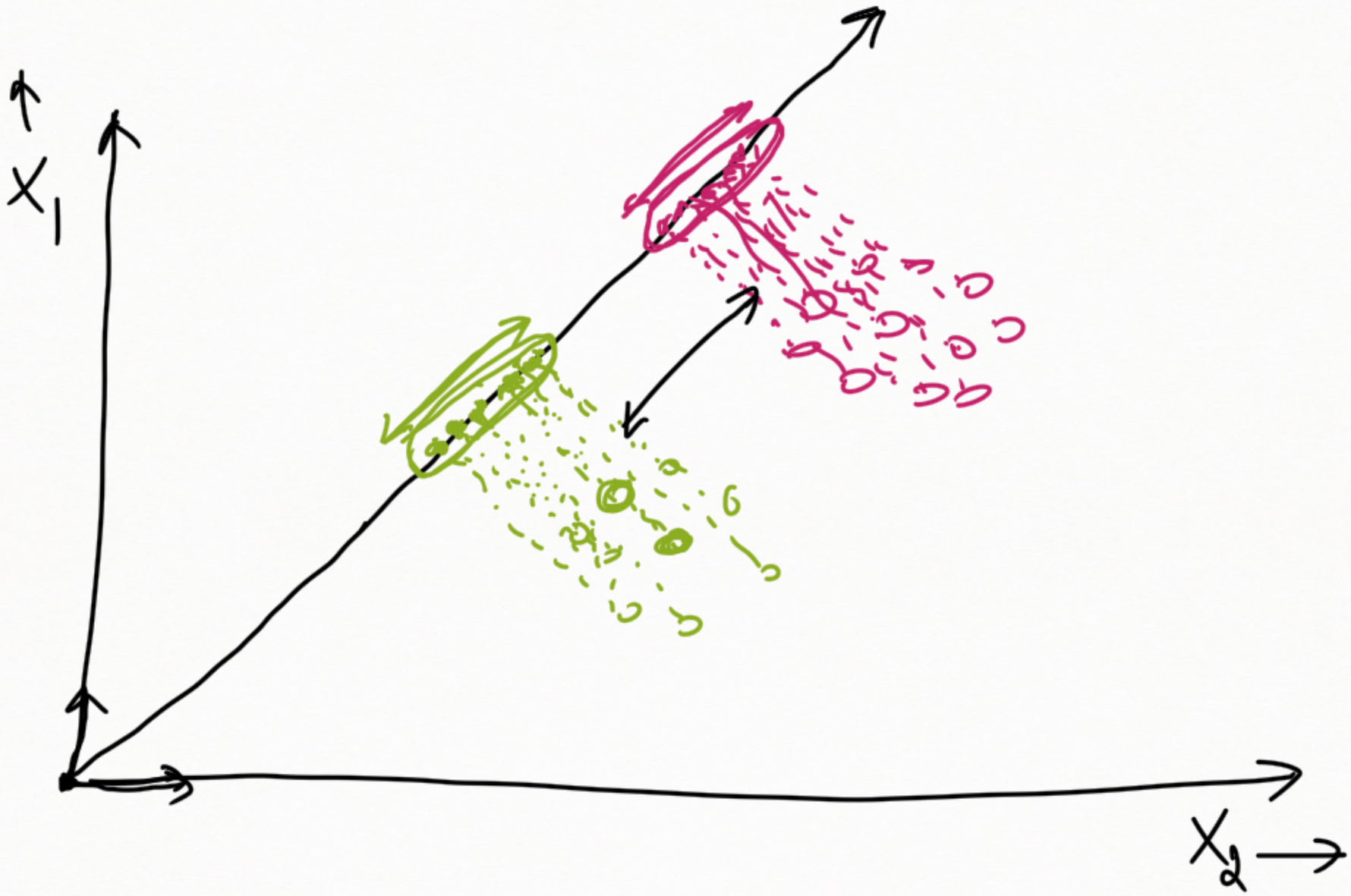
$\hat{\theta}_0 < \hat{\theta}_1 < \hat{\theta}_{n-1}$

$\hat{\theta}_3 < \hat{\theta}_4$

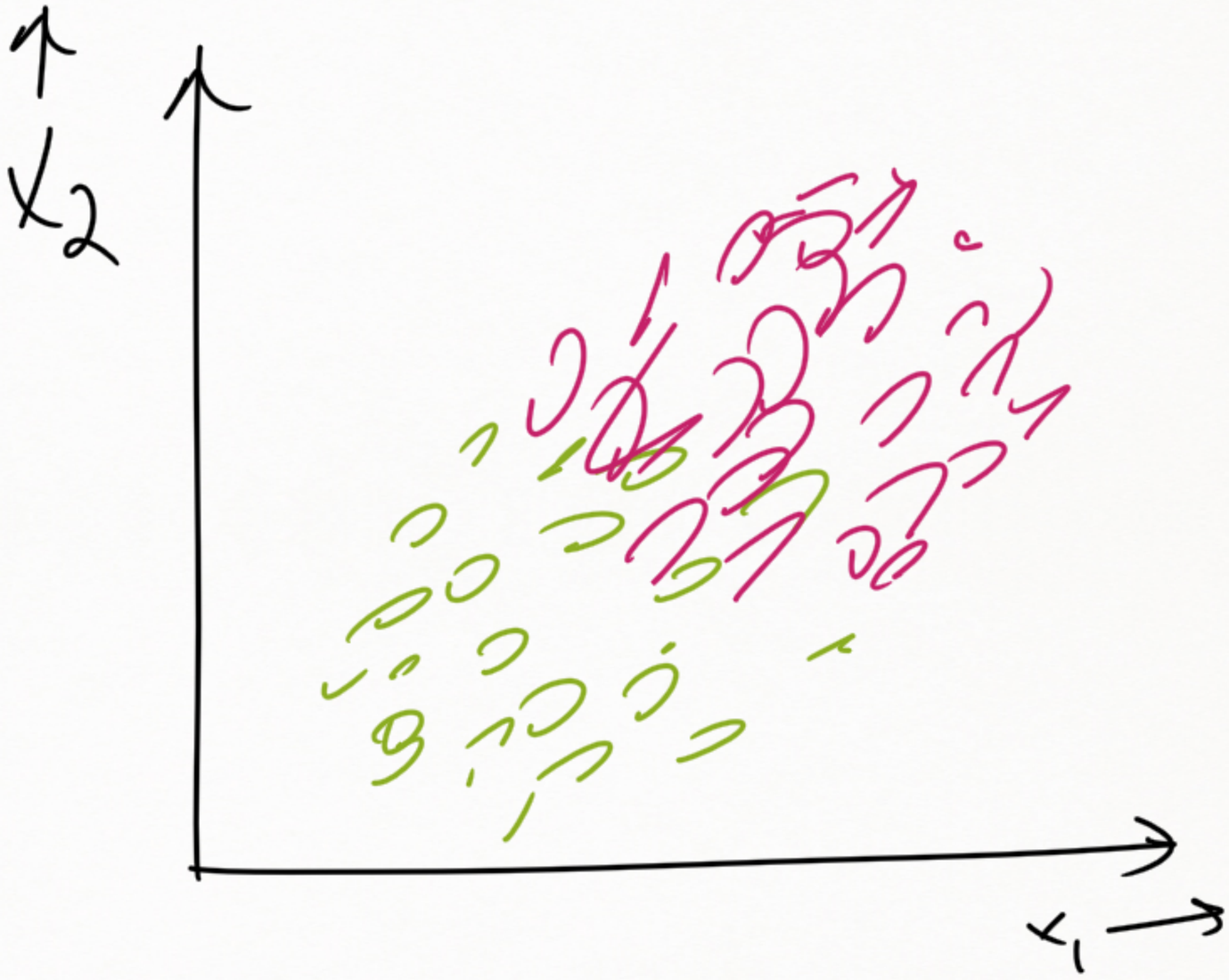
LDA → Dimensionality Reduction Technique

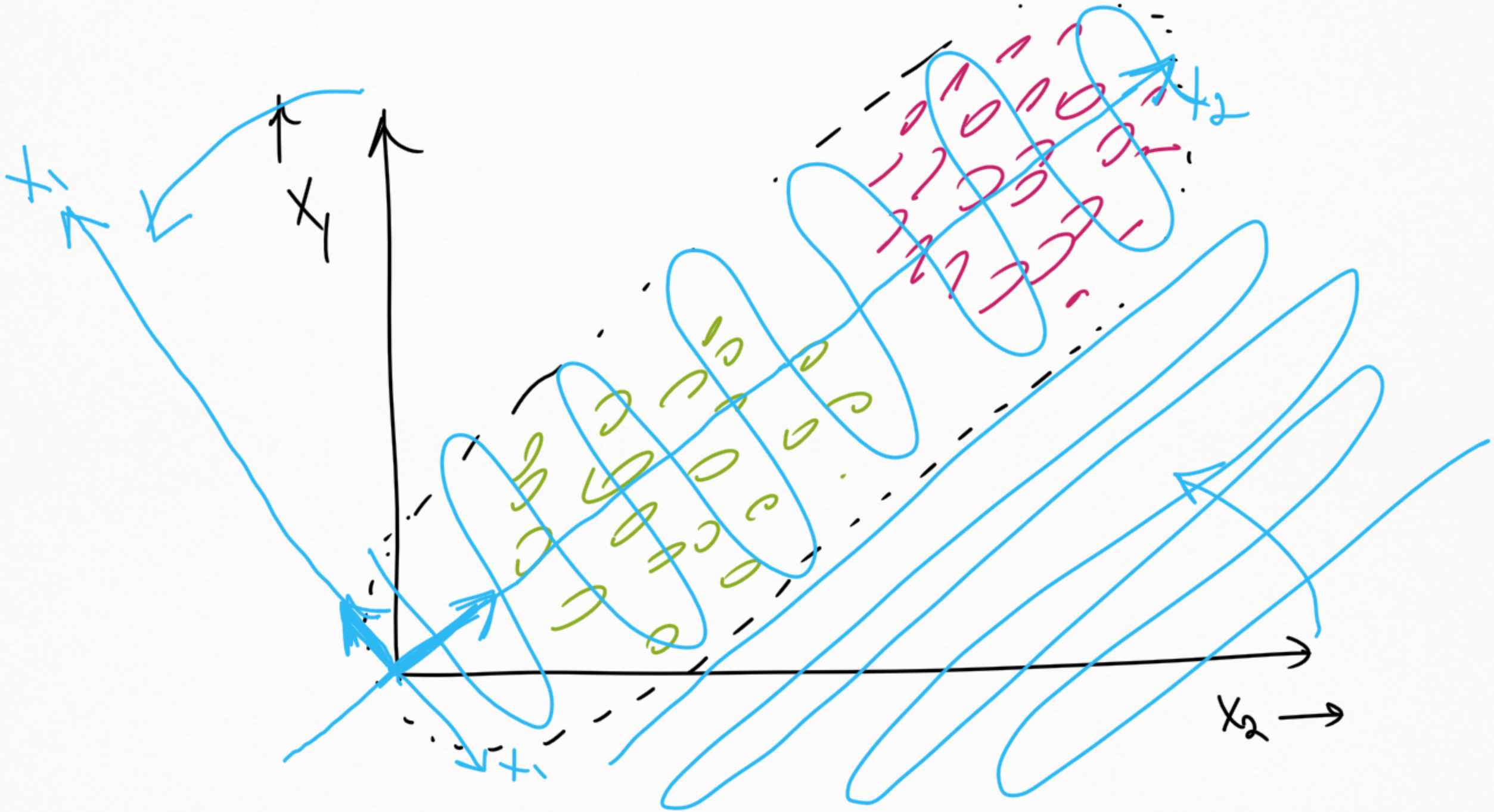
$$y = f(X) = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \dots\dots + \hat{\theta}_N x_N \rightarrow \text{Linear Discriminant Function}$$

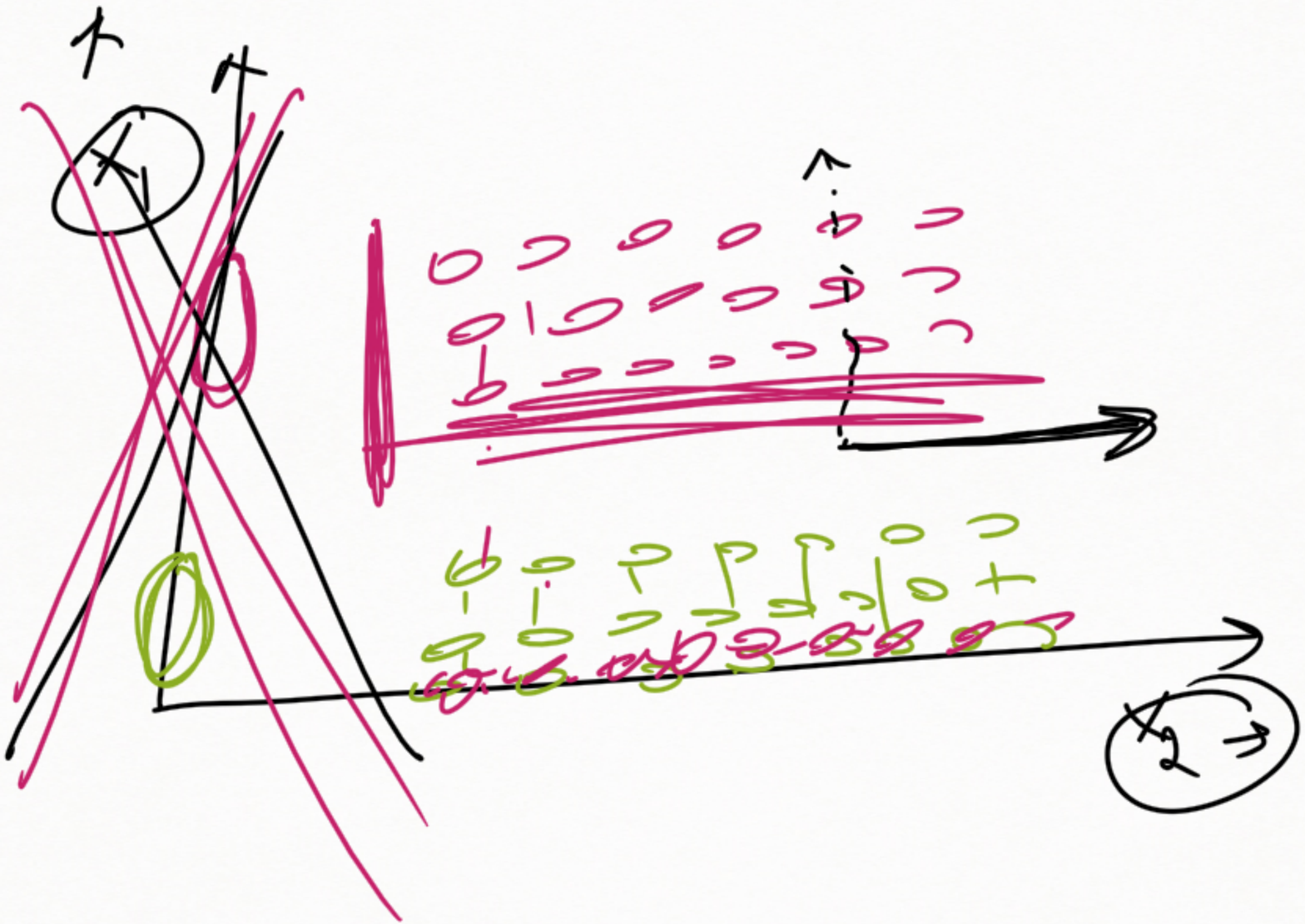
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_N \end{bmatrix}^T \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \Rightarrow \hat{\theta}^T \cdot X$$


X_1
X_2

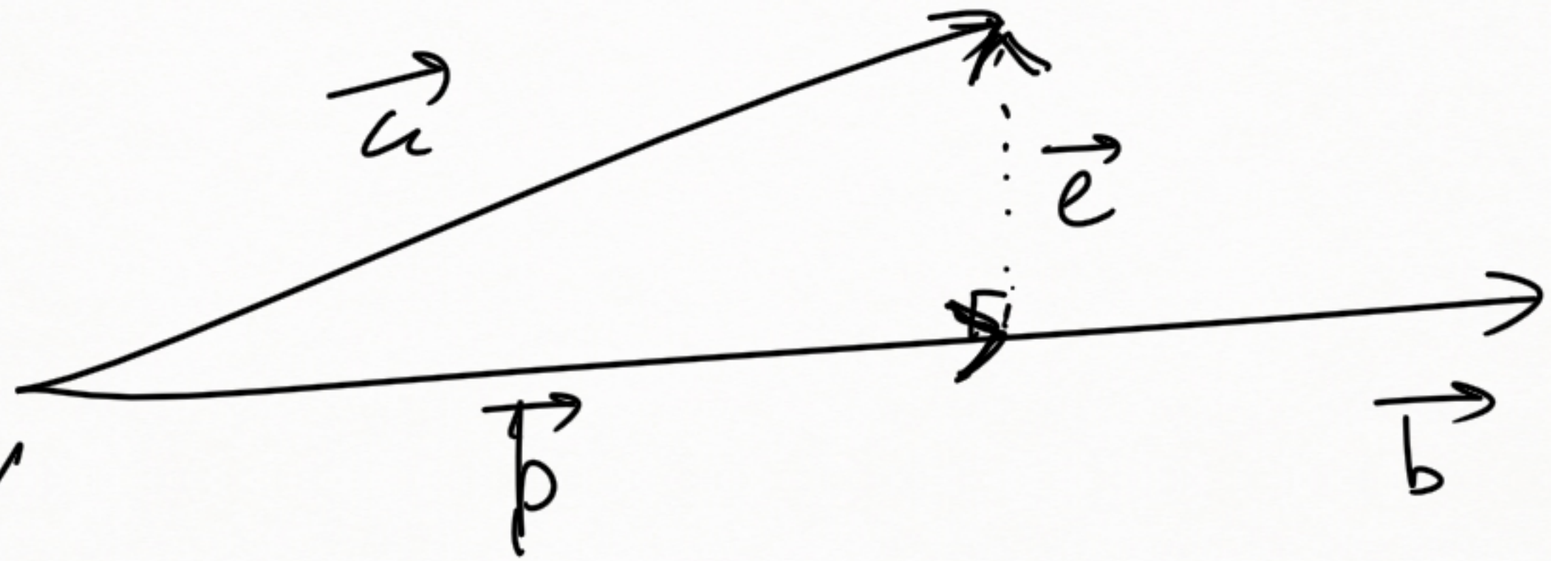
LDA v/s ICA




x_1'
x_1
x_2
x_1
$x_2 \rightarrow$


x_2

x_1
x_2



$$\boxed{\lambda = \vec{b}^T \cdot \vec{a}}$$

$\hookrightarrow$ Ubrem be von projektion $\vec{a}$ on unit-vector $\vec{b}$

$$\vec{p} = \lambda \cdot \vec{b}$$

$$\vec{a} = \vec{p} + \vec{e}$$

$$\vec{e} = (\vec{a} - \vec{p})$$

$$\vec{e} \perp \vec{b}$$

$$\vec{b}^T \cdot (\vec{a} - \vec{p}) = 0$$

$$\vec{b}^T \cdot \vec{a} - \vec{b}^T \cdot \vec{p} = 0$$

$$\vec{b}^T \cdot \vec{a} - \vec{b}^T \cdot \lambda \cdot \vec{b} = 0 \Rightarrow \boxed{\lambda = \frac{\vec{b}^T \cdot \vec{a}}{\vec{b}^T \cdot \vec{b}}}$$

if $\vec{b}$ is a unit vector

$$\boxed{\lambda = \vec{b}^T \cdot \vec{a}}$$

$$\vec{b}^T \cdot \vec{b} = \|\vec{b}\|_2^2 = 1^2 = 1$$

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

$$y - \hat{\theta}^T x = 0$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_N \end{bmatrix}^T \cdot \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \hat{\theta}^T \cdot X = 0$$

is the equation of the decision boundary

$$y = \vec{b}^T \cdot \vec{a}$$

$\hat{\theta}$

Regularized Discriminant Analysis :-

Classification in LDA :-

$$\boxed{\log_e \frac{P(C=1|X)}{P(C=0|X)}}$$

$$f(X) = \hat{\theta} X = \boxed{\log_e \frac{P(C=0|X)}{P(C=1|X)}}$$

If $f(X) > 0$ then the X belongs to class 0 (class 1)

else X belongs to class 1 (class 0)

In LDA - $\hat{\Sigma}_0 = \hat{\Sigma}_1 = \Sigma_p$

$$\boxed{\Sigma_p = \frac{N_0 \cdot \hat{\Sigma}_0 + N_1 \cdot \hat{\Sigma}_1}{N_0 + N_1}}$$

$$\hat{\theta} \longrightarrow \text{LDA} \longrightarrow f(x) \longrightarrow \hat{\theta}^T x$$

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_N \end{bmatrix} \quad x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

Regularised Discriminant Analysis :-

1. It is used when the Σ_p calculated in the case of LDA is also singular matrix.

2. It is used when along with condition 1, in the case of QDA, $\hat{\Sigma}_0$ as well $\hat{\Sigma}_1$ is a singular matrix

$$f(x) = \text{Quadratic} = f(\hat{\Sigma}_0^{-1}, \hat{\Sigma}_1^{-1})$$

Now, we will apply RDA :-

$$\boxed{\alpha \rightarrow \text{Hyperparameter}} \quad 0 \leq \alpha \leq 1$$

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

$\alpha = 1$

$\alpha = 0$

$$\boxed{\Sigma_0^{RDA} = \Sigma_p}$$

$$\boxed{\Sigma_0^{RDA} = \hat{\Sigma}_0}$$

Pooled Matrix of zero class

$$\boxed{\Sigma_1^{RDA} = \alpha \cdot \Sigma_p + (1-\alpha) \cdot \hat{\Sigma}_1}$$

$$y = f(x) = \text{having constants } \boxed{\hat{\mu}_0, \hat{\mu}_1, \Sigma_0^{RDA}, \Sigma_1^{RDA}}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad \hat{\theta}_0 = f($$

Q. $y = f(X) = \hat{\theta}^T X \Rightarrow \hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2$

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\boxed{\hat{\Sigma}_0 = \hat{\Sigma}_1 = \Sigma_p}$$

$$\boxed{\Sigma_p = \frac{N_0 \cdot \hat{\Sigma}_0 + N_1 \cdot \hat{\Sigma}_1}{N_0 + N_1}}$$

$$\underline{\underline{\theta}} \quad y = f(x) = \hat{\theta}^T x' \Rightarrow \hat{\theta}_0, \hat{\theta}_1, \underline{\underline{\hat{\theta}_2, \hat{\theta}_3}}, \hat{\theta}_4$$

$$x \rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\hat{\theta}^T \rightarrow \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \\ \hat{\theta}_4 \end{bmatrix}$$

$$x' \rightarrow \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_2^2 \end{bmatrix}$$

Quadratic Discriminant function

Q: $y = f(X) \rightarrow \hat{\theta}^T X \longrightarrow$ How it will <u>look</u> like?

In case of RDA

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

$$\boxed{\Sigma_1^{RDA} = \alpha \cdot \Sigma_p + (1-\alpha) \cdot \hat{\Sigma}_1}$$

$$\boxed{P(C=0 \mid x) = \frac{1}{1+e^{-(\theta_0+\theta^T x)}}}$$

Discriminant function

Sigmoid Function