Basic Example of Tumor Classification :-

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

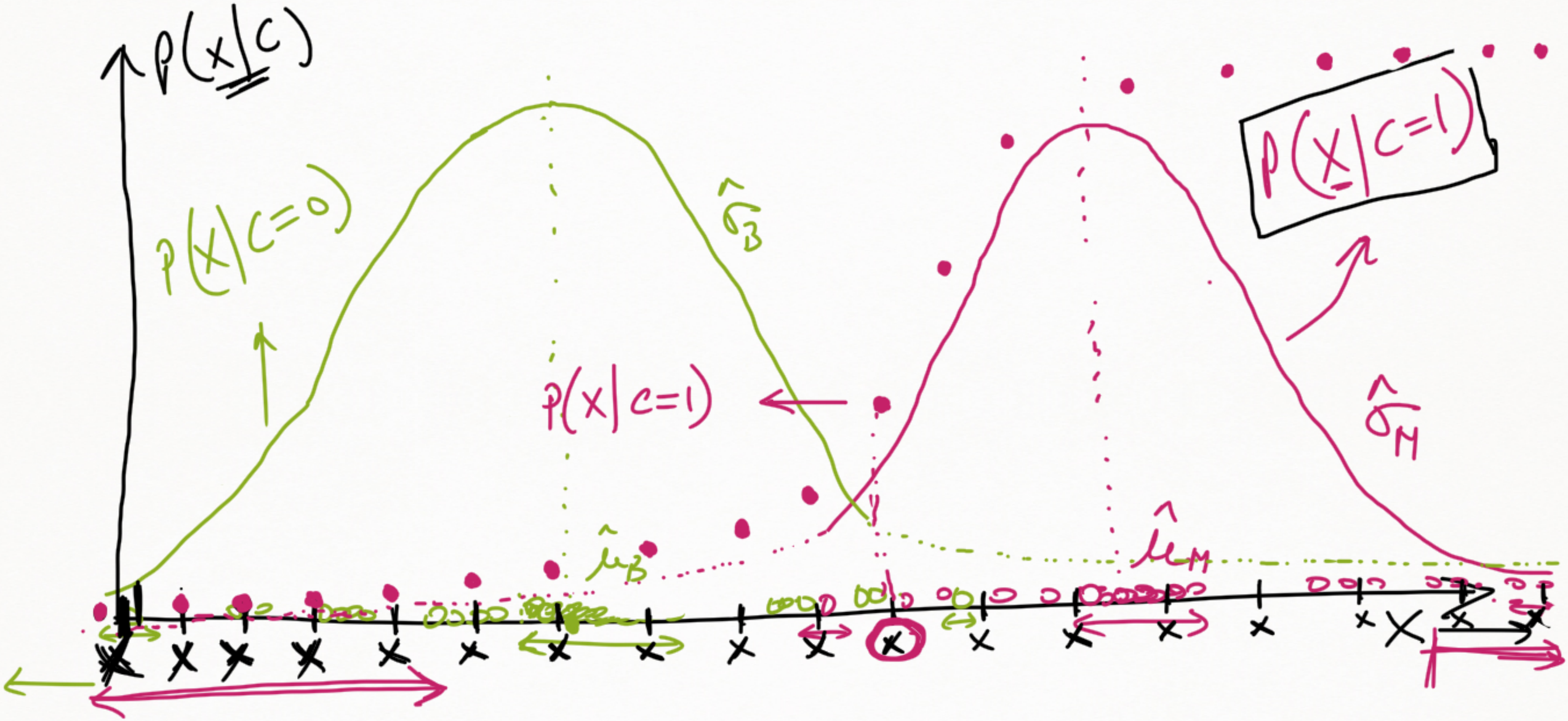
2 columns

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)
$\hat{\sigma}_B$
$\hat{\sigma}_M$
$\hat{\mu}_B$
$\hat{\mu}_M$

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

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

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

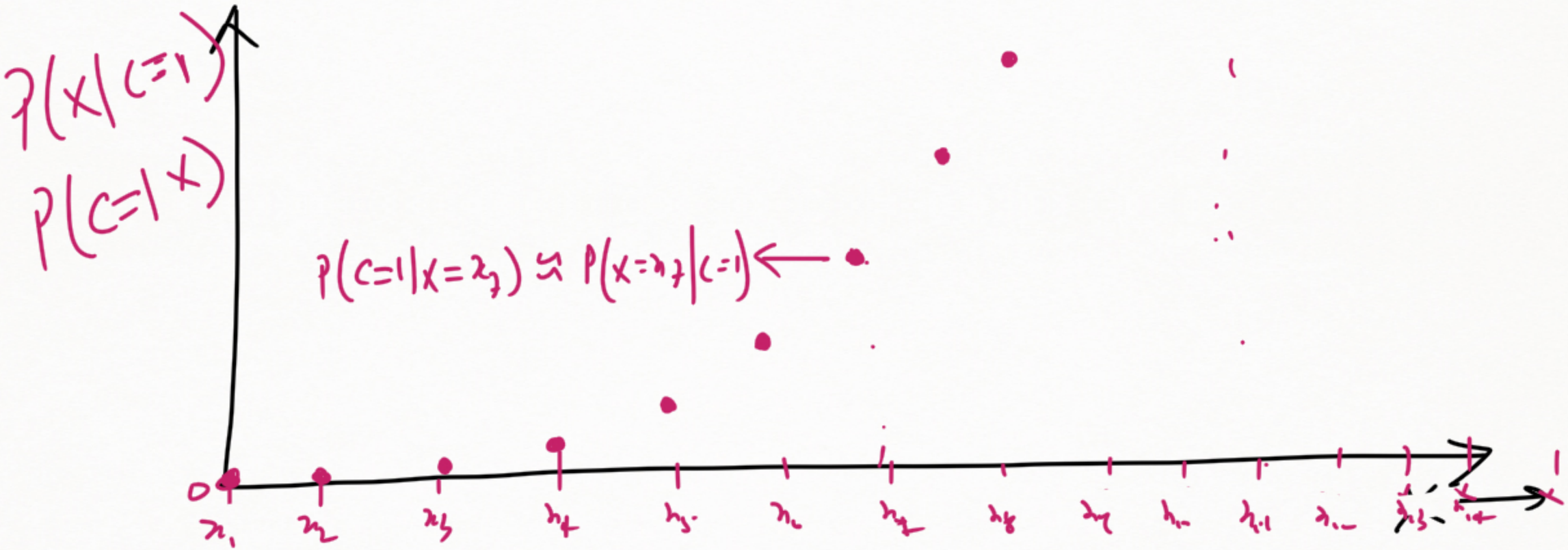
$\boxed{0.57}$

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

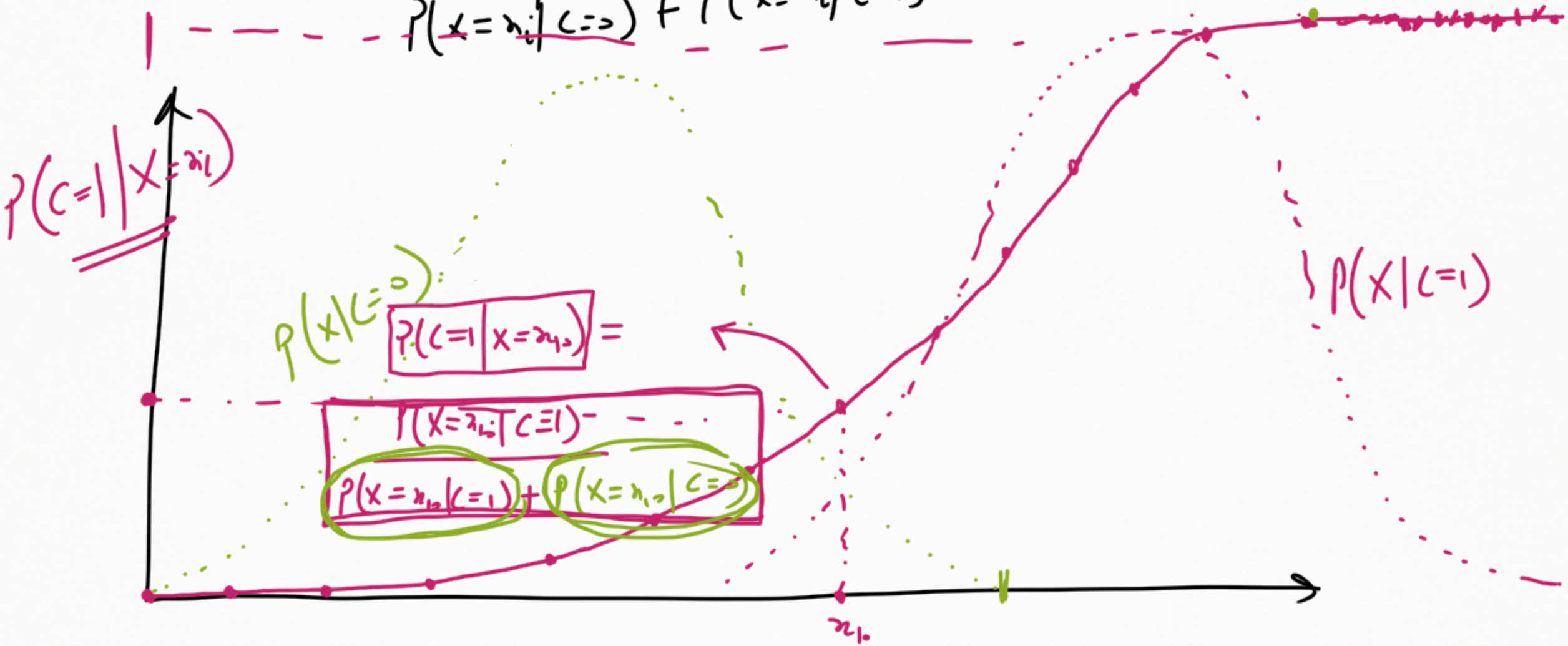
Denominator which remains constant cut for both $\underbrace{P(C=1|x)}$ and $\underbrace{P(C=0|x)}$

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

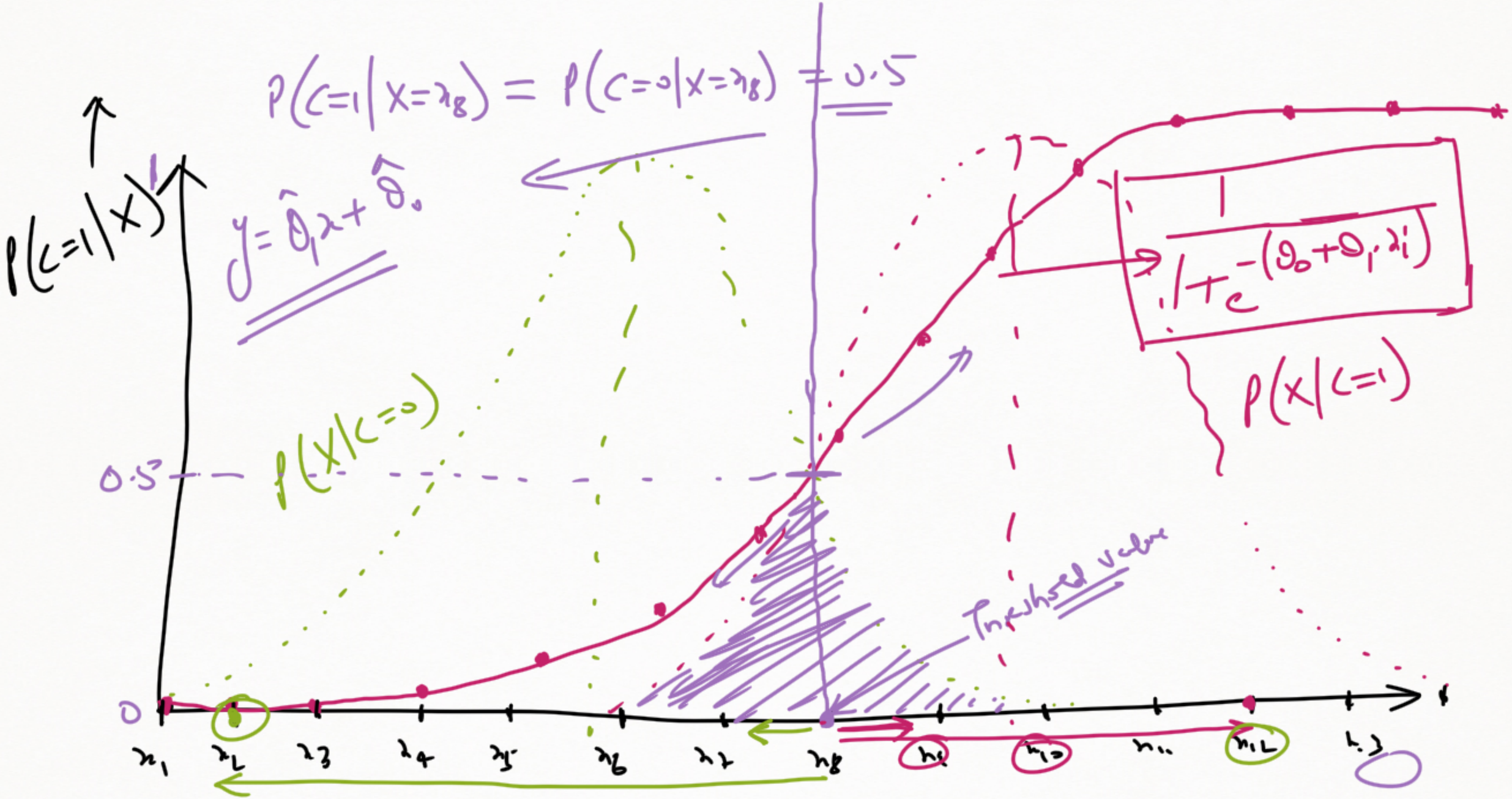
$$P(c=1|x=x_n) \approx P(x=x_n|c=1)$$



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



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


$P(C=1|X=x_8) = P(C=0|X=x_8) = 0.5$
$P(C=1|X)$
$y = \hat{\theta}_1 x + \hat{\theta}_0$
$P(X|C=0)$
$\frac{1}{1+e^{-(\theta_0+\theta_1 x_i)}}$
$P(X|C=1)$
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 $\theta_0 = -\hat{\theta}_0$, $\theta_1 = -\hat{\theta}_1$

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

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

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

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

Sigmoid function

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

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

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

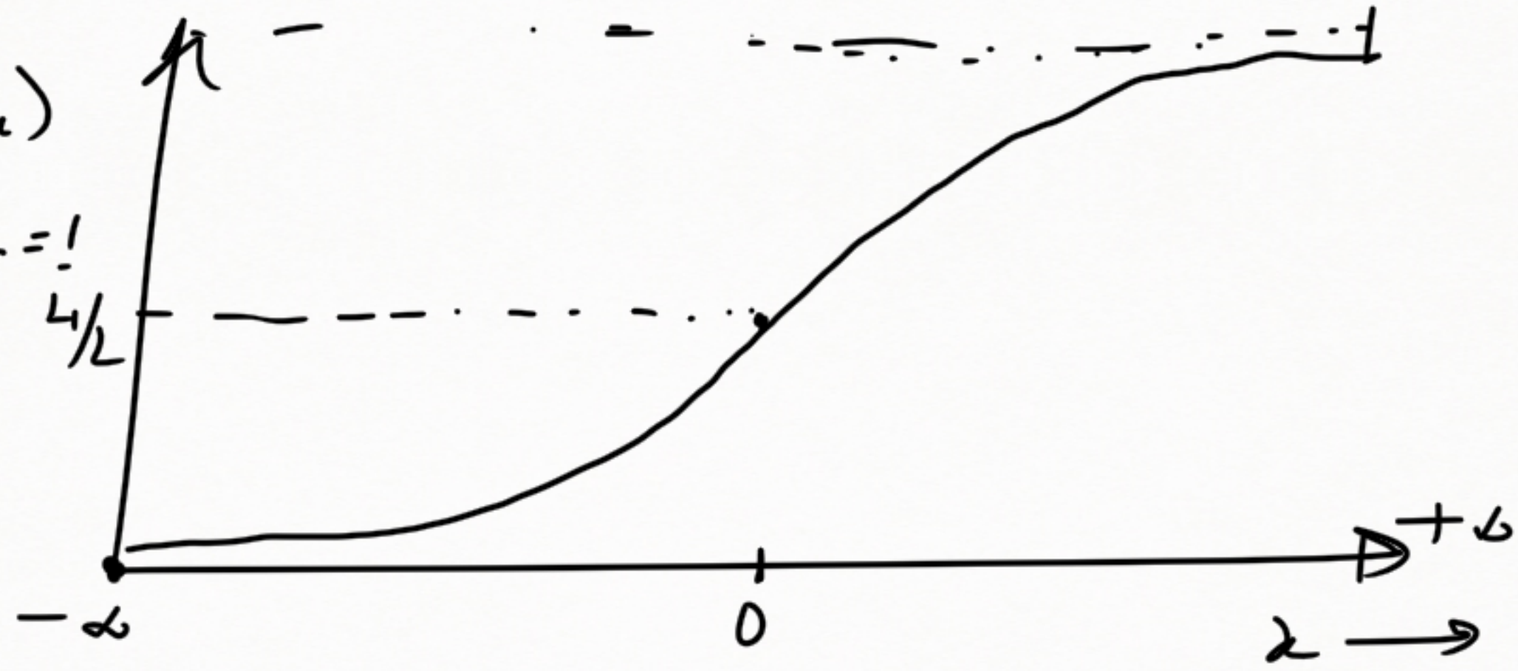
$$\frac{1}{1+\infty} = \frac{1}{\infty} = \underline{\underline{0}}$$

$x=0$

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

$x=+\infty$

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



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

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

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

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

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

$$\boxed{x_8 = \frac{\left(\hat{\mu}_B^L - \hat{\mu}_M^L\right)}{\dfrac{2\cancel{\sigma_p^L}}{\dfrac{\left(\hat{\mu}_B - \hat{\mu}_M\right)}{\cancel{\sigma_p^L}}}} = \frac{\hat{\mu}_B + \hat{\mu}_M}{2}}$$

Inflection point

$(0,1)$

$x_i > x_8$

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

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

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

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

If $x > x_g$ then –

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

Linear Discriminant function

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

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

$= \sigma_p^2$

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

If $x_i < x_g$ then –

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

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

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

$$\underbrace{\log_e \frac{p(C=0 \mid X=x_i)}{p(C=0 \mid X=x_i)}}_{y_i} = (\hat{\theta}_0 + \hat{\theta}_1 x_i)$$

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

$$\boxed{y_i = \theta_1 x_i + \theta_0}$$

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

QDA

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