CLASIFICACIÓN

- Y toma valores en un conjunto no ordenado
  - color de ojos ∈ {brown, azules o verdes}
  - pingüinos ∈ {emperador, }
  - iris ∈ {setosa, virginica y versicolor}

son no ordenados

virginica, setosa, versicolor

- experiencia de usuario ∈ {mala, neutra, buena}

0 1 2

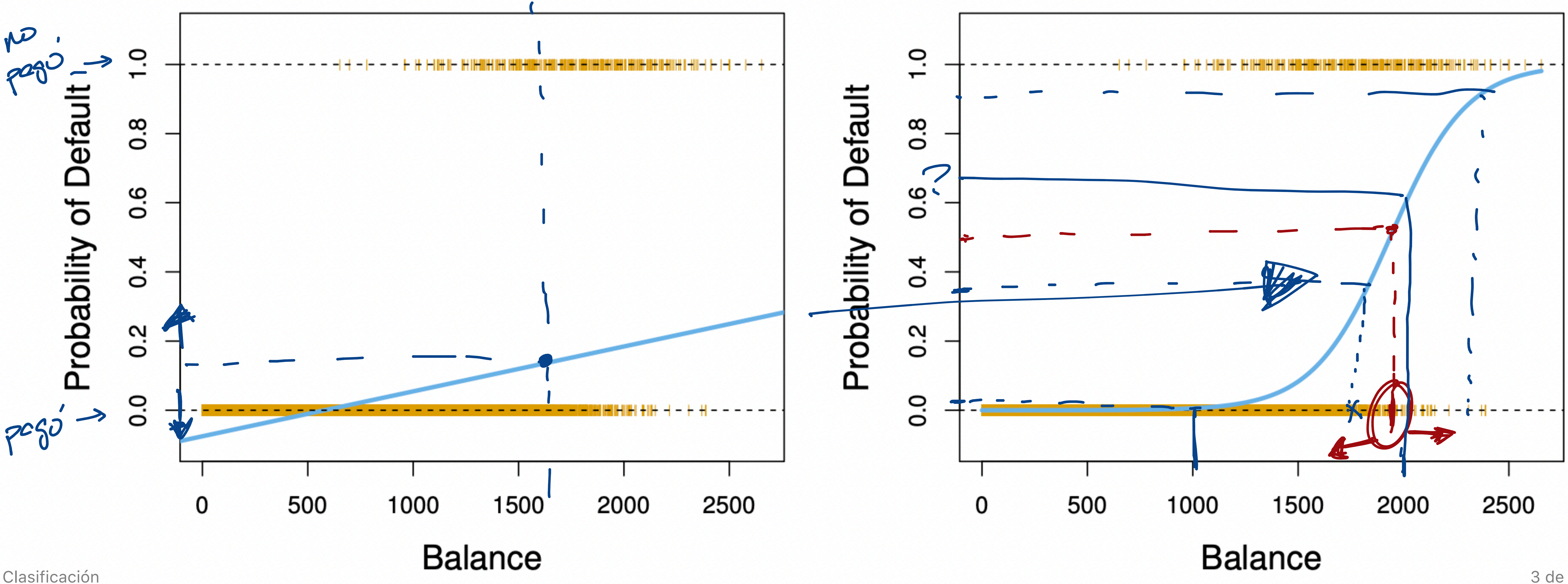
Como ya vimos al hacer clustering:

- clasificar (decir pertenece a tal clase)
- podemos querer dar probabilidades de pertenencia
- quiero clasificar pero usando algún modelo de probabilidad

"k-medios"

"k-medios difuso"

"EM"


no pagó
pagó
Probability of Default
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
Balance
?

$$Y = \begin{cases} 0 & \text{si} \quad \text{no} \\ 1 & \text{si} \quad \text{si} \end{cases}$$

$\hat{Y} > 0.5 \rightarrow$ si

$\hat{Y} \leq 0.5 \leftarrow$ no.

Train	Test
	$Y \in \{0,1\}$
	$\hat{Y} \in [0,1]$

		Predicción	
		si	no
Valor verdader.	si	n_{11}	n_{12}
	no	n_{21}	n_{22}

Error de tipo I.

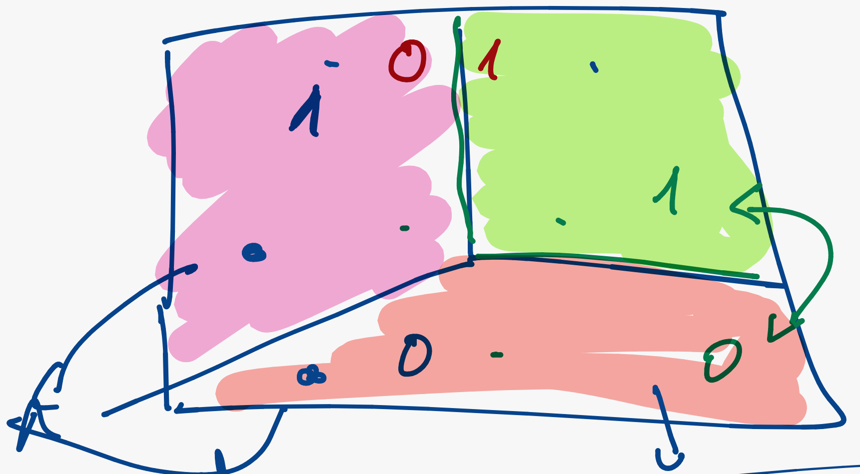
$\alpha =$ Tasa de Falsos positivos

Error de tipo II

$\beta =$ Tasa de Falsos negativos

$$\Pr(Y=1 \mid X=x) = E(Y \mid X=x)$$

$Y \sim Ber(p)$



modelo que
me de probabilidad
x pertenencia a
una o a otra

$$1 - P(Y = \text{rosado} \mid X) - P(Y = \text{verde} \mid X)$$

REGRESIÓN LOGÍSTICA

$$\boxed{P(Y=1 \mid X=x) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}$$

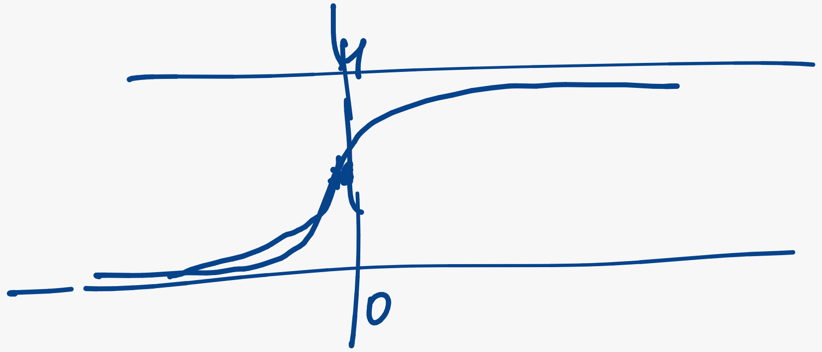
(En el ejemplo: prob. de default conociendo el balance).

$$h(u) = \frac{e^{u}}{1+e^{u}}$$

$$h: \mathbb{R} \to (0,1)$$

$$\lim_{u\to-\infty} \frac{e^{u} \to 0}{1+e^{u} \searrow 0} = 0$$

$$\lim_{u\to+\infty} \frac{e^{u}}{1+e^{u}} = 1$$



$$\boxed{\underline{\log}\left(\frac{P(Y=1|X)}{1-P(Y=1|X)}\right) = \beta_0 + \beta_1 X}$$

$$h^{-1}(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$$

↑

logit , log odds

↳ ln

default ~ balance $\quad \boxed{\text{Default}}$

$$P(Y=1 \mid X=x) = \frac{e^{\beta_0+\beta_1 x}}{1-e^{\beta_0+\beta_1 x}}$$

$$\hat{\beta}_0 = -10.6513$$

$$\hat{\beta}_1 = 0.0055$$

0

$\uparrow$

$$P(Y=1 \mid X=1000) = \frac{e^{-10.6513+0.0055\times 1000}}{1-e^{-10.6513+0.0055\times 1000}} \approx 0.006$$

$$P(Y=1 \mid X=2000) = \frac{e^{-10.6513+0.0055\times 2000}}{1-e^{-10.6513+0.0055\times 2000}} \approx 0.586$$

$\downarrow$

1

default ~ student

$$\mathbb{P}(Y = Si \mid estudiante = si) = \frac{e^{-3.504 + 0.405 \cdot 1}}{1 - e^{3.504 + 0.405 . 1}} \approx 0\ 0431$$

$$\shortparallel\ \frac{e^{\beta_0 + \beta_1}}{1 - e^{\beta_0 + \beta_1}}$$

$$\mathbb{P}(Y = Si \mid estudiante = no) = \frac{e^{-3.504 + 0.405 \cdot 0}}{1 - e^{-3.504 + 0.405 \cdot 0}} \approx 0.0292$$

$$\shortparallel\ \frac{e^{\beta_0}}{1 - e^{\beta_0}}$$

$\hat{\beta}_0 = -3.504$

$\hat{\beta}_1 = 0.405$

$\text{logit}(\text{default}) = \beta_0 + \beta_1\,\text{balance} + \beta_2\,\text{income} + \beta_3\,\text{student} + \varepsilon$

$\hat\beta_0 = -10.87$

$\hat\beta_1 = -0.0057$

$\cancel{\hat\beta_2 = -3\times10^{-6}}$ ← p-valor 0.711

$\hat\beta_3 = -0.6468$

$$P(Y=1|x) = \frac{e^{-10.87 - 0.0057\,\underline{bal} - 0.6468\,st}}{1 - e^{\boxed{}}}$$

Regresión logística para más de dos clases ⇐ Regresión multinomial.

$$\mathbb{P}(Y=k \mid X) = \frac{e^{\beta_{0k}+\beta_{1k}X_1+\dots+\beta_{pk}X_p}}{\sum_{l=1}^{K} e^{\beta_{0l}+\beta_{1l}X_1+\dots+\beta_{pl}X_p}} \quad \longleftarrow$$

↗ una regresión lineal para cada clase

ANÁLISIS DISCRIMINANTE.

Tenemos k clases (conocidas)

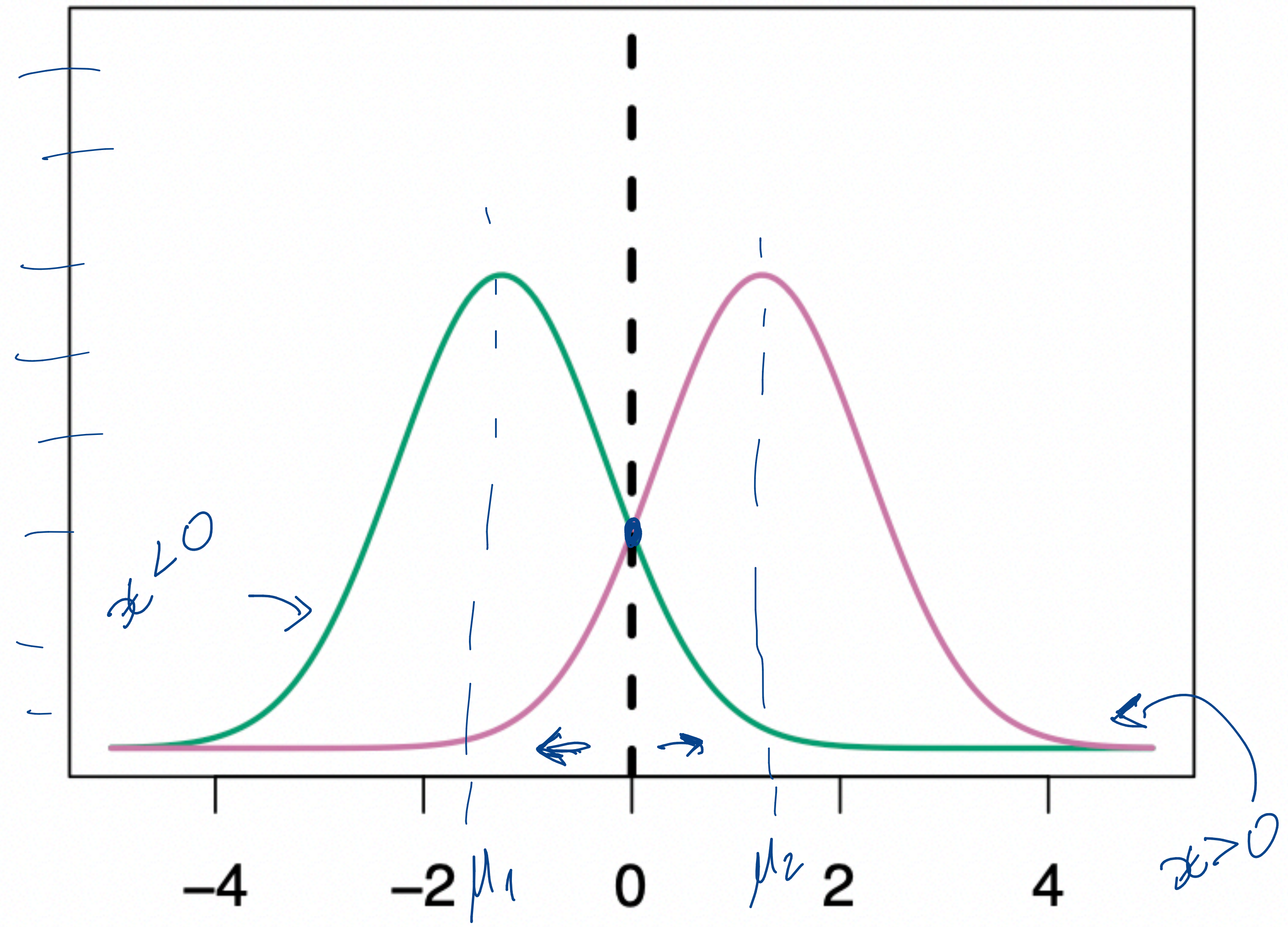
Queremos calcular $\mathbb{P}(Y \mid X)$ usando Bayes

Vamos a suponer $X \sim N$ dentro de cada clase.

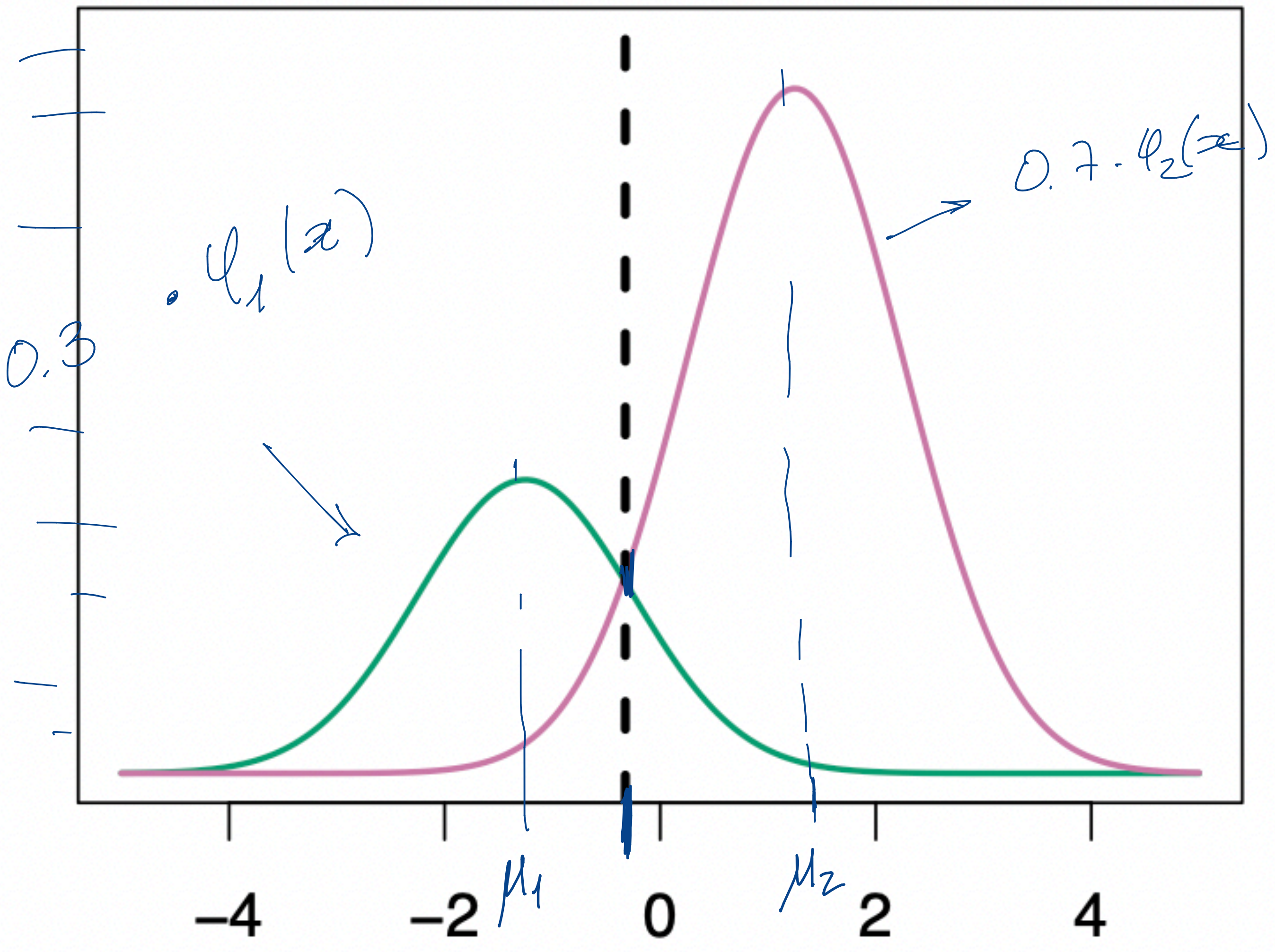
- Si todas las clases tienen igual $\underset{\Sigma}{\sigma}$ → Análisis discriminante lineal
- Si tienen diferentes σ → Análisis discriminante cuadrático.

$$=: \frac{\overline{Y=k)} \cdot \overbrace{\mathbb{P}(Y=k)}^{\pi_k}}{} = \boxed{\frac{\pi_k \cdot \varphi_k(x)}{\sum_{\ell=1}^{K} \pi_\ell \cdot \varphi_\ell(x)}}$$

prior



π_1=.5, π_2=.5



π_1=.3, π_2=.7

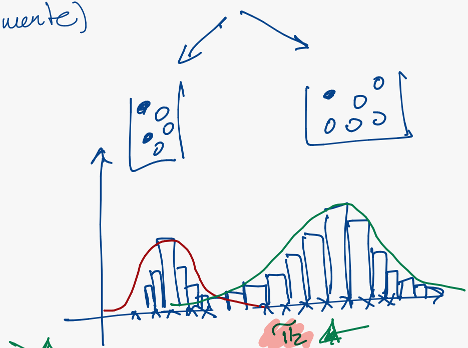
Expectation Maximization (intuitivamente)

FLASH BACK

Inicialización

Expectation (asignación)

Maximization (actualización)



$$p(x) = \overset{\pi_1}{p}\cdot \varphi_1(x) + \overset{\pi_2}{(1-p)}\,\varphi_2(x)$$

$$\varphi_1 \downarrow \sim \mathcal{N}(\underline{\mu_1}, \underline{\sigma_1}) \qquad \varphi_2 \downarrow \mathcal{N}(\underline{\mu_2}, \underline{\sigma_2})$$

$$\varphi_1 \to p(x \mid \bullet)$$

$$\Theta = (p, \underline{\mu_1, \sigma_1}, \underline{\mu_2, \sigma_2})$$

$$\mathbb{P}(Y=k \mid X=x) = \frac{\pi_k \cdot \varphi_k(x)}{\boxed{\sum_{l=1}^{K} \pi_l \cdot \varphi_l(x)}} \nleftarrow \mathbb{P}(X=x)$$

$$\mathbb{P}(Y=k_1 \mid X=x) \overset{?}{>} \mathbb{P}(Y=k_2 \mid X=x)$$

$$\pi_{k_1} \cdot \varphi_{k_1}(x) \overset{?}{>} \pi_{k_2} \varphi_{k_2}(x)$$

En una variable $p=1$

$$\varphi_k(x) = \frac{1}{\sqrt{2\pi}\,\sigma_k} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2}$$

$$\mathbb{P}(X=x \mid Y=k) \sim \mathcal{N}(\mu_k, \sigma_k^2)$$

Suponiendo que $\sigma_k = \sigma \ \forall k$

$$\mathbb{P}(Y=k \mid X=x) \overset{\cancel{\mathbb{P}(X=x)}}{\propto} \pi_k \cdot \frac{1}{\sqrt{2\pi}\,\sigma_k} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2}$$

$$\log \mathbb{P}(Y=k_1 \mid X=x) = \log(\pi_k) + \cancel{\log\left(\frac{1}{\sqrt{2\pi}\,\sigma_k}\right)} - \frac{1}{2}\left(\frac{x-\mu_k}{\boxed{\sigma_k}}\right)^2$$

$$\|$$

$$\log(\pi_k) - \cancel{\frac{1}{2}\frac{x^2}{\sigma^2}} + \frac{x\cdot\mu_k}{\sigma^2} - \frac{\mu_k^2}{\sigma^2}$$

$\circ\ X$

$\log \mathbb{P}(Y=k_1 \mid X=x)$

$\log \mathbb{P}(Y=k_2 \mid X=x)$

$\vdots$

$\log \mathbb{P}(Y=k_K \mid X=x)$

$$\boxed{\delta_k(x) = x \cdot \frac{\mu_k}{\sigma^2} \boxed{- \frac{\mu_k^2}{\sigma^2} + \log(\pi_k)}}$$

DISCRIMINANTE LINEAL

Ejemplo $K = 2$

$$\delta_1(x) = x \cdot \frac{\mu_1}{\sigma^2} - \frac{\mu_1^2}{2\sigma^2} + \log(\pi_1)$$

$$\delta_2(x) = x \cdot \frac{\mu_2}{\sigma^2} - \frac{\mu_2^2}{2\sigma^2} + \log(\pi_2)$$

¿Cuándo $\delta_1(x) < \delta_2(x)$?

$\delta_1(x) = \delta_2(x) \Leftarrow$ boundary

$$\delta_1(x) = x\cdot\frac{\mu_1}{\sigma^2} - \frac{\mu_1^2}{2\sigma^2} + \log(\pi_1)$$

$$\delta_2(x) = x\cdot\frac{\mu_2}{\sigma^2} - \frac{\mu_2^2}{2\sigma^2} + \log(\pi_2)$$

$$x\cdot\frac{\mu_1}{\sigma^2} - \frac{\mu_1^2}{2\sigma^2} + \log(\pi_1) = x\cdot\frac{\mu_2}{\sigma} - \frac{\mu_2^2}{2\sigma^2} + \log\pi_2$$

$$\left|\; + \frac{\mu_1^2}{2\sigma^2} - \log(\pi_1) \quad - \frac{x\mu_2}{\sigma}\right.$$

$$\left|\; A^2 - B^2 = (A+B)(A-B)\right.$$

$$x\cdot\left(\frac{\mu_1-\mu_2}{\sigma^2}\right) = \frac{\mu_1^2-\mu_2^2}{2\boxed{\sigma^2}} + \log(\pi_2) - \log(\pi_1)$$

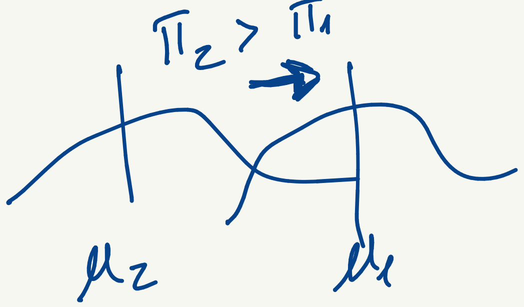
$$\left|\; : \frac{\mu_1-\mu_2}{\sigma^2}\right.$$

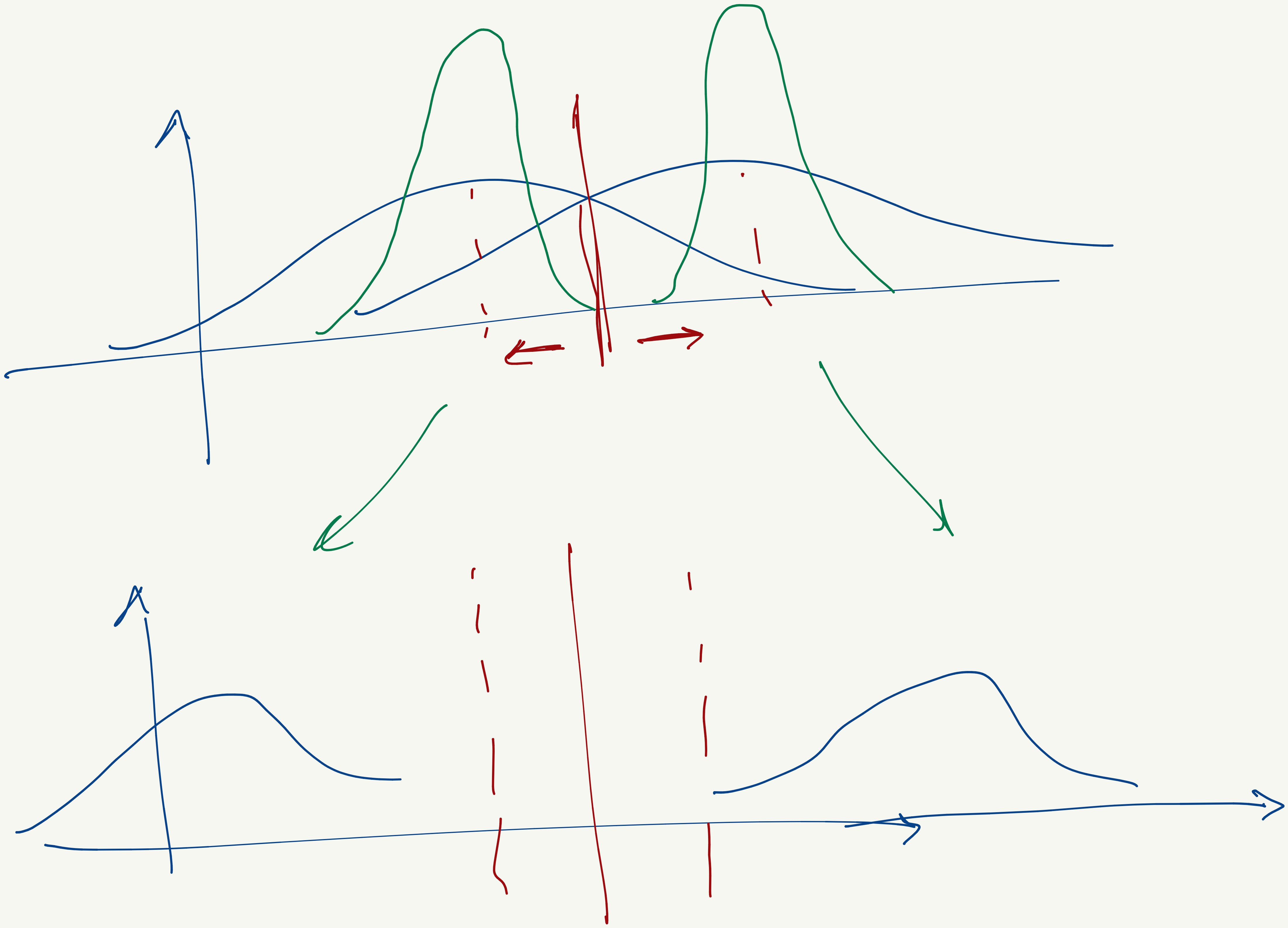
$$x\cdot\left(\cancel{\frac{\mu_1-\mu_2}{\sigma}}\right) = \frac{(\mu_1+\mu_2)}{2}\cdot\cancel{\frac{(\mu_1-\mu_2)}{\sigma^2}} + \log\left(\frac{\pi_2}{\pi_1}\right)\cdot\frac{\sigma^2}{\mu_1-\mu_2}$$

$$x = \frac{\mu_1+\mu_2}{2} + \log\left(\frac{\pi_2}{\pi_1}\right)\cdot\boxed{\frac{\sigma^2}{\mu_1-\mu_2}}$$

$$\pi_1 = \pi_2 = \frac{1}{2}$$

$$x = \frac{\pi_1+\pi_2}{2}$$





PARÁMETROS A ESTIMAR

K clases y n_k individuos en la clase k, y n el total

Training set

$$\hat{\pi}_k = \frac{n_k}{n}$$

$$\hat{\mu}_k = \frac{1}{n_k} \cdot \sum_{i: y_i = k} x_i$$ ← estimo promediando solamente dentro de la clase

$$\hat{\sigma}^2 = \frac{1}{n-K} \sum_{k=1}^{K} \sum_{y_i = k} (x_i - \hat{\mu}_k)^2$$

$$= \sum_{k=1}^{K} \frac{n_k - 1}{n-K} \cdot \hat{\sigma}_k^2$$

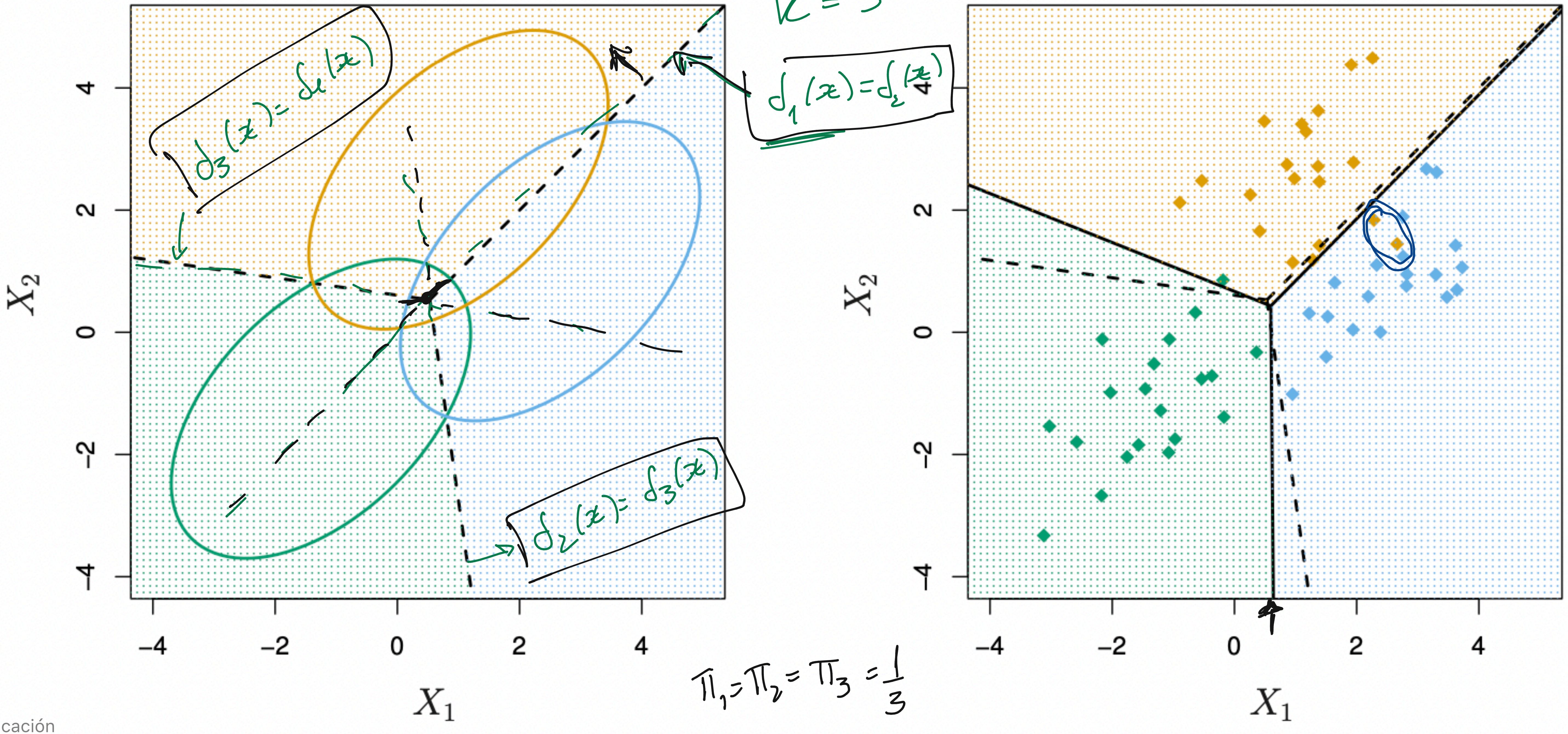
donde $\hat{\sigma}_k^2 = \frac{1}{n_k - 1} \cdot \sum_{i: y_i = k} (x_i - \hat{\mu}_k)^2$

↑ estimación de la varianza intra clase.

$\underline{p > 2}$

$$\varphi(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

$$\delta_k(x) = x^T \cdot \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \underline{\log \pi_k}$$



$p = 2$
$K = 3$

$\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$

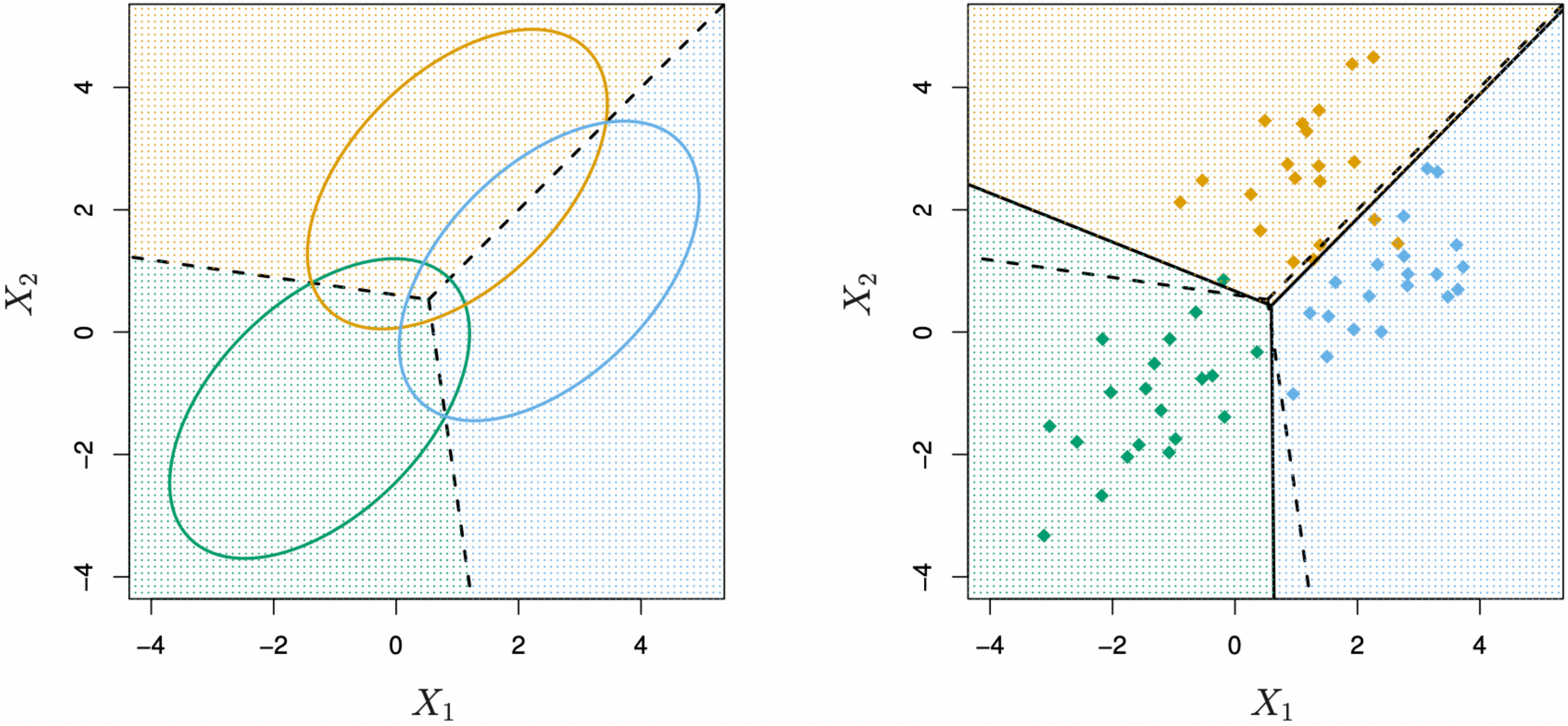
De los $\delta_k(x)$ podemos volver a las probabilidades.

$$P(Y=k \mid X=x) = \frac{\pi_k \cdot \varphi_k(x)}{\sum_{\ell=1}^{K} \pi_\ell \cdot \varphi_\ell(x)}$$

$$\delta_k(x) = \log(\pi_k \cdot \varphi_k(x))$$

$$\pi_k \cdot \varphi_k(x) = e^{\delta_k(x)}$$

$$= \frac{e^{\delta_k(x)}}{\sum_{\ell=1}^{K} \delta_\ell(x)}$$



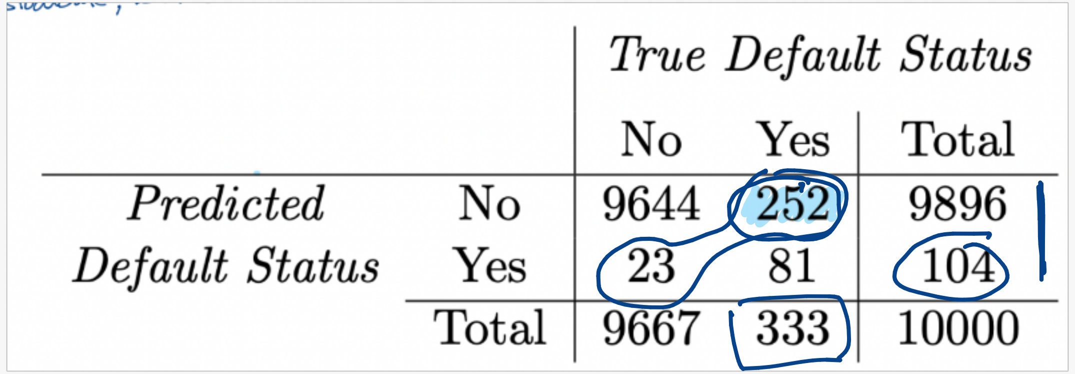
4
2
0
−2
−4
X_2
−4
−2
0
2
4
X_1
4
2
0
−2
−4
X_2
−4
−2
0
2
4
X_1

Default

Estudiante, Balance

LDA on Credit Data

		True Default Status		
		No	Yes	Total
Predicted Default Status	No	9644	252	9896
	Yes	23	81	104
	Total	9667	333	10000



		True Default Status		
		No	Yes	Total
Predicted Default Status	No	9644	252	9896
	Yes	23	81	104
	Total	9667	333	10000

Error total de predicción:

$$\frac{23+252}{10,000} = 0.0275$$

$\boxed{2.75\% \text{ de error}}$

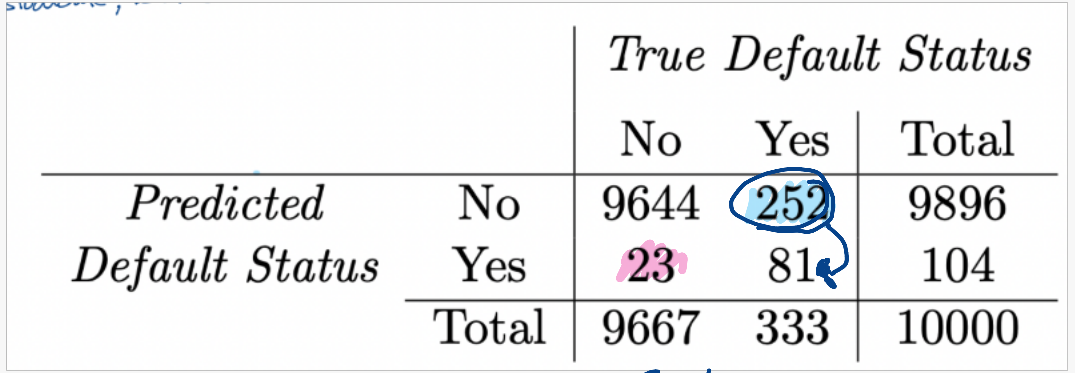
TRAINIG

Overfitting?

Nadie va a entrar en default!

Error total de predicción: $\frac{333}{10.000}$ 3.33% de error

En este caso (default) el error total no me dice mucho.



		True Default Status		
		No	Yes	Total
Predicted	No	9644	252	9896
Default Status	Yes	23	81	104
	Total	9667	333	10000

23 /9667

FALSOS NEGATIVOS

FALSOS POSITIVOS

tasa FN $\frac{252}{333} = \frac{FN}{\text{Positivos}}$

tasa FP $= \frac{23}{9667} = \frac{FP}{\text{Negativos}}$

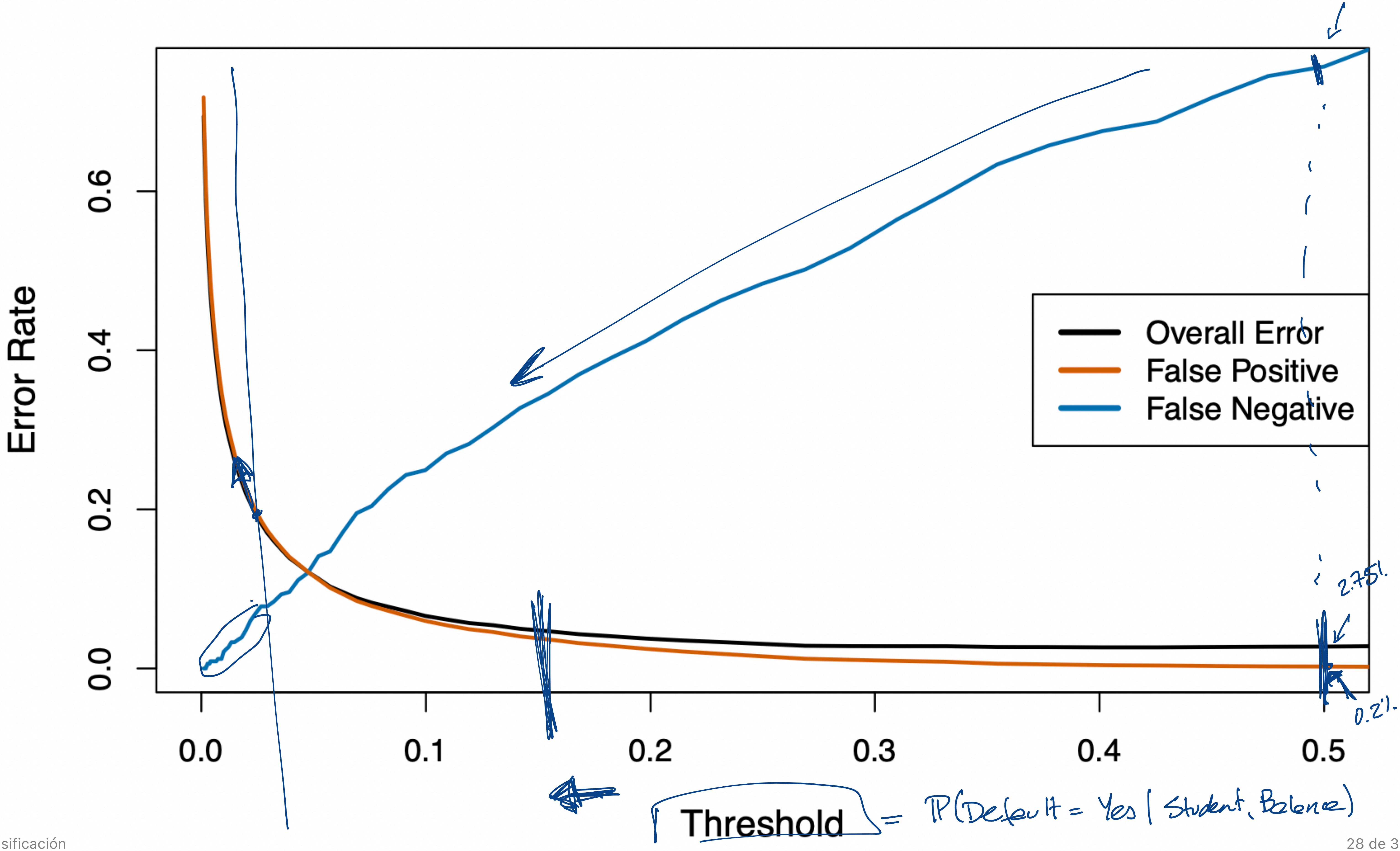
$\approx 0.2\%$

$$\delta_Y(x) > \delta_N(x)$$

$$\mathbb{P}(S=Y \mid X=x) > \mathbb{P}(S=N \mid X=x) \iff \mathbb{P}(S=Y \mid X=x) > 0.5$$

threshold

$$\mathbb{P}(\text{Default} = \text{Yes} \mid \text{Student}, \text{Balance}) > \boxed{0.5}$$



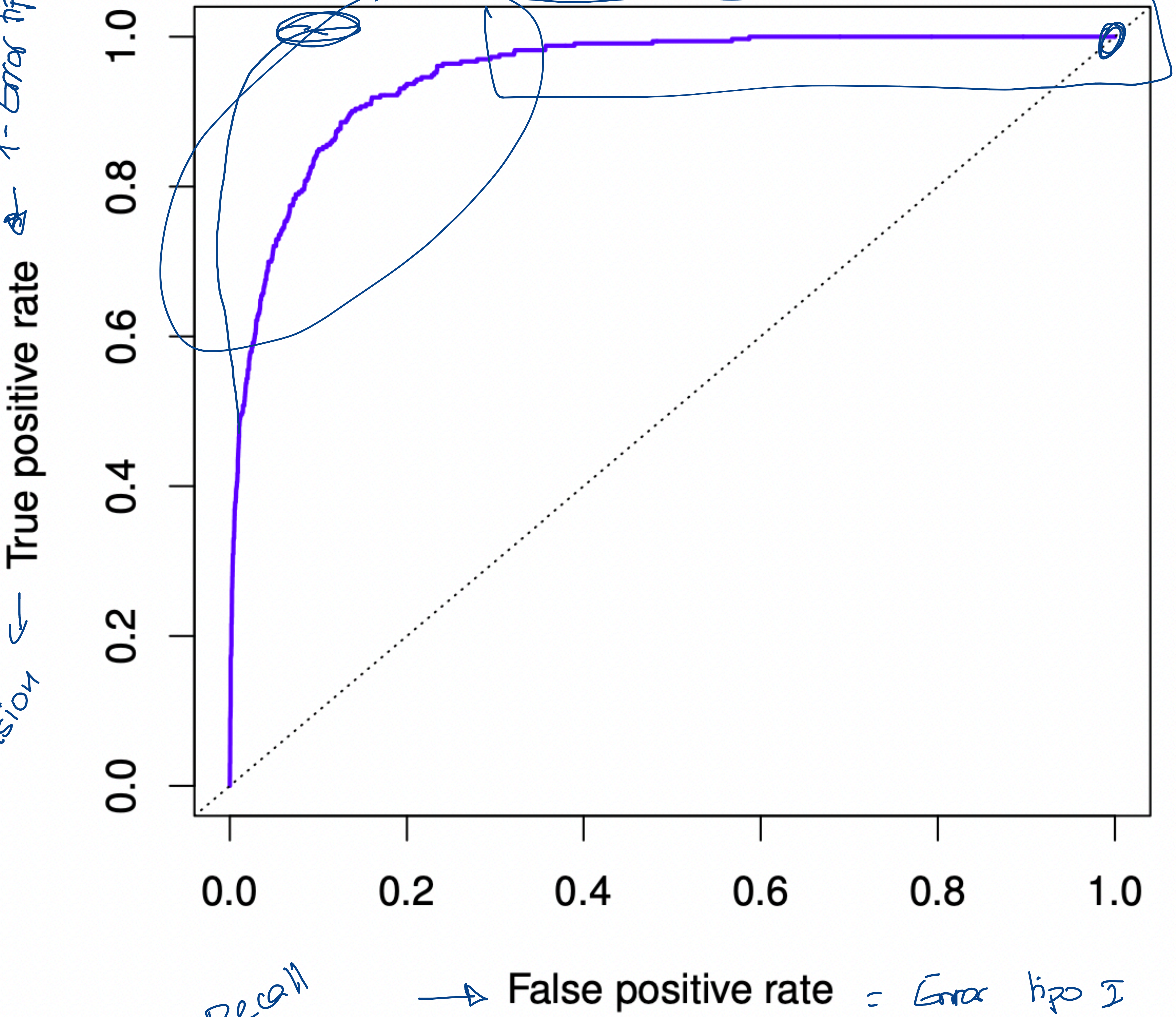
2.75%

0.2%

= $\mathbb{P}$(Default = Yes | Student, Balance)

ROC Curve

Receiver Operating Characteristics



1 - Error tipo II ←

Precision ←

Recall

→ False positive rate = Error tipo I

DISCRIMINANTE CUADRÁTICO

$$\varphi_k(x) \sim N(\mu_k, \Sigma_k)$$

$$\delta_k(x) = -\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k) + \log \pi_k - \frac{1}{2}\log|\Sigma_k|$$

$\hat{\pi}_k$

$\hat{\mu}_k$

$\sigma^2 \rightarrow \sigma_k^2$

Σ_k