Bloque 2
Aprendizaje Automático

Tema 2:
Aprendizaje Supervisado:

Funciones discriminantes lineales.
Perceptrón.

Tema 2.1- Funciones discriminantes lineales. Perceptron.

1. Representación de espacios vectoriales.
2. Funciones lineales.
3. Clasificadores lineales.
4. Fronteras y regiones de decisión.
5. Clasificadores equivalentes.
6. Algoritmo del Perceptron.
7. Estimación empírica del error.

Bibliografía

- R.O. Duda, D.G. Stork, P.E. Hart. Pattern Classification. Wiley, 2001.
- Kevin P. Murphy. Probabilistic Machine Learning: An Introduction. MIT Press, 2022

1. Representación de espacios vectoriales

Tenemos un conjunto de **N** muestras (observaciones, datos, objetos)

Consideramos un conjunto de **características** para las muestras; por ejemplo:

1. climatología/luminosidad/seguridad en el dataset de viajes del tema 1
2. tamaño/forma/color de ejemplos de imágenes de animales
3. En un dataset de flores -> longitud del pétalo, anchura del pétalo, longitud del sépalo, anchura del sépalo

REPRESENTACIÓN VECTORIAL DE LAS MUESTRAS

VECTOR DE CARACTERÍSTICAS

$$\begin{array}{l} \text{long. pétalo}\\ \text{anch. pétalo}\\ \text{long. sépalo}\\ \text{anch. sépalo} \end{array} \quad \underset{\text{muestra 1}}{\begin{pmatrix}5.1\\3.4\\1.4\\0.2\end{pmatrix}} \quad \underset{\text{muestra 2}}{\begin{pmatrix}4.9\\3.0\\1.4\\0.6\end{pmatrix}} \quad \underset{\text{muestra 3}}{\begin{pmatrix}7.0\\3.2\\4.7\\1.4\end{pmatrix}} \quad \underset{\text{muestra 4}}{\begin{pmatrix}3.7\\3.3\\6.0\\2.5\end{pmatrix}} \quad \underset{\text{muestra 5}}{\begin{pmatrix}6.5\\2.4\\5.3\\1.8\end{pmatrix}}$$

N=5 muestras   4 dimensiones ➔ 4 características ➔ $\mathbb{R}^4$

1. Representación de espacios vectoriales

$$\begin{array}{ccccc}
\begin{pmatrix}5.1\\3.4\\1.4\\0.2\end{pmatrix} & \begin{pmatrix}4.9\\3.0\\1.4\\0.6\end{pmatrix} & \begin{pmatrix}7.0\\3.2\\4.7\\1.4\end{pmatrix} & \begin{pmatrix}3.7\\3.3\\6.0\\2.5\end{pmatrix} & \begin{pmatrix}6.5\\2.4\\5.3\\1.8\end{pmatrix} \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\text{clavel} & \text{clavel} & \text{tulipán} & \text{rosa} & \text{tulipán}
\end{array} \qquad \text{clase}$$

N=5 muestras

En un espacio de representación de 4 dimensiones $\mathbb{R}^4$

Para un problema de 3 clases

1. Representación de espacios vectoriales

¿Cuál es nuestro objetivo?

$$
\begin{array}{ccccc}
\begin{pmatrix}5.1\\3.4\\1.4\\0.2\end{pmatrix} & \begin{pmatrix}4.9\\3.0\\1.4\\0.6\end{pmatrix} & \begin{pmatrix}7.0\\3.2\\4.7\\1.4\end{pmatrix} & \begin{pmatrix}3.7\\3.3\\6.0\\2.5\end{pmatrix} & \begin{pmatrix}6.5\\2.4\\5.3\\1.8\end{pmatrix} \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\text{clavel} & \text{clavel} & \text{tulipán} & \text{rosa} & \text{tulipán}
\end{array}
$$

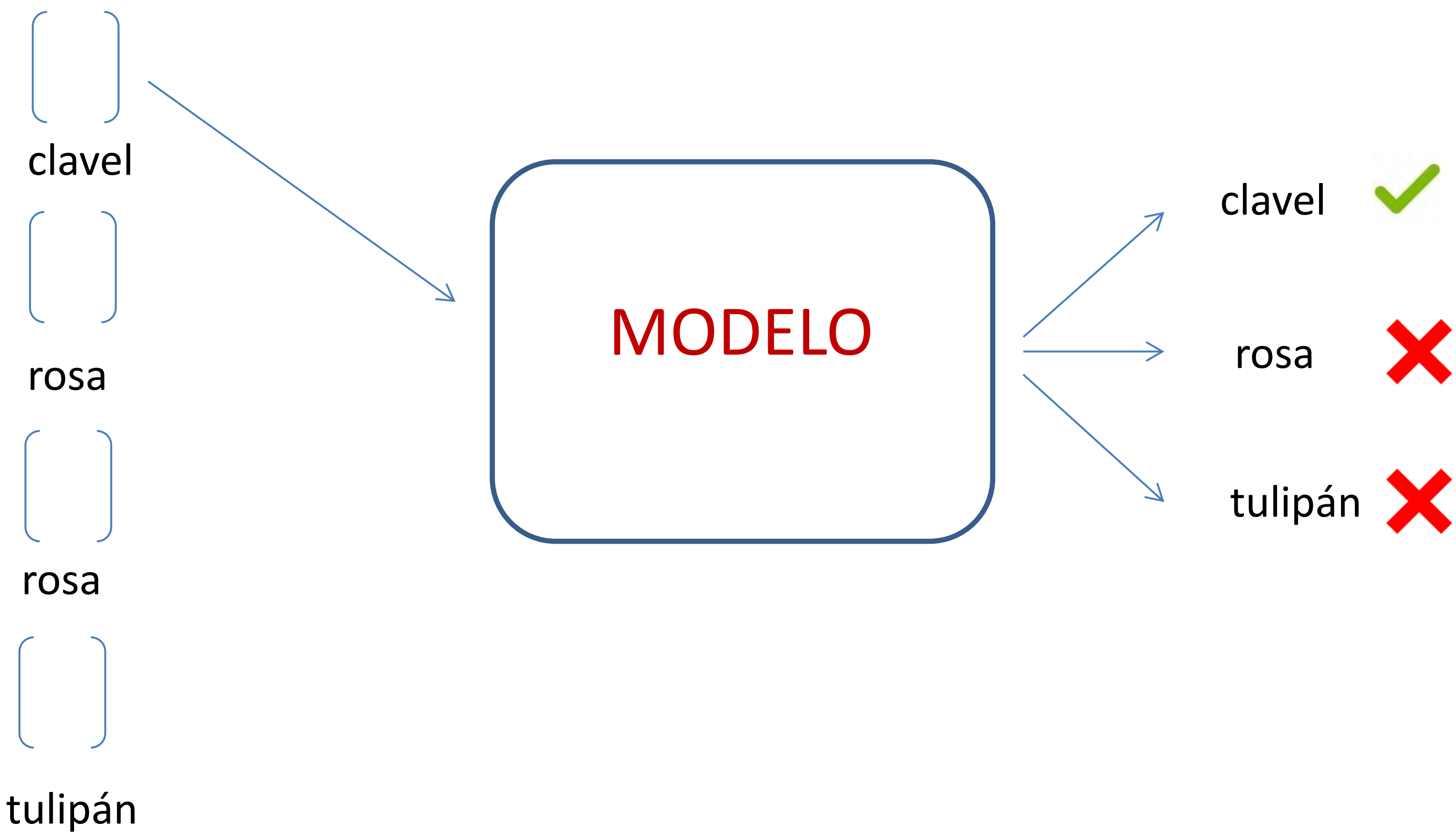
Algoritmo de Aprendizaje

$\Downarrow$

MODELO (matemático)

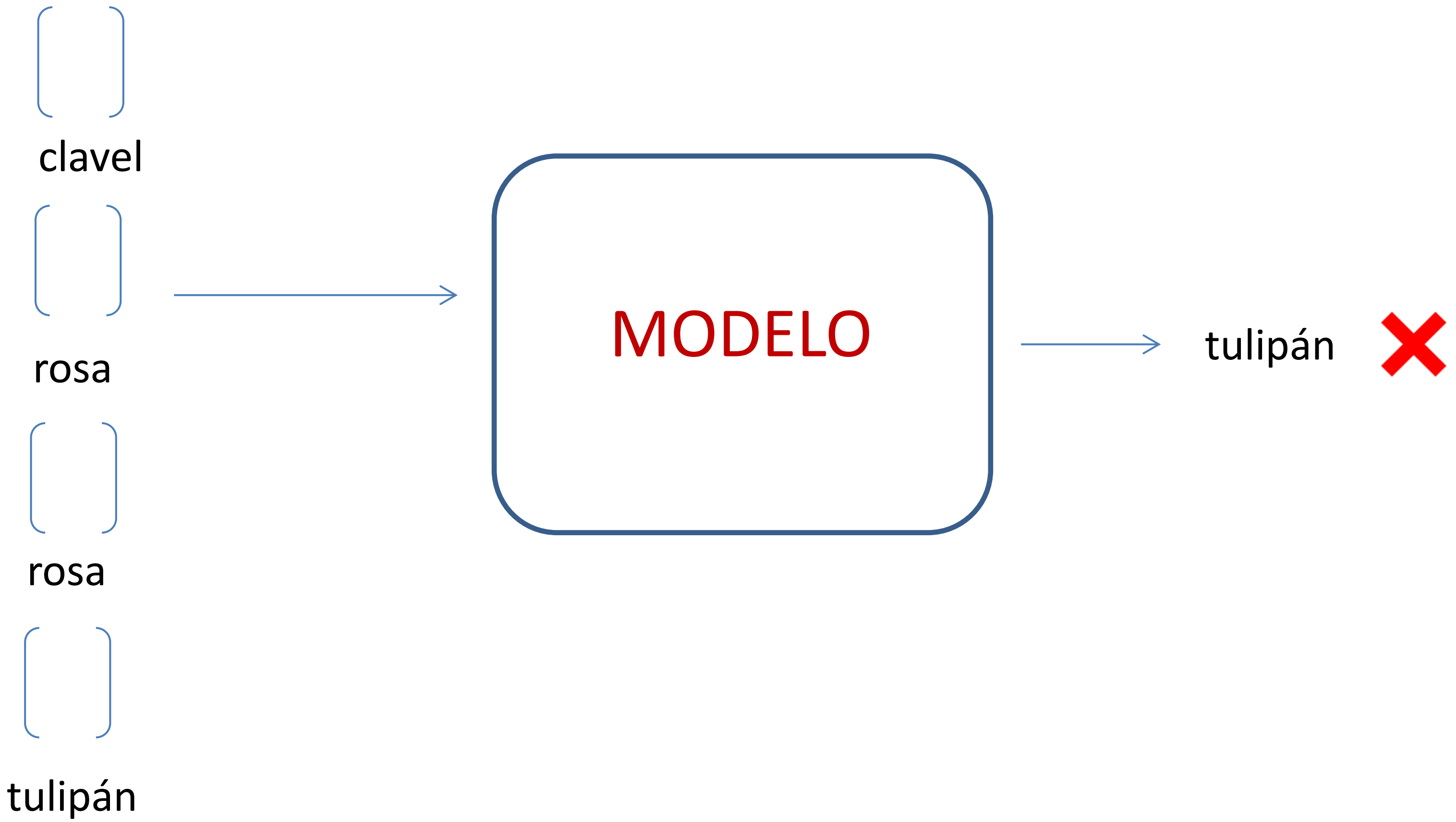
ENTRENAMIENTO

1. Representación de espacios vectoriales



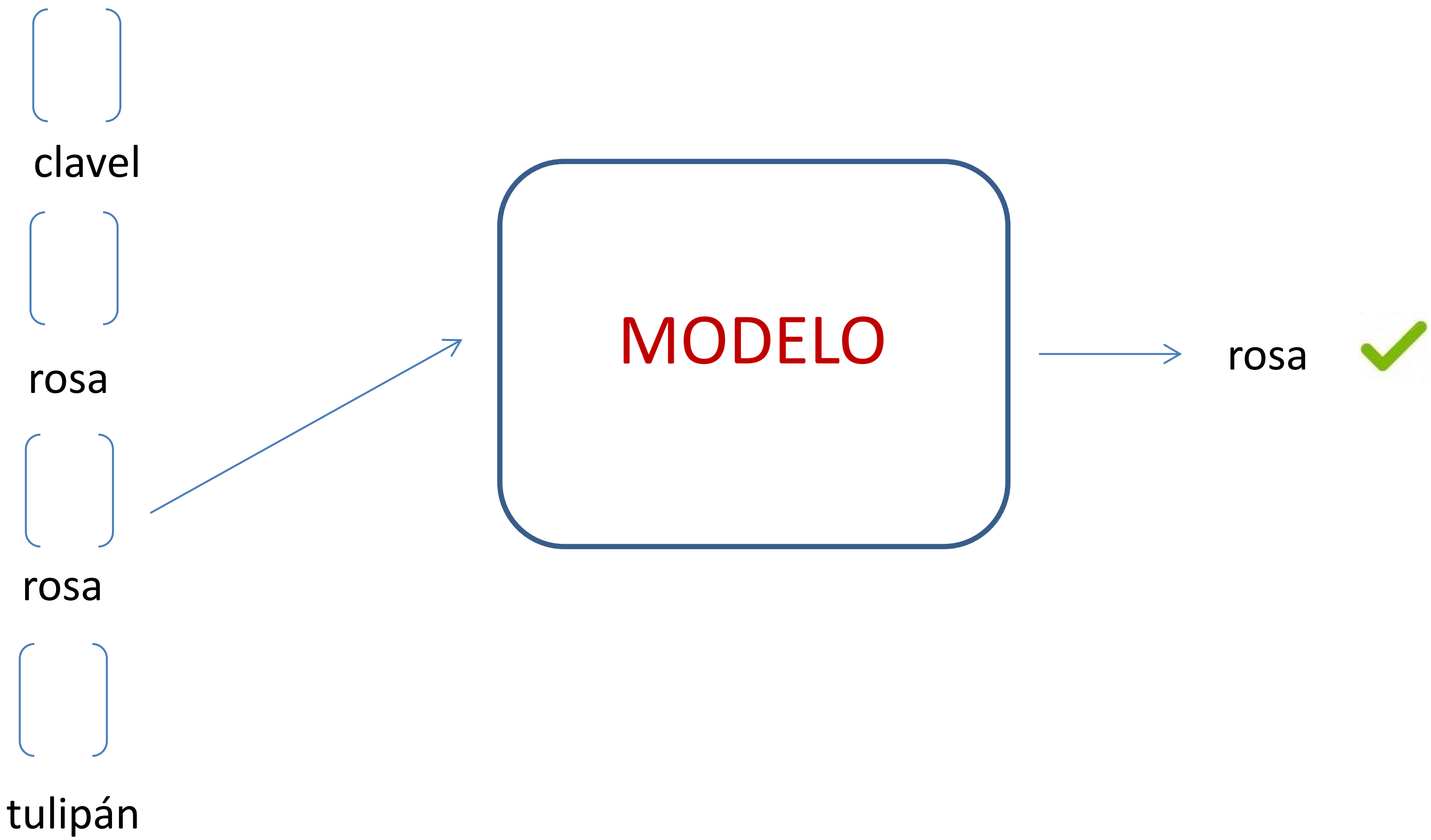
nuevas muestras
(conjunto de test)

1. Representación de espacios vectoriales



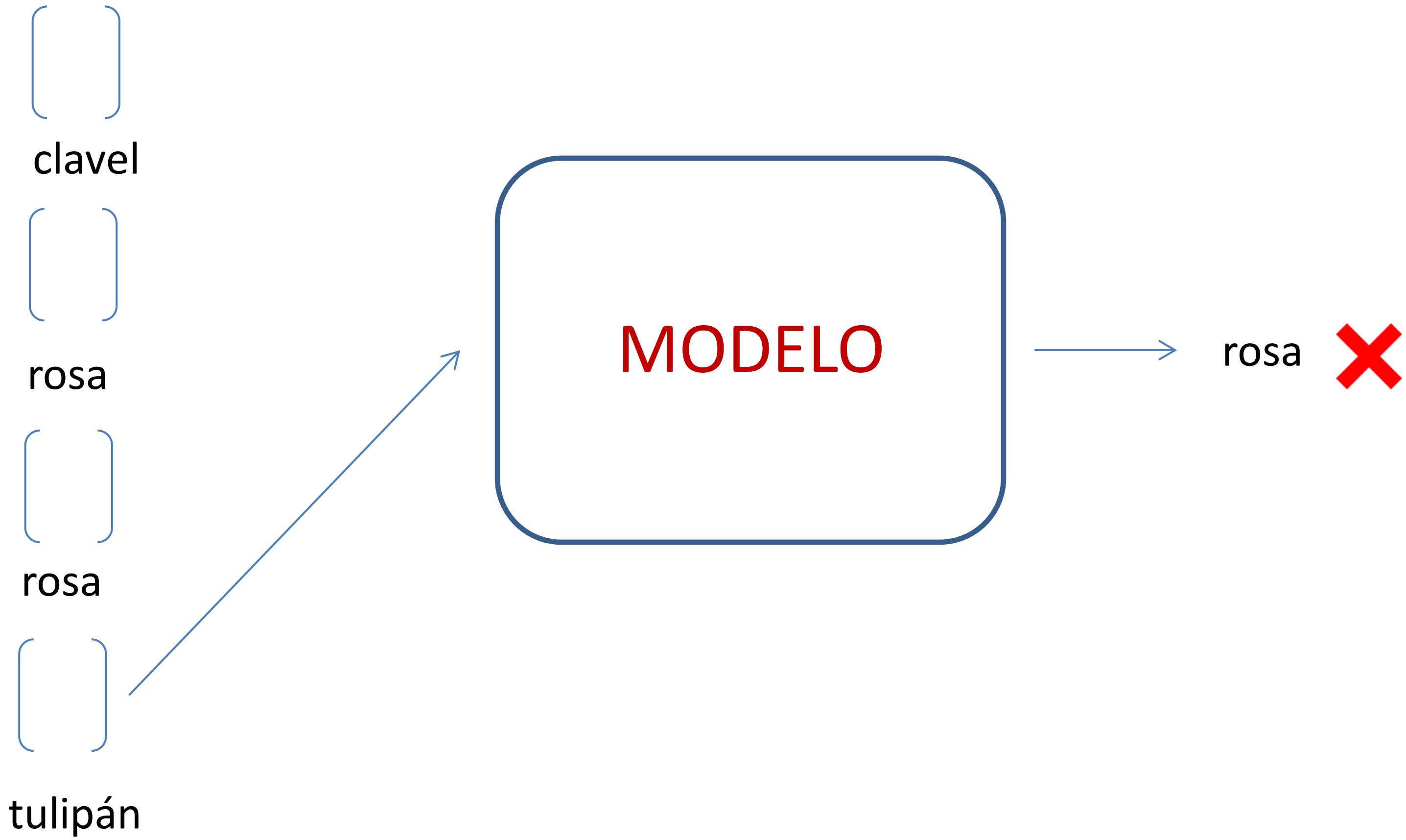
nuevas muestras
(conjunto de test)

1. Representación de espacios vectoriales



nuevas muestras
(conjunto de test)

1. Representación de espacios vectoriales



nuevas muestras
(conjunto de test)

1. Representación de espacios vectoriales

REPRESENTACIÓN VECTORIAL DEL ESPACIO MUESTRAL

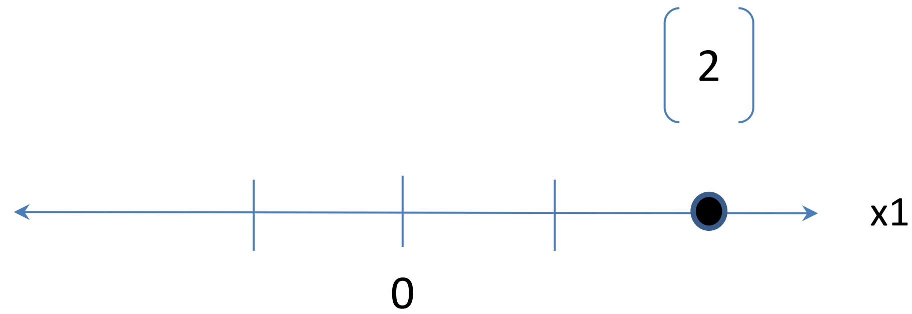
$$\begin{pmatrix} x1 \end{pmatrix} \quad \begin{pmatrix} x1 \\ x2 \end{pmatrix} \quad \begin{pmatrix} x1 \\ X2 \\ x3 \end{pmatrix} \quad \begin{pmatrix} x1 \\ X2 \\ X3 \\ x4 \end{pmatrix} \quad \ldots$$

$$\mathbb{R} \quad \mathbb{R}^2 \quad \mathbb{R}^3 \quad \mathbb{R}^4$$

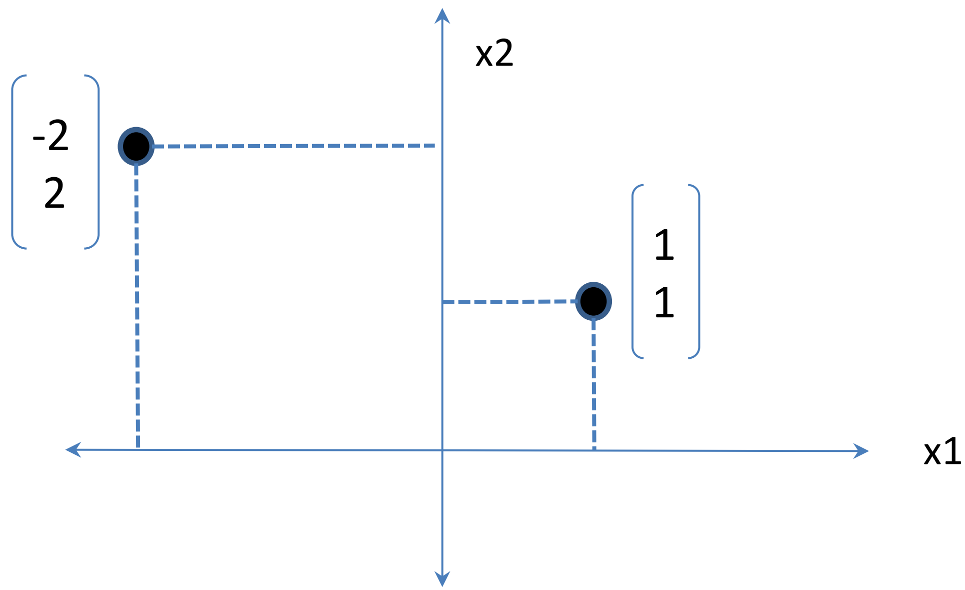
1. Representación de espacios vectoriales

REPRESENTACIÓN GRÁFICA DE ESPACIOS VECTORIALES

$\begin{pmatrix} x1 \end{pmatrix}$ $\qquad \mathbb{R}$



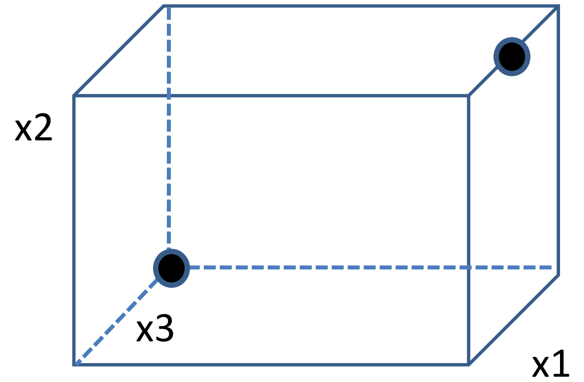
$\begin{pmatrix} x1 \\ x2 \end{pmatrix}$ $\qquad \mathbb{R}^2$



1. Representación de espacios vectoriales

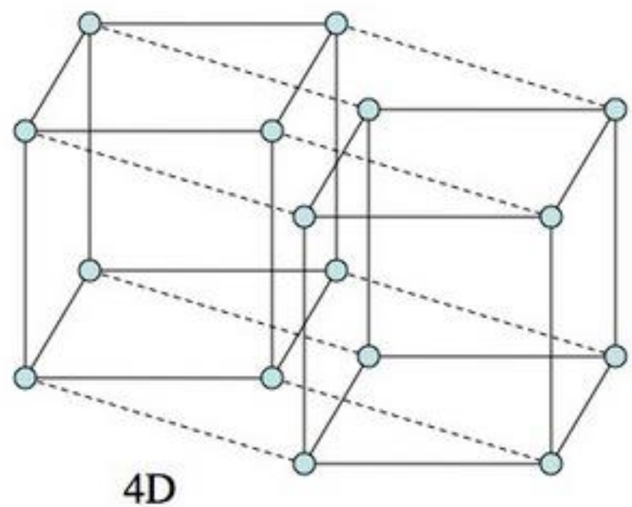
REPRESENTATCIÓN GRÁFICA DE ESPACIOS VECTORIALES

$$\begin{pmatrix} x1 \\ x2 \\ x3 \end{pmatrix} \qquad \mathbb{R}^3$$



$$\begin{pmatrix} 2.5 \\ 1.7 \\ 1.3 \end{pmatrix}$$

$$\begin{pmatrix} x1 \\ X2 \\ X3 \\ x4 \end{pmatrix} \qquad \mathbb{R}^4$$



1. Representación de espacios vectoriales

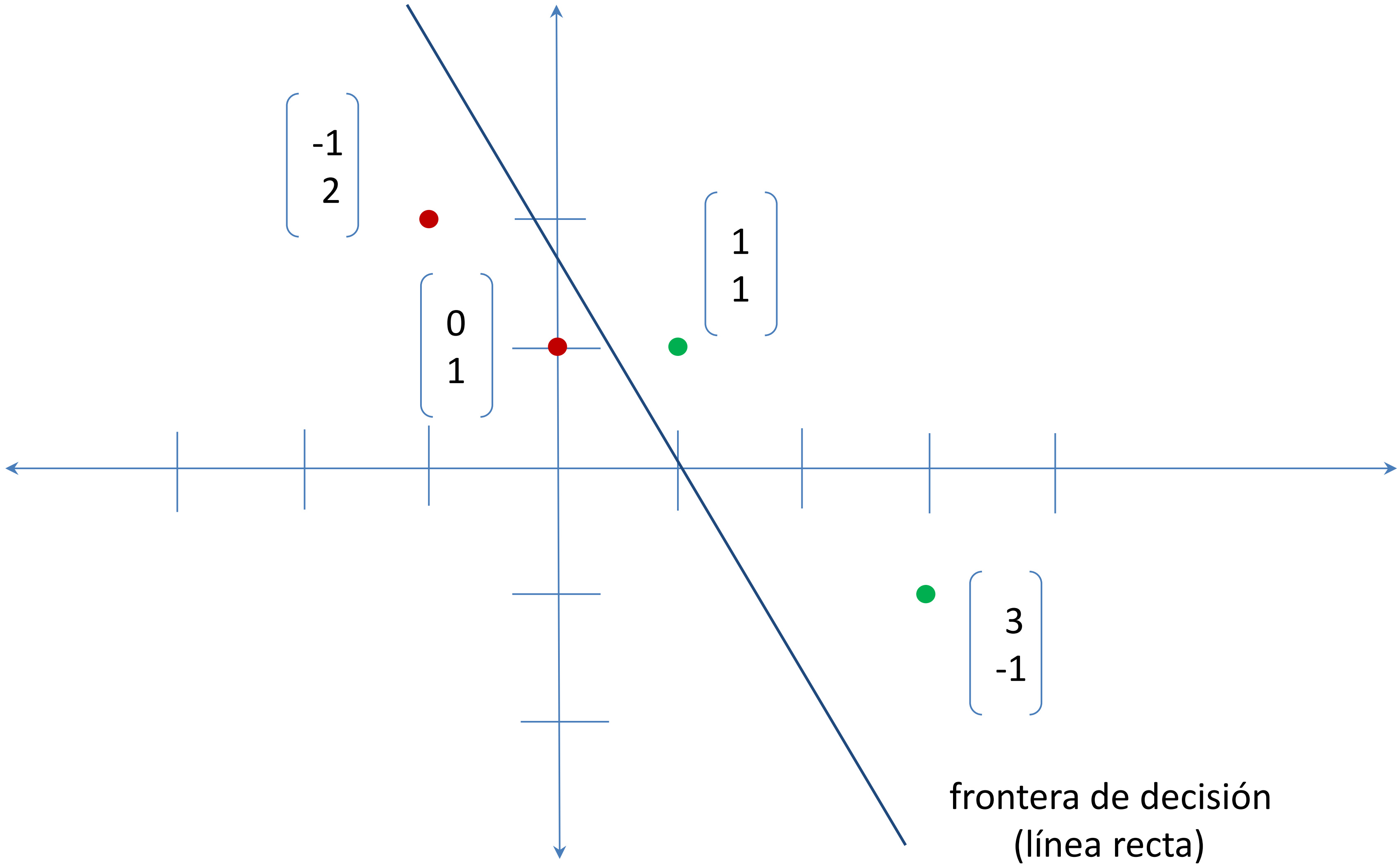
Ejemplo en un espacio de 2 dimensiones $\mathbb{R}^2$

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \begin{bmatrix} -1 \\ 2 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

A	A	B	B

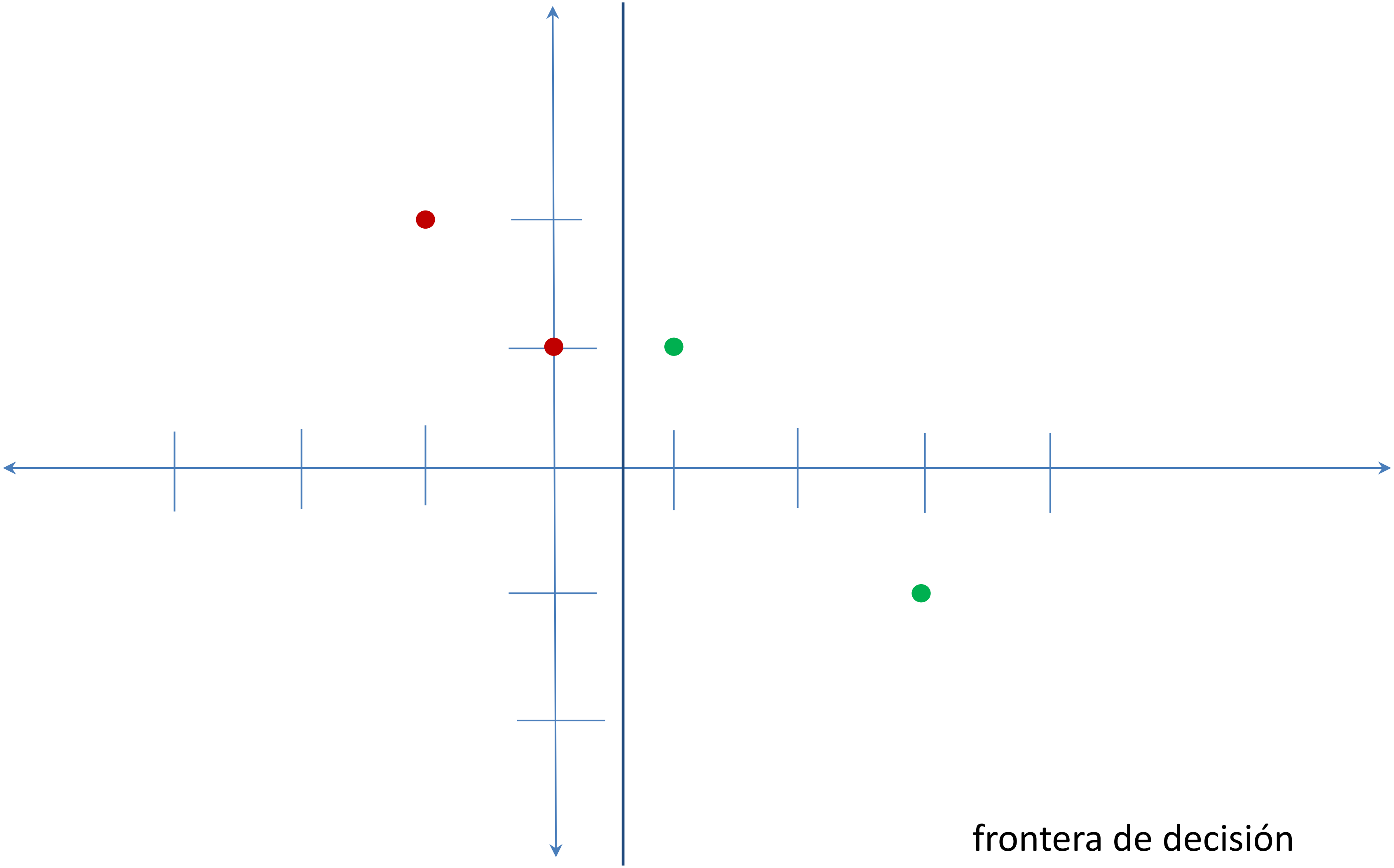
Clase/etiqueta de las muestras

1. Representación de espacios vectoriales



Las dos clases son linealmente separables

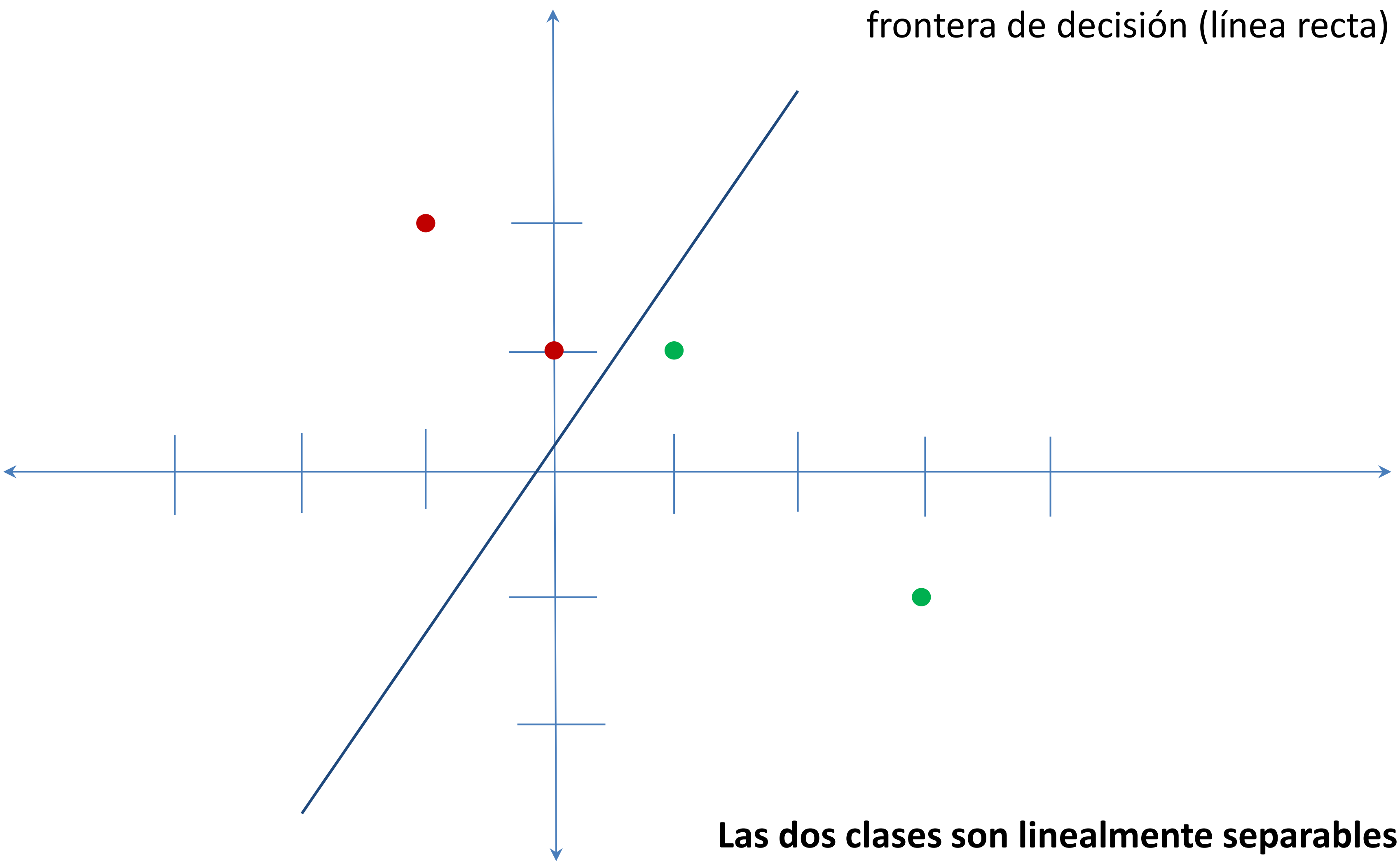
1. Representación de espacios vectoriales



frontera de decisión
(línea recta)

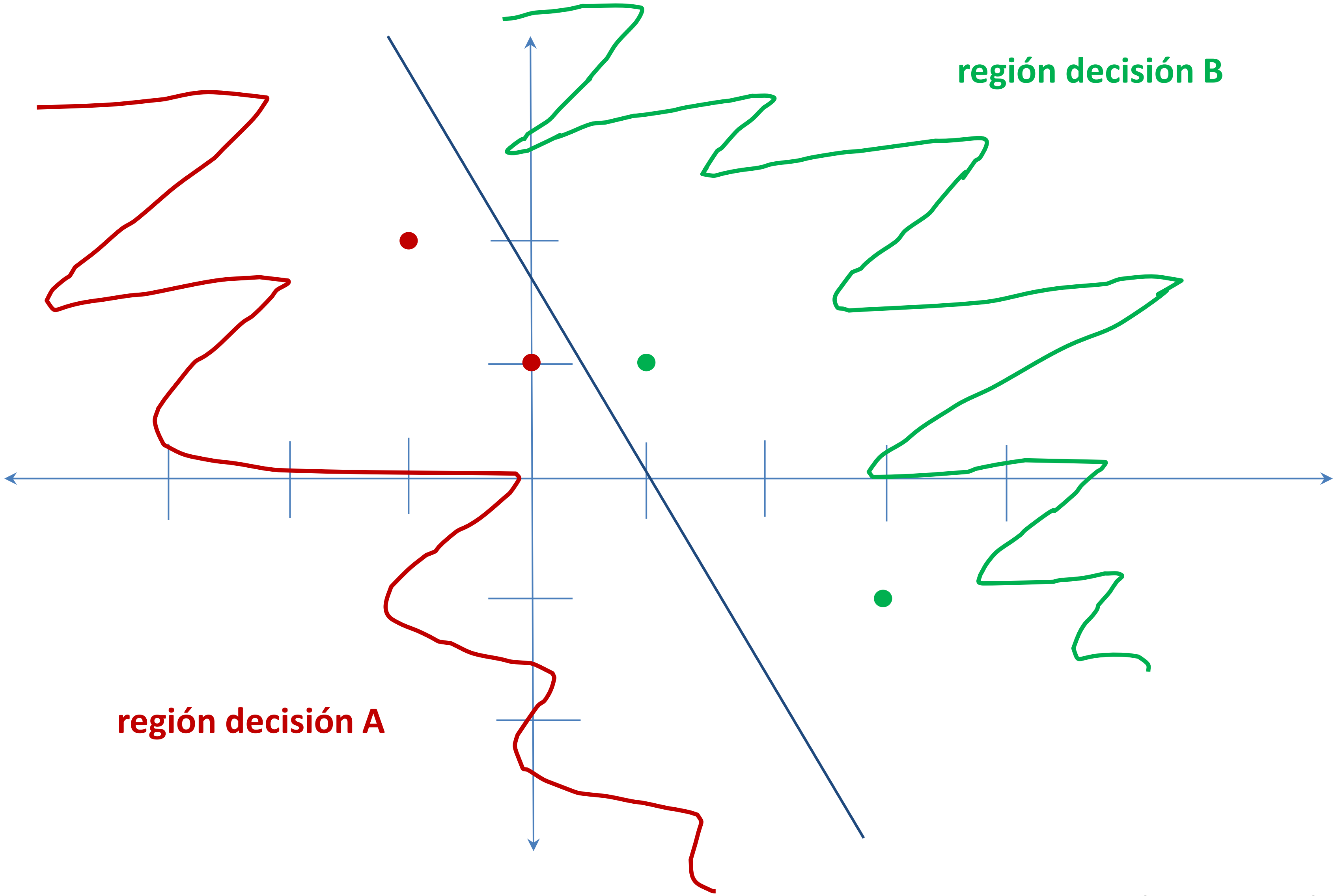
Las dos clases son linealmente separables

1. Representación de espacios vectoriales



Las dos clases son linealmente separables

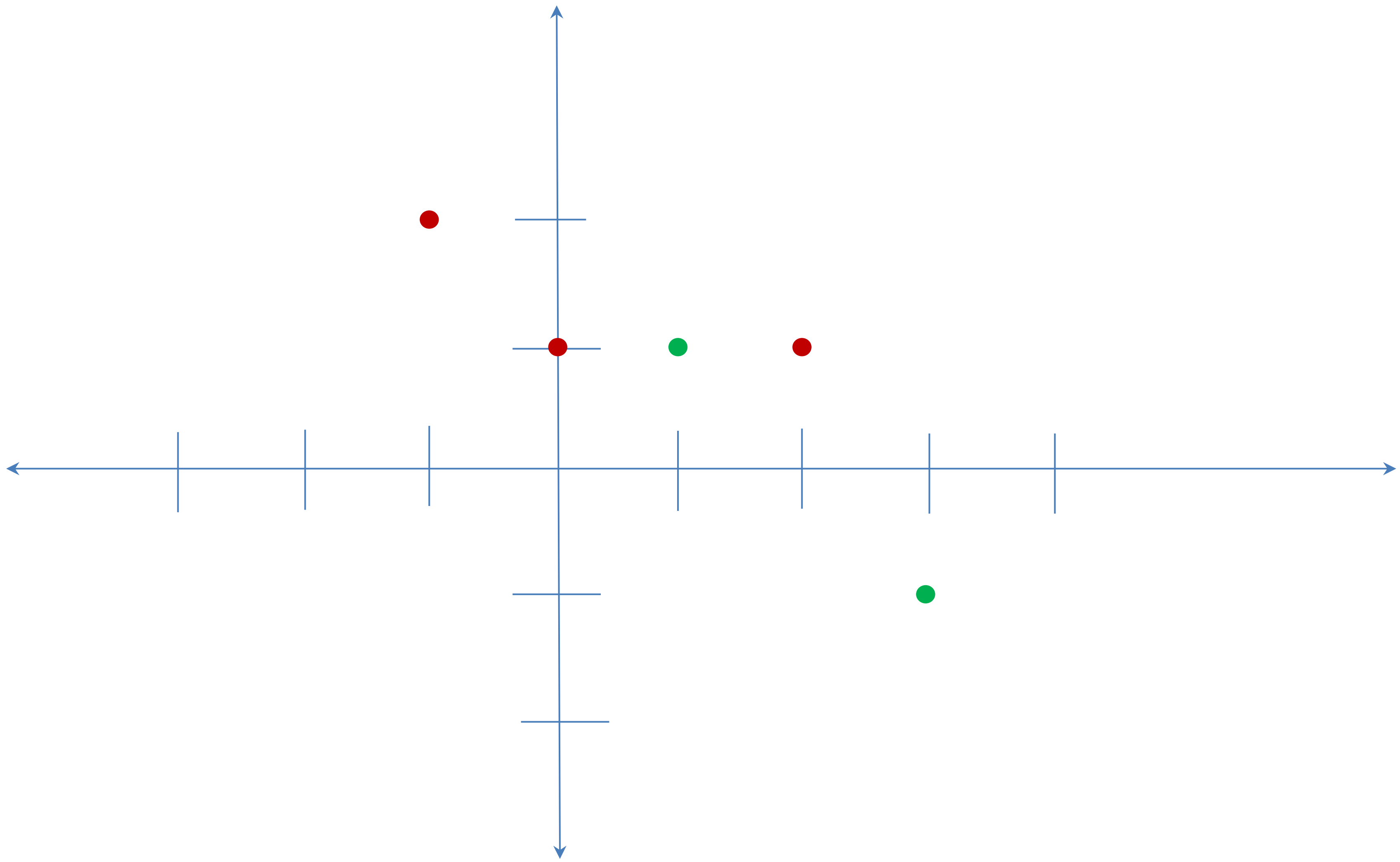
1. Representación de espacios vectoriales



Las dos clases son linealmente separables

frontera de decisión (línea recta)

1. Representación de espacios vectoriales



Las dos clases no son linealmente separables

2. Funciones lineales discriminantes

pesos

$$y = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$$

variables discriminantes = características de las muestras

término independiente

$$\begin{pmatrix} w0 & w1 & w2 & \ldots & wn \end{pmatrix}$$

vector pesos
(función lineal)

vector de características
de las muestras

$$\begin{pmatrix} x1 \\ x2 \\ . \\ . \\ xn \end{pmatrix}$$

3. Clasificadores lineales

CLASIFICADOR LINEAL:

una función lineal para **cada clase** de mi espacio muestral

La idea es que los valores de una variable discriminante $\mathbf{x_i}$ para las muestras de una clase **A** estén próximos entre sí y alejados de los valores de $\mathbf{x_i}$ de las muestras de la clase **B** (y cualquier otra clase que no sea A)

Diseñar un algoritmo que aprenda
una función lineal por clase (un clasificador lineal)

➔ aprender los pesos de las funciones lineales
➔ tal que los pesos de las variables discriminen entre clases

EJEMPLO de un clasificador lineal

Muestras **de 2-dimensiones** ($\mathbb{R}^2$) para un problema de **2 clases** (clase A y B)

$$g_A(y) = -1 - 3y_1 + y_2 \longrightarrow (-1 \quad -3 \quad 1)$$

$$g_B(y) = -3 + 3y_1 - y_2 \longrightarrow (-3 \quad 3 \quad -1)$$

Supongamos la siguiente muestra $y = \begin{pmatrix} 0.3 \\ 3.5 \end{pmatrix}$ que sabemos que pertenece a la clase A

¿Cómo clasifica la muestra **y** mi clasificador lineal?

3. Clasificadores lineales

$$g_A(y) = -1 - 3y_1 + y_2$$

$$g_B(y) = -3 + 3y_1 - y_2$$

$$y = \begin{array}{cc} \begin{pmatrix} 0.3 \\ 3.5 \end{pmatrix} & \begin{array}{c} y1 \\ y2 \end{array} \end{array}$$

A

$$g_A(y) = -1 - 3 \cdot 0.3 + 3.5 = 1.6 \qquad g_B(y) = -3 + 3 \cdot 0.3 - 3.5 = -5.6$$

$$\begin{pmatrix} -1 & -3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0.3 \\ 3.5 \end{pmatrix} = 1.6 \qquad \begin{pmatrix} -3 & 3 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0.3 \\ 3.5 \end{pmatrix} = -5.6$$

Mi modelo clasifica la muestra **y** como clase A

3. Clasificadores lineales

Si $g_A(y) > g_B(y)$ entonces la muestra se clasifica como clase A

Si $g_A(y) < g_B(y)$ entonces la muestra se clasifica como clase B

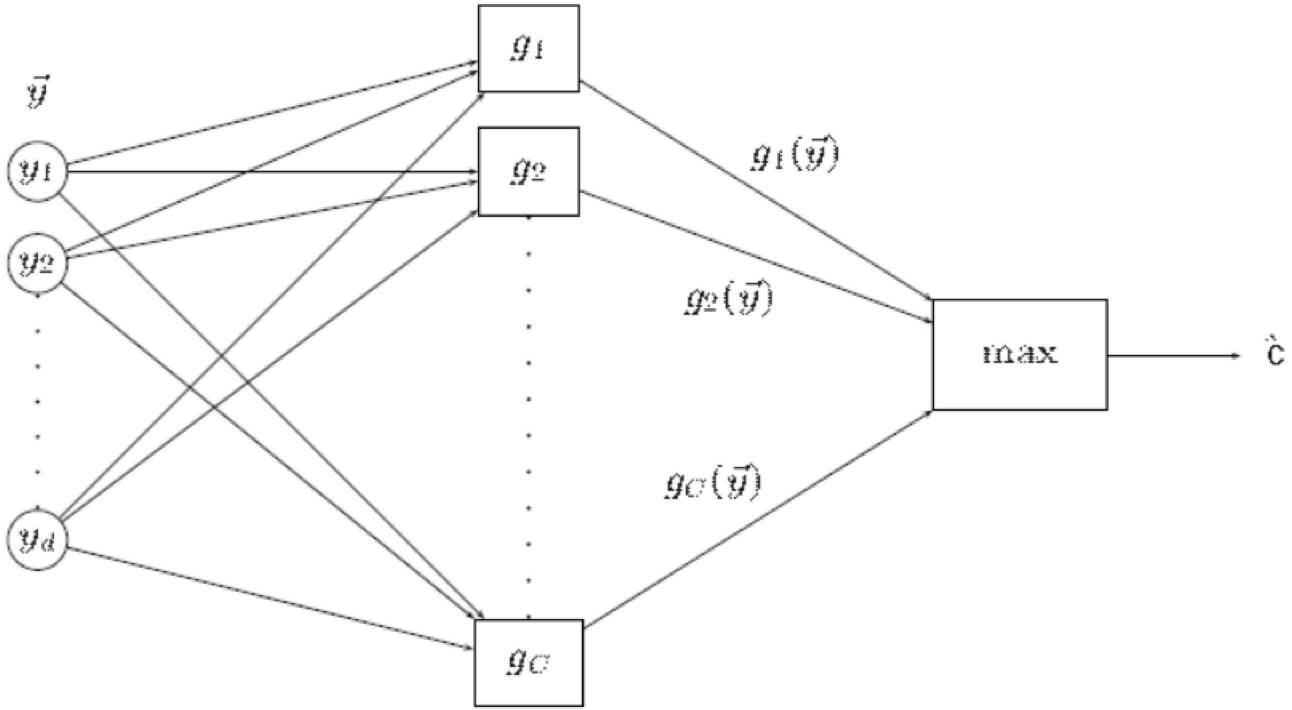
Si $g_A(y) = g_B(y)$ entonces la muestra está en la frontera que separa las dos clases

En el caso de más de dos clases, la muestra se clasifica como la clase cuya función lineal devuelve el valor más alto

3. Clasificadores lineales

Todo clasificador G en C clases puede expresarse mediante C *funciones discriminantes* $g_c : E \to \mathbb{R}, \ 1 \leq c \leq C$, y la correspondiente *regla de clasificación*:

$$G = (g_1, g_2, \ldots, g_C), \quad \hat{c} = G(y) \equiv \underset{1 \leq c \leq C}{\text{argmax}} \ g_c(y)$$



¿cómo podemos calcular las regiones de decisión que define un clasificador?

Calcular la frontera de decisión entre dos clases

$$g_A(y) = -1 - 3y_1 + y_2$$

$$g_B(y) = -3 + 3y_1 - y_2$$

1) Igualamos las dos funciones discriminantes: esto resulta en una ecuación lineal que define la **frontera de decisión**

2) Dibujamos la ecuación lineal resultante

3) Determinamos las **regiones de decisión**

4. Fronteras y regiones de decisión

Vamos a calcular **la frontera de decisión** y **regiones de decisión** del **clasificador lineal**

$$g_A(y) = -1 - 3y_1 + y_2$$

$$g_B(y) = -3 + 3y_1 - y_2$$

$$-1 - 3y_1 + y_2 = -3 + 3y_1 - y_2$$

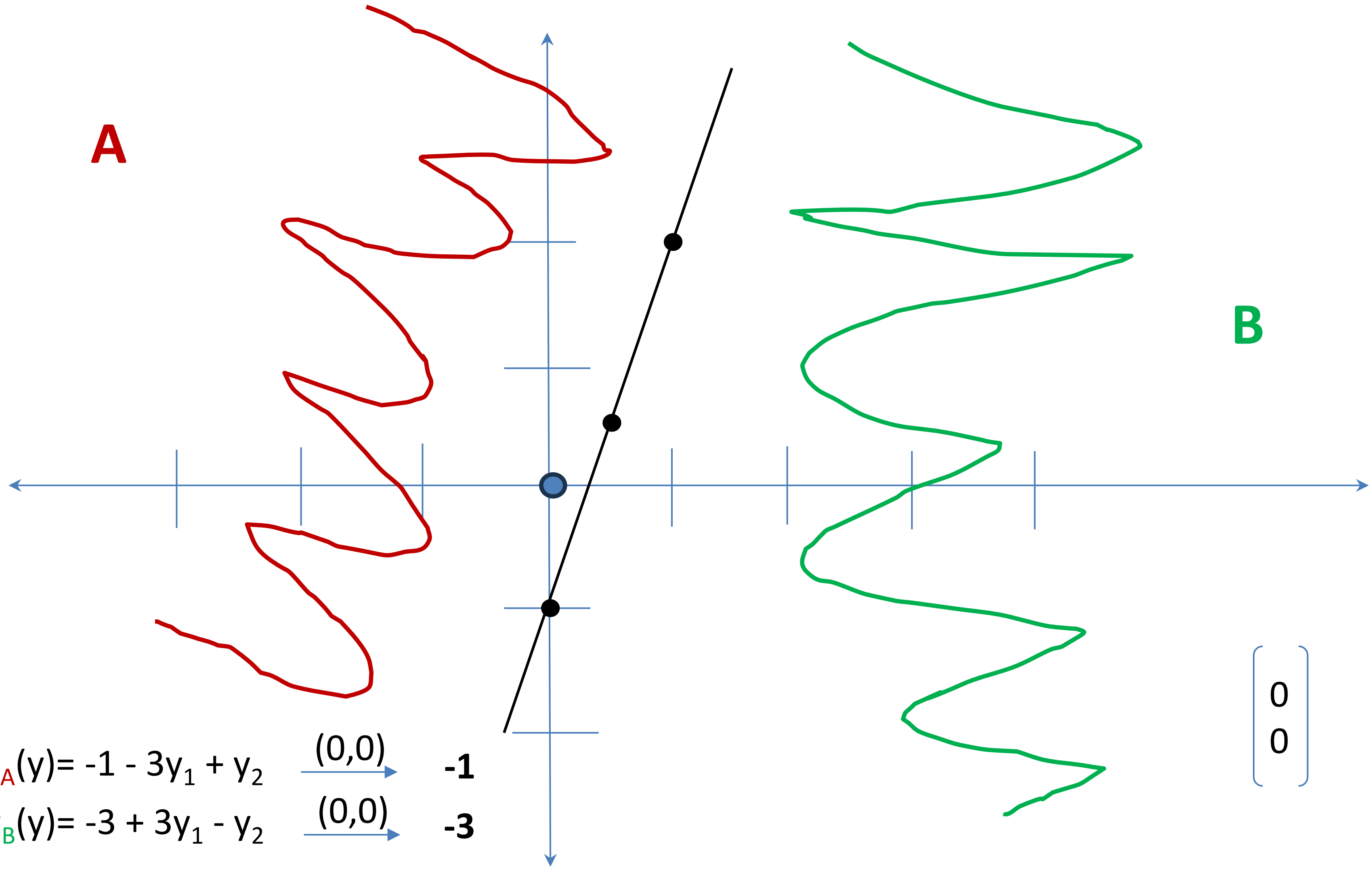
$$2y_2 = 6y_1 - 2$$

$$\boxed{\mathbf{y_2 = 3y_1 - 1}} \longrightarrow$$

ecuación que define la frontera de decisión

4. Fronteras y regiones de decisión

$\mathbf{y_2 = 3y_1 - 1}$ → Ecuación de la frontera de decisión entre A y B



$g_A(y) = -1 - 3y_1 + y_2 \xrightarrow{(0,0)} \mathbf{-1}$

$g_B(y) = -3 + 3y_1 - y_2 \xrightarrow{(0,0)} \mathbf{-3}$

4. Fronteras y regiones de decisión

Un clasificador divide el espacio de representación en C *regiones de decisión*, $R_1, \dots, R_C$:

$$R_j = \{\mathbf{y} \in E : \quad g_j(\mathbf{y}) > g_i(\mathbf{y}) \quad i \neq j,\ 1 \leq i \leq C\}$$

- *Frontera de Decisión entre dos clases* i, j:
  Lugar geométrico de los puntos $\mathbf{y} \in E$ para los que $g_i(\mathbf{y}) = g_j(\mathbf{y})$

  En general son *Hipersuperficies* definidas por las ecuaciones:

  $$g_i(\mathbf{y}) - g_j(\mathbf{y}) = 0 \qquad i \neq j,\ \ 1 \leq i, j \leq C$$

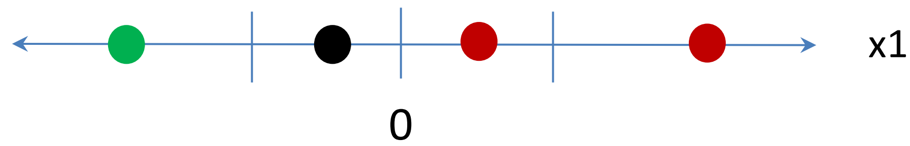
  - Si $E \equiv \mathbb{R}^3$ las fronteras son superficies (ej. *planos*)
  - Si $E \equiv \mathbb{R}^2$ las fronteras son líneas (ej. *rectas*)
  - Si $E \equiv \mathbb{R}$ las fronteras son puntos

- *Frontera de Decisión de una clase* i:
  Lugar geométrico de los puntos $\mathbf{y} \in E$ para los que:

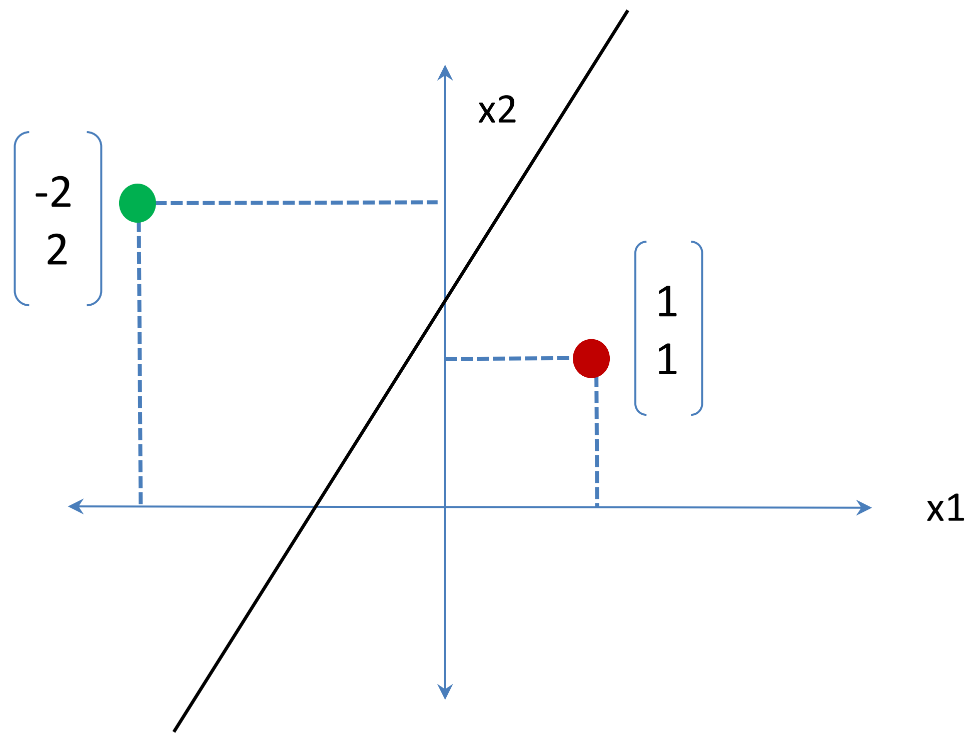
  $$g_i(\mathbf{y}) = \max_{j \neq i} g_j(\mathbf{y})$$

4. Fronteras y regiones de decisión

$\begin{pmatrix} x1 \end{pmatrix}$ $\mathbb{R}$

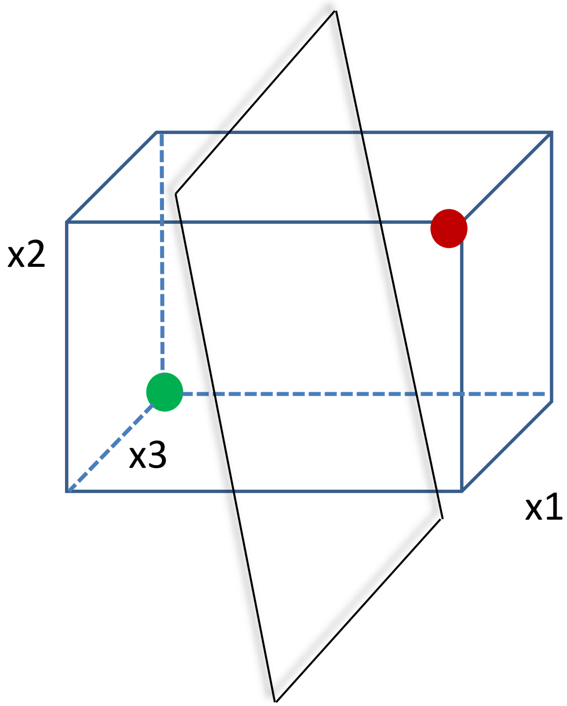


$\begin{pmatrix} x1 \\ x2 \end{pmatrix}$ $\mathbb{R}^2$



4. Fronteras y regiones de decisión

$$\begin{pmatrix} x1 \\ X2 \\ x3 \end{pmatrix} \qquad \mathbb{R}^3$$



$$\begin{pmatrix} 2.5 \\ 1.7 \\ 1.3 \end{pmatrix}$$

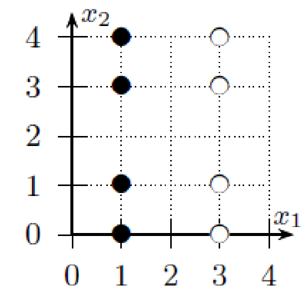
44 ☐ Sea un problema de clasificación en cuatro clases de objetos representados en $\mathbb{R}^3$. Se tiene un clasificador cuyas funciones discriminantes son lineales con vectores de pesos (en notación homogénea):

$$\mathbf{a}_1 = (-2, 1, 2, 0)^t \qquad \mathbf{a}_2 = (0, 2, 2, 0)^t \qquad \mathbf{a}_3 = (1, 1, 1, 0)^t \qquad \mathbf{a}_4 = (3, 0, 0, 2)^t$$

Indica a qué clase se asignarà el objeto $\mathbf{x} = (1, 2, 2)^t$ (*no* en notación homógenea).

A) 1.
B) 2.
C) 3.
D) 4.

18 ☐ En la figura de la derecha se representan cuatro muestras de aprendizaje bidimensionales de 2 clases: ○ y ●. Si nuestro criterio de aprendizaje es la minimización del número de errores de clasificación (sobre las muestras de aprendizaje), elegiremos como vector de pesos de cada una de las clases...



A) $\mathbf{a}_\circ = (3,1,1)^t$ y $\mathbf{a}_\bullet = (1,2,1)^t$

B) $\mathbf{a}_\circ = (1,1,2)^t$ y $\mathbf{a}_\bullet = (3,1,1)^t$

C) $\mathbf{a}_\circ = (3,1,1)^t$ y $\mathbf{a}_\bullet = (1,1,2)^t$

D) $\mathbf{a}_\circ = (1,2,1)^t$ y $\mathbf{a}_\bullet = (3,1,1)^t$

7 ☐ En un problema de clasificación en dos clases se tienen los siguientes puntos en dos dimensiones: $\mathbf{x}_1 = (1,1)^t, \mathbf{x}_2 = (2,2)^t, \mathbf{x}_3 = (2,0)^t$; $\mathbf{x}_1$ y $\mathbf{x}_2$ pertenencen a la clase A y $\mathbf{x}_3$ a la clase B. Teniendo en cuenta que se emplea un clasificador basado en funciones discriminantes lineales con vectores de pesos $\mathbf{w}_A$ y $\mathbf{w}_B$ asociados a las clases A y B respectivamente, indica cuál de las siguientes afirmaciones es *falsa*:

A) Se puede encontrar una función discriminante lineal que clasifique $\mathbf{x}_1$, $\mathbf{x}_2$ y $\mathbf{x}_3$ con error$=2/3$.
B) Los pesos $\mathbf{w}_A = (1,-1,1)^t$ y $\mathbf{w}_B = (1,2,-4)^t$ clasifican $\mathbf{x}_1$, $\mathbf{x}_2$ y $\mathbf{x}_3$ sin error.
C) Los pesos $\mathbf{w}_A = (1,-1,1)^t$ y $\mathbf{w}_B = (1,2,-4)^t$ clasifican $\mathbf{x}_1$, $\mathbf{x}_2$ y $\mathbf{x}_3$ con error$=1/3$.
D) Se puede encontrar una función discriminante lineal que clasifique $\mathbf{x}_1$, $\mathbf{x}_2$ y $\mathbf{x}_3$ con error$=1/3$.

40 ☐ En la figura de la derecha se representan las fronteras de decisión de un clasificador en 3 clases. ¿Cuales de los siguientes vectores de pesos definen dichas fronteras?

A) $\mathbf{w}_1 = (0,0,1)^t \quad \mathbf{w}_2 = (0,1,0)^t$ y $\mathbf{w}_3 = (1,0,0)^t$

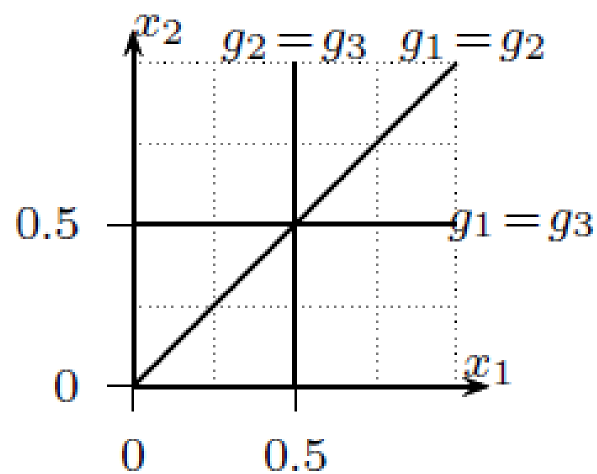
B) $\mathbf{w}_1 = (0,0,1)^t \quad \mathbf{w}_2 = (0,1,0)^t$ y $\mathbf{w}_3 = (0.5,0,0)^t$

C) $\mathbf{w}_1 = (0.5,0,0)^t \quad \mathbf{w}_2 = (0,1,0)^t$ y $\mathbf{w}_3 = (0,0,1)^t$

D) $\mathbf{w}_1 = (0,0,1)^t \quad \mathbf{w}_2 = (1,0,0)^t$ y $\mathbf{w}_3 = (0,1,0)^t$

g1=g2 → x1=x2 o x2=x1 solo puede ser A o B

g2=g3 → x1=0.5 solo puede ser B



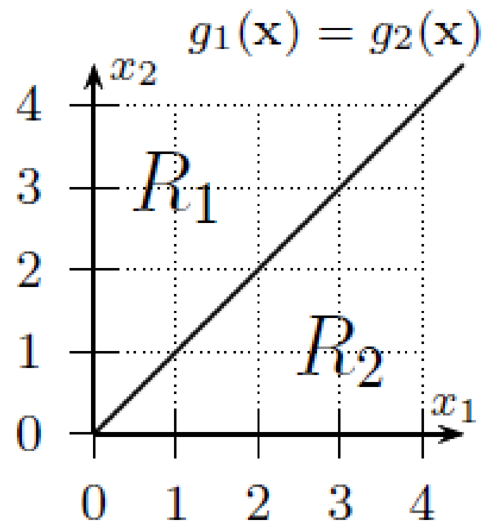
36 $\square$ En la figura se representan frontera y regiones de decisión de un clasificador binario. ¿Cuál de los siguientes pares de vectores de pesos corresponde al clasificador de la figura?

A) $\mathbf{w}_1 = (-1, -1, -2)^t$ y $\mathbf{w}_2 = (-1, -2, -1)^t$

B) $\mathbf{w}_1 = (\ \ 1, -1, -2)^t$ y $\mathbf{w}_2 = (\ \ 0, -2, -1)^t$

C) $\mathbf{w}_1 = (\ \ 1,\ \ 1,\ \ 2)^t$ y $\mathbf{w}_2 = (\ \ 1,\ \ 2,\ \ 1)^t$

D) $\mathbf{w}_1 = (-1,\ \ 1,\ \ 2)^t$ y $\mathbf{w}_2 = (\ \ 0,\ \ 2,\ \ 1)^t$



5. Clasificadores equivalentes

$g_A(y) = \ldots.$ $\qquad$ $g'_A(y) = \ldots.$

$g_B(y) = \ldots.$ $\qquad$ $g'_B(y) = \ldots.$

son equivalentes si:

- definen la misma **frontera de decisión**
- definen las mismas **regiones de decisión**

Esto siempre se puede resolver gráficamente

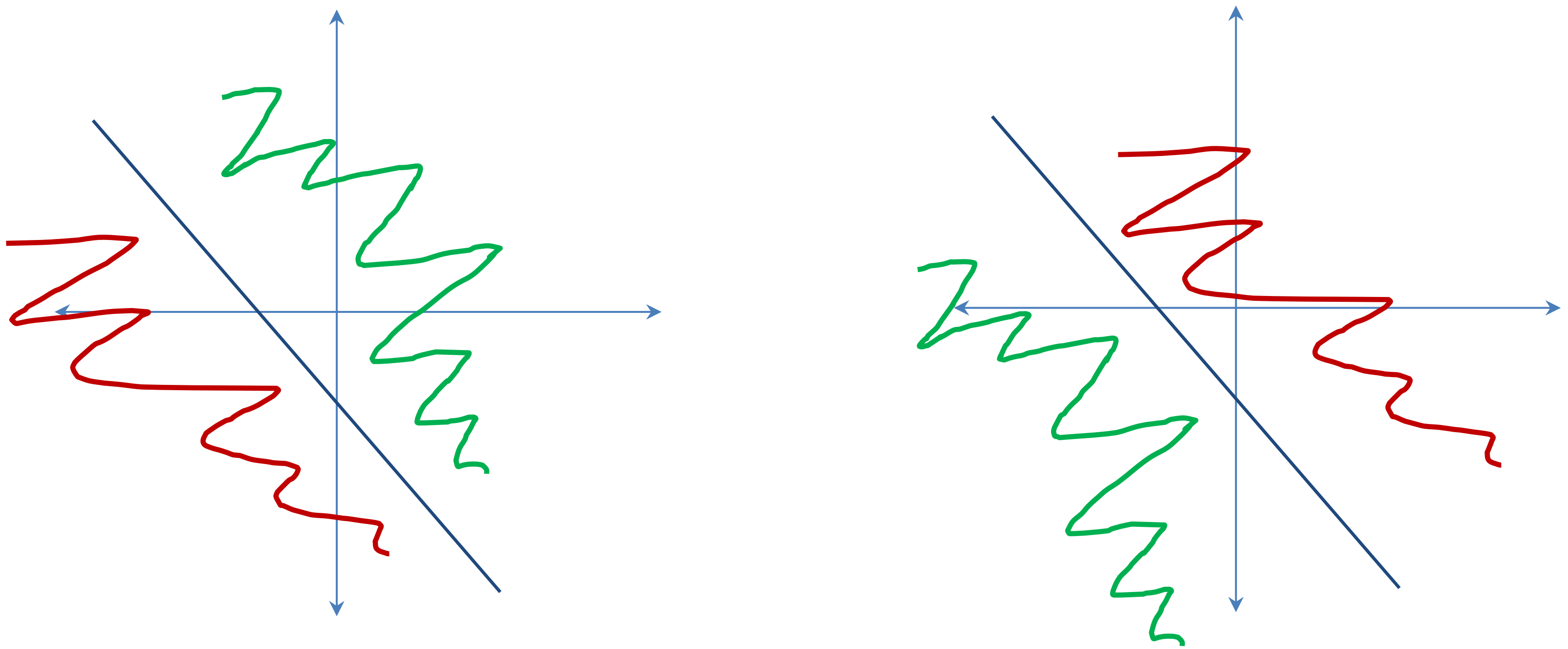
5. Clasificadores equivalentes

$g_A(y)= \ldots.$

$g_B(y)= \ldots.$

$g'_A(y)= \ldots.$

$g'_B(y)= \ldots.$



No son equivalentes

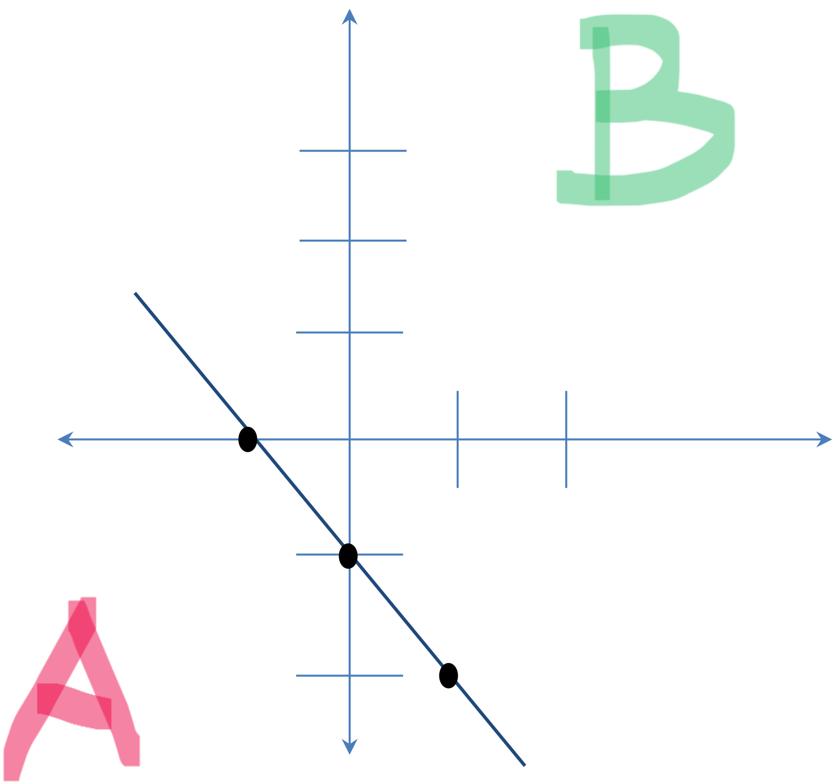
5. Clasificadores equivalentes

$g_A(y) = (-1\ -2\ 1)$

$g_B(y) = (0\ -1\ 2)$

$$-1 - 2y_1 + y_2 = -y_1 + 2y_2$$

$$\mathbf{-1 - y_1 = y_2}$$



$g_A(y) = (-3\ -2\ -1)$

$g_B(y) = (-2\ -1\ 0)$

A $\quad -1 - 2y_1 + y_2 \quad \begin{pmatrix}0\\0\end{pmatrix} \Rightarrow -1$

B $\quad -y_1 + 2y_2 \quad \begin{pmatrix}0\\0\end{pmatrix} \Rightarrow 0$

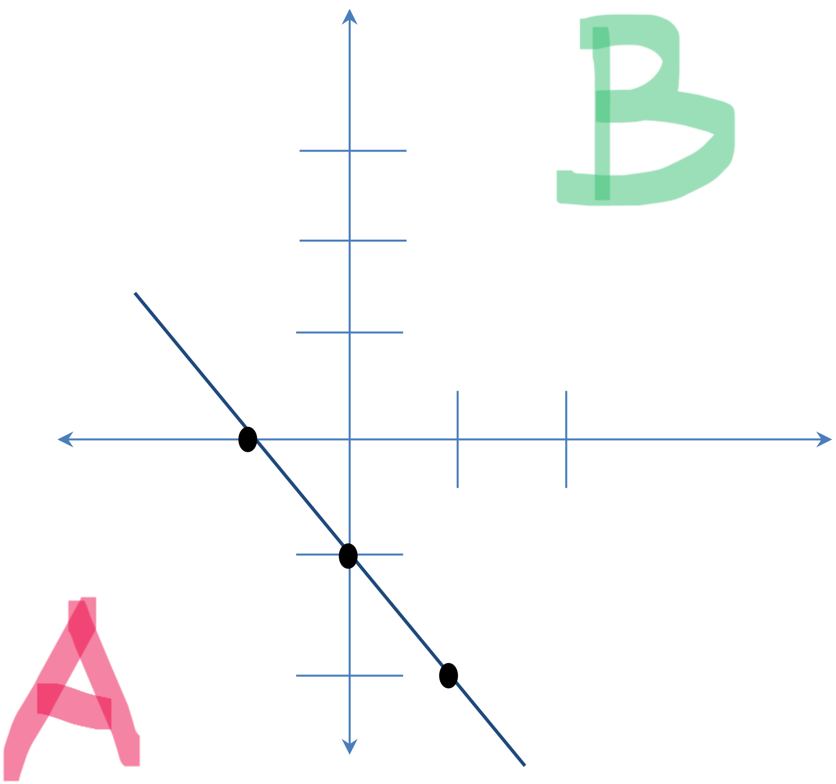
5. Clasificadores equivalentes

$g_A(y) = (-1\ -2\ 1)$

$g_B(y) = (0\ -1\ 2)$

$$-1 - 2y_1 + y_2 = -y_1 + 2y_2$$

$$\mathbf{-1 - y_1 = y_2}$$

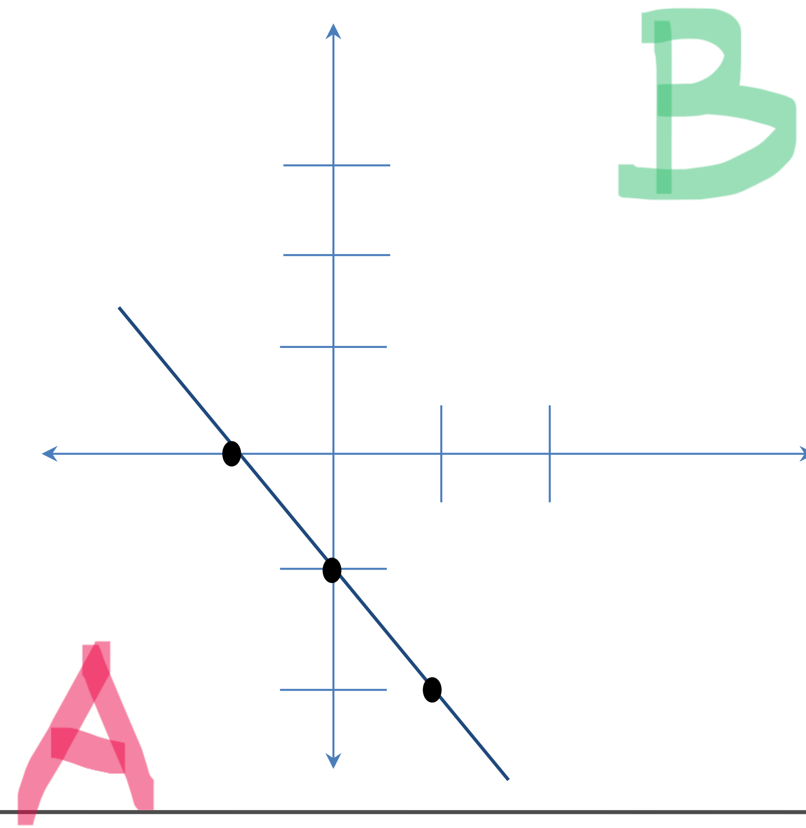


$g_A(y) = (-3\ -2\ -1)$

$g_B(y) = (-2\ -1\ \ 0)$

$$-3 - 2y_1 - y_2 = -2 - y_1$$

$$\mathbf{-1 - y_1 = y_2}$$



Algunas veces, se puede determinar si dos clasificadores son equivalentes analizando las funciones lineales de ambos clasificadores

G	**G'**
(0 -1 1)	(1 -1 1)
(0 1 -1)	(1 1 -1)

Si podemos encontrar **G' = G · a + b**, donde **a > 0**
y **G = G' · a' + b'**, donde **a' > 0**
entonces los dos clasificadores son equivalentes
(a, a', b and b' son constantes)

5. Clasificadores equivalentes

G

$$\begin{pmatrix}0 & -1 & 1\end{pmatrix}$$
$$\begin{pmatrix}0 & 1 & -1\end{pmatrix}$$

G'

$$\begin{pmatrix}1 & -1 & 1\end{pmatrix}$$
$$\begin{pmatrix}1 & 1 & -1\end{pmatrix}$$

G' = G · a + b

> 0

$$\begin{pmatrix}1 & -1 & 1\end{pmatrix} = \begin{pmatrix}0 & -1 & 1\end{pmatrix} \cdot \boxed{a} + \boxed{b}$$
$$\begin{pmatrix}1 & 1 & -1\end{pmatrix} = \begin{pmatrix}0 & 1 & -1\end{pmatrix} \cdot \boxed{a} + \boxed{b}$$

→ añadir término independiente

$$\begin{pmatrix}1 & -1 & 1\end{pmatrix} = \begin{pmatrix}0 & -1 & 1\end{pmatrix} \cdot 1 + 1$$
$$\begin{pmatrix}1 & 1 & -1\end{pmatrix} = \begin{pmatrix}0 & 1 & -1\end{pmatrix} \cdot 1 + 1$$

5. Clasificadores equivalentes

G

$$\begin{matrix}(0 & -1 & 1)\\(0 & 1 & -1)\end{matrix}$$

G’

$$\begin{matrix}(1 & -1 & 1)\\(1 & 1 & -1)\end{matrix}$$

G = G’ · a’ + b’

$$\begin{matrix}(0 & -1 & 1) & = & (1 & -1 & 1) & \cdot & \overbrace{\boxed{\begin{matrix}a'\\a'\end{matrix}}}^{>0} & + & \boxed{\begin{matrix}b'\\b'\end{matrix}}\\(0 & 1 & -1) & = & (1 & 1 & -1) & \cdot & & + & \end{matrix}$$

→ añadir término independiente

$$\begin{matrix}(0 & -1 & 1) & = & (1 & -1 & 1) & \cdot & 1 & + & (-1)\\(0 & 1 & -1) & = & (1 & 1 & -1) & \cdot & 1 & + & (-1)\end{matrix}$$

5. Clasificadores equivalentes

Se puede inferir una relación de **a** y **b** con **a'** y **b'**

a' = 1/a
b' = - b/a

Por tanto, ¿cómo podemos obtener clasificadores equivalentes?

$$g_i'(\mathbf{x}) = a \cdot g_i(\mathbf{x}) + b \qquad \text{con } a > 0 \qquad 1 \leq i \leq C$$

Otra forma de encontrar clasificadores equivalentes:

$$g_i'(\mathbf{x}) = ln\, g_i(\mathbf{x}) \qquad \text{con } g_i(\mathbf{x}) > 0 \qquad 1 \leq i \leq C$$

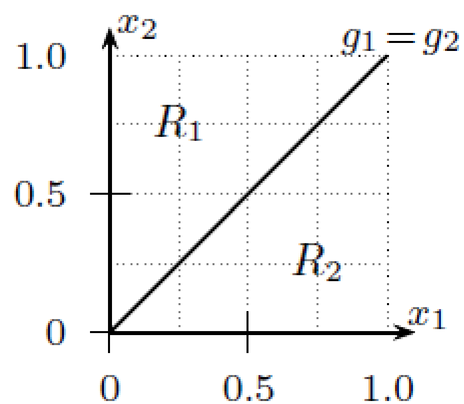
49 ☐ Dado el clasificador en dos clases definido por su frontera y regiones de decisión de la figura de la derecha, ¿cuál de los siguientes vectores de pesos *no* define un clasificador equivalente al dado?

A) $\mathbf{w}_1 = (0, 0, 1)^t$ y $\mathbf{w}_2 = (0, 1, 0)^t$.

B) $\mathbf{w}_1 = (0, 1, 0)^t$ y $\mathbf{w}_2 = (0, 0, 1)^t$.

C) $\mathbf{w}_1 = (0, -1, 0)^t$ y $\mathbf{w}_2 = (0, 0, -1)^t$.

D) Todos los vectores de pesos anteriores definen clasificadores equivalentes.



10 $\square$

(Examen de SIN del 30 de enero de 2013)
Para un problema de clasificación de dos clases en $\Re^2$ se han construido tres clasificadores distintos. Uno está formado por las dos funciones discriminantes lineales siguientes: $g_1(y) = 3 + 4\, y_1 - 2\, y_2$ y $g_2(y) = -3 + 1.5\, y_1 + 5\, y_2$. El segundo clasificador por $g'_1(y) = 6 + 8\, y_1 - 4\, y_2$ y $g'_2(y) = -6 + 3\, y_1 + 10\, y_2$. El tercero por $g''_1(y) = -6 - 8\, y_1 + 4\, y_2$ y $g''_2(y) = 6 - 3\, y_1 - 10\, y_2$. ¿Los tres clasicadores son equivalentes? es decir ¿definen las mismas fronteras de decisión?

A) (g_1, g_2) y (g'_1, g'_2) son equivalentes.
B) Los tres son equivalentes.
C) (g_1, g_2) y (g''_1, g''_2) son equivalentes.
D) (g'_1, g'_2) y (g''_1, g''_2) son equivalentes.

G1	G1’	G1’’
(3 4 -2)	(6 8 -4)	(-6 -8 4)
(-3 1.5 5)	(-6 3 10)	(6 -3 -10)

(6 8 -4) = (3 4 -2) · 2 + 0
(-6 3 10) = (-3 1.5 5) · 2 + 0

Respuesta: A

66 ☐ Dado el clasificador en 2 clases definido por sus vectores de pesos $\mathbf{w}_1 = (-2, 3, 3)^t$, $\mathbf{w}_2 = (0, 2, -2)^t$ en notación homogénea, ¿cuál de los siguientes conjuntos de vectores **no** define un clasificador equivalente al dado?

A) $\mathbf{w}_1 = (1, 3, 3)^t$, $\mathbf{w}_2 = (3, 2, -2)^t$

B) $\mathbf{w}_1 = (-4, 6, 6)^t$, $\mathbf{w}_2 = (0, 4, -4)^t$

C) $\mathbf{w}_1 = (-1, 6, 6)^t$, $\mathbf{w}_2 = (3, 4, -4)^t$

D) $\mathbf{w}_1 = (2, -3, -3)^t$, $\mathbf{w}_2 = (0, -2, 2)^t$

41 ☐ Sea un clasificador lineal para dos clases, $\circ$ y $\bullet$, de vectores de pesos $\mathbf{a}_\circ = (2, -5, 4)^t$ y $\mathbf{a}_\bullet = (5, 1, 1)^t$, respectivamente. ¿Cuál de las siguientes afirmaciones es correcta?

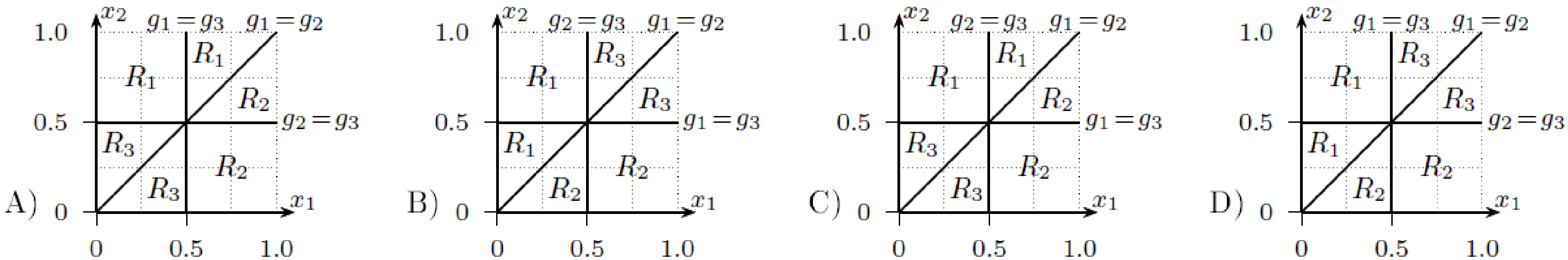
A) Los vectores de pesos $\mathbf{a}_\circ = (3, 4, 1)^t$ y $\mathbf{a}_\bullet = (2, 2, 2)^t$ definen la misma frontera de decisión que los del enunciado.

B) Los vectores de pesos $\mathbf{a}_\circ = (-2, 5, -4)^t$ y $\mathbf{a}_\bullet = (-5, -1, -1)^t$ definen un clasificador equivalente al del enunciado.

C) El punto $\mathbf{x}' = (1, 2)^t$ pertenece a la clase $\circ$.

D) El punto $\mathbf{x}' = (-2, 0)^t$ se encuentra en la frontera de decisión.

48 ☐ Sea un clasificador en tres clases basado en las funciones discriminantes lineales bidimensionales de vectores de pesos: $\mathbf{w}_1 = (0, 0, 1)^t$, $\mathbf{w}_2 = (0, 1, 0)^t$ y $\mathbf{w}_3 = (0.5, 0, 0)^t$. Indica cuál de las figuras dadas a continuación es coherente con las fronteras y regiones de decisión que define dicho clasificador.

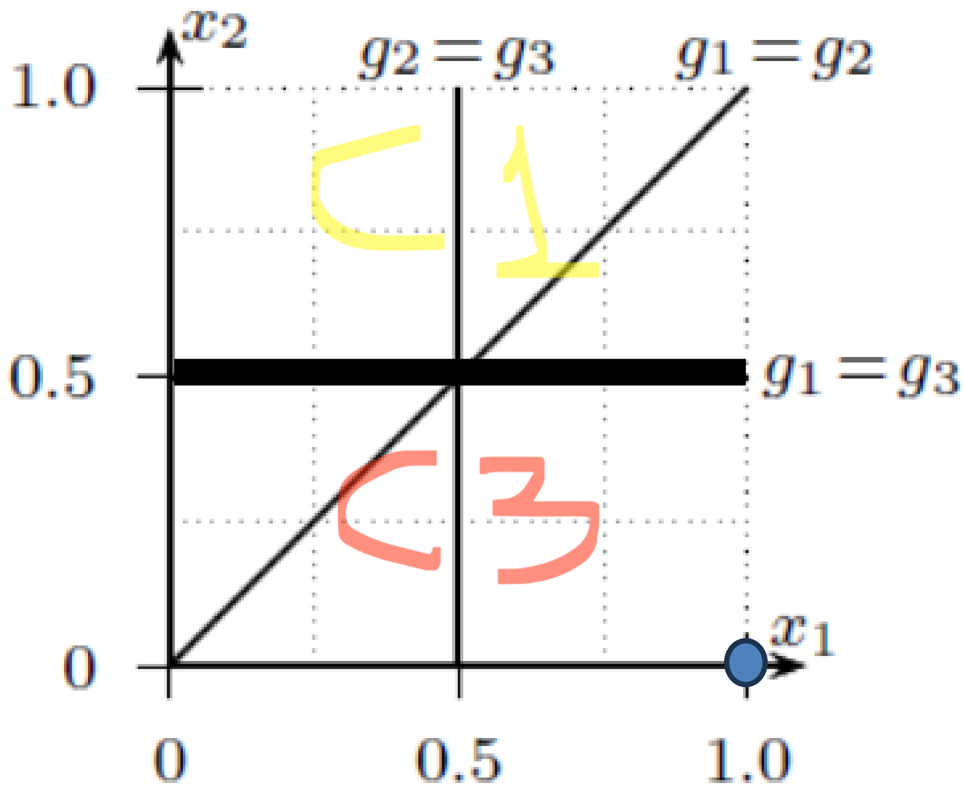


w1=(0 0 1) g1(x) = 0+0+x2 ➔ g1(x)=x2

w2=(0 1 0) g2(x) = 0+x1+0 ➔ g2(x)=x1

w3=(0.5 0 0) g3(x)= 0.5+0+0 ➔ g3(x)=0.5

Si igualamos las funciones lineales de la clase 1 (g1) con clase 3 (g3) nos sale la frontera de decisión x2=0.5, así que la respuesta tiene que ser B o C

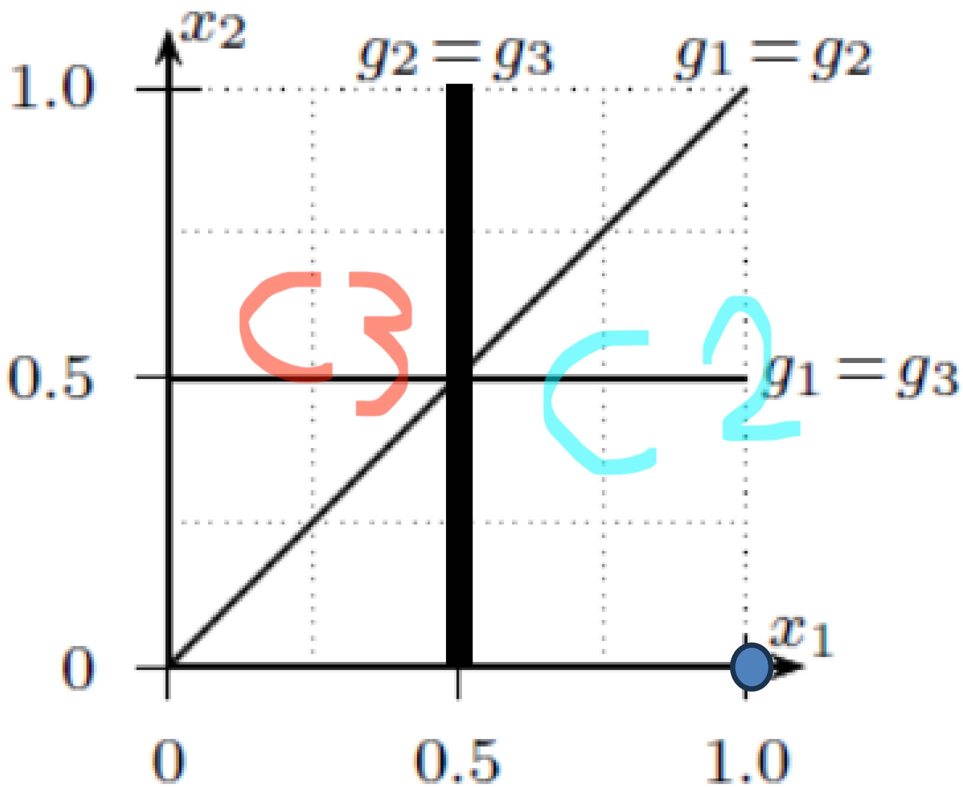


$$s = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{array}{l} \rightarrow x1 \\ \rightarrow x2 \end{array}$$

w1=(0 0 1) y1= 0+0+x2 ➔ g1(x)=x2

w3=(0.5 0 0) y3= 0.5+0+0 ➔ g3(x)=0.5

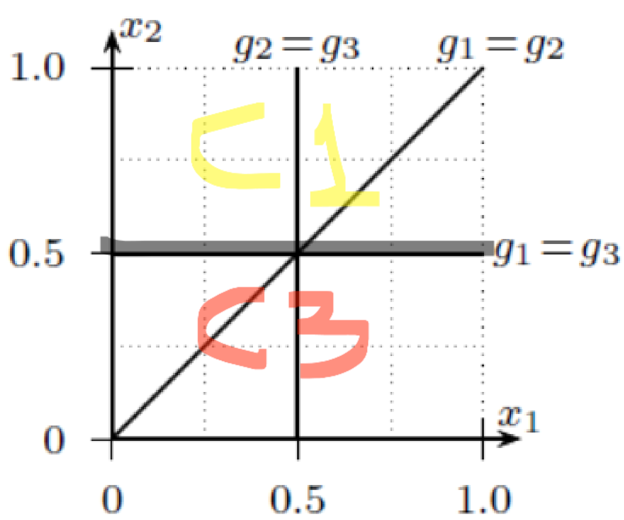
Viendo las gráficas, sabemos que la respuesta tiene que ser C. No obstante, vamos a completar el cálculo de las regiones.



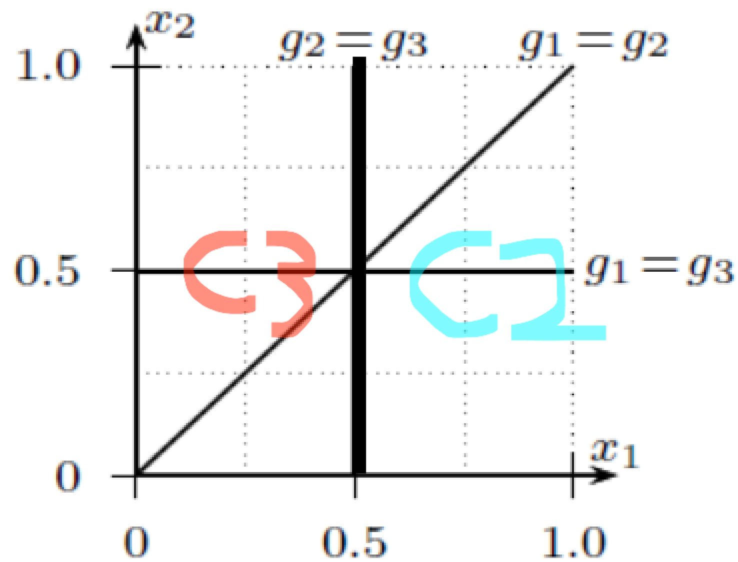
$$s = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{array}{l} \rightarrow x1 \\ \rightarrow x2 \end{array}$$

w2=(0 1 0) y2= 0+x1+0 ➔ g2(x)=x1

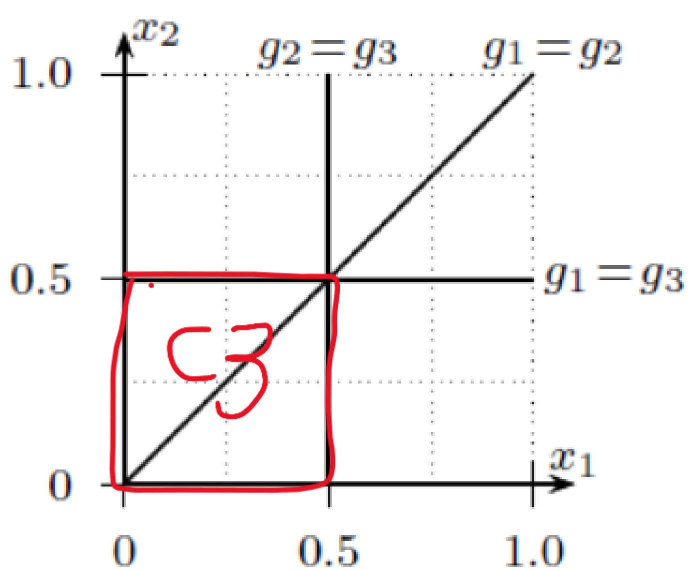
w3=(0.5 0 0) y3= 0.5+0+0 ➔ g3(x)=0.5



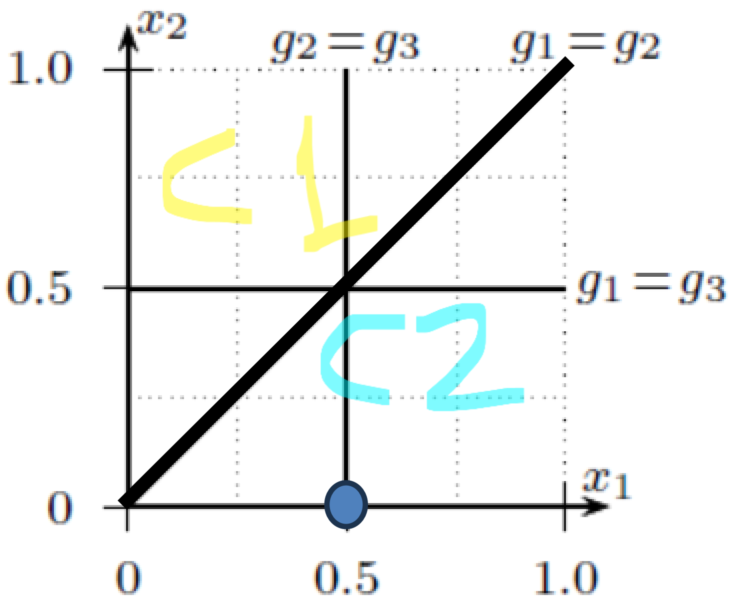
x_2
$g_2 = g_3$
$g_1 = g_2$
1.0
0.5
0
C1
C3
$g_1 = g_3$
x_1
0
0.5
1.0



x_2
$g_2 = g_3$
$g_1 = g_2$
1.0
0.5
0
C3
C2
$g_1 = g_3$
x_1
0
0.5
1.0



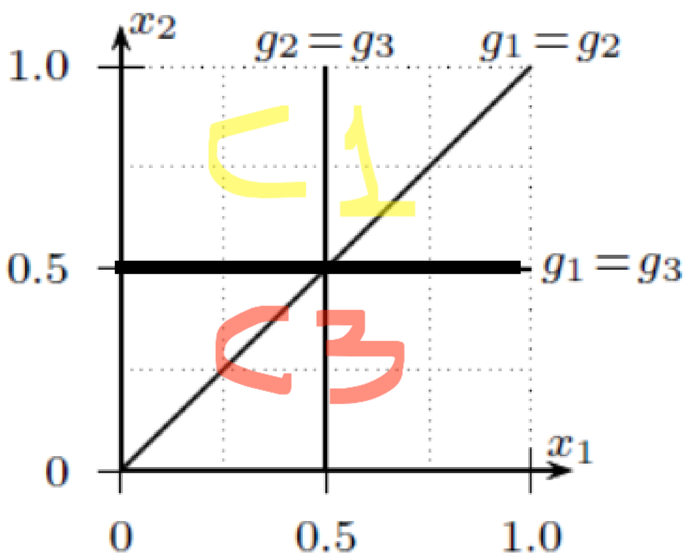
x_2
$g_2 = g_3$
$g_1 = g_2$
1.0
0.5
0
C3
$g_1 = g_3$
x_1
0
0.5
1.0

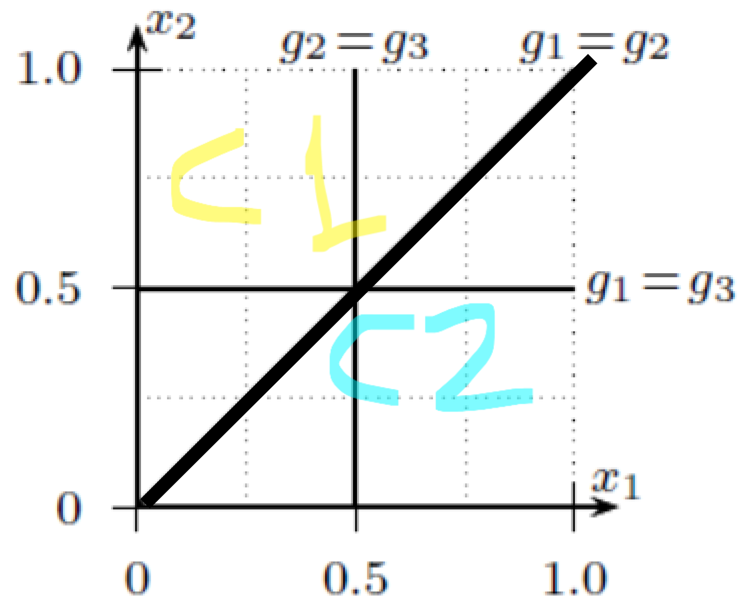


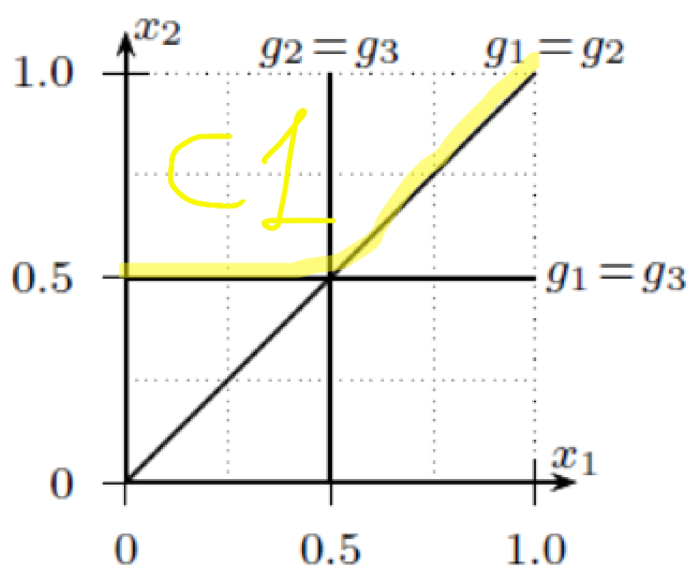
$$s = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix} \begin{matrix} \rightarrow x1 \\ \rightarrow x2 \end{matrix}$$

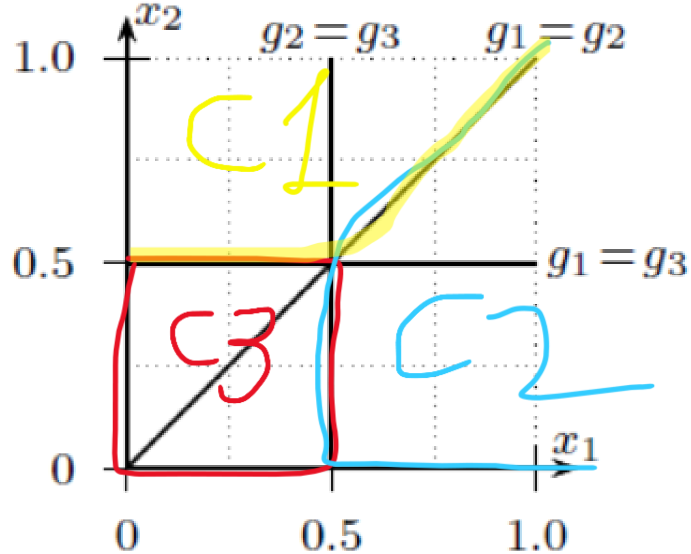
w1=(0 0 1) y1= 0+0+x2 ➔ g1(x)=x2

w2=(0 1 0) y2= 0+x1+0 ➔ g2(x)=x1


x_2
1.0
0.5
0
$g_2 = g_3$
$g_1 = g_2$
$g_1 = g_3$
C1
C3
x_1
0
0.5
1.0


x_2
1.0
0.5
0
$g_2 = g_3$
$g_1 = g_2$
$g_1 = g_3$
C1
C2
x_1
0
0.5
1.0


x_2
1.0
0.5
0
$g_2 = g_3$
$g_1 = g_2$
$g_1 = g_3$
C1
x_1
0
0.5
1.0


x_2
$g_2 = g_3$
$g_1 = g_2$
1.0
0.5
0
$g_1 = g_3$
C1
C3
C2
x_1
0
0.5
1.0

38 $\square$

Se tiene un problema de clasificación en 3 clases, $c = 1, 2, 3$, para objetos representados mediante vectores de 2 características reales, $\mathbf{x} = (x_1, x_2)^t \in \mathbb{R}^2$. Considérese un clasificador lineal de vectores de pesos (en notación homogénea): $\mathbf{w}_1 = (w_{10}, w_{11}, w_{12})^t = (2, 0, 0)^t$, $\mathbf{w}_2 = (0, 1, 1)^t$ y $\mathbf{w}_3 = (0, 1, -1)^t$. La región de decisión de la clase 1 correspondiente a este clasificador es:

A) $\{\mathbf{x} : x_1 \geq 0 \wedge x_2 < -x_1 + 2\} \cup \{\mathbf{x} : x_1 < 0 \wedge x_2 < x_1 + 2\}$.
B) $\{\mathbf{x} : x_2 \geq 0 \wedge x_2 < -x_1 + 2\} \cup \{\mathbf{x} : x_2 < 0 \wedge x_2 > x_1 - 2\}$.
C) $\{\mathbf{x} : x_1 \geq 0 \wedge x_2 < -x_1 + 1\} \cup \{\mathbf{x} : x_1 < 0 \wedge x_2 < x_1 + 1\}$.
D) $\{\mathbf{x} : x_2 \geq 0 \wedge x_2 < -x_1 + 1\} \cup \{\mathbf{x} : x_2 < 0 \wedge x_2 > x_1 - 1\}$.

w1 = (2 0 0)
w2 = (0 1 1)
w3 = (0 1 -1)

g1(x)=2
g2(x)=x1+x2
g3(x)=x1-x2

Frontera decisión clase 1 – clase 2

2=x1+x2

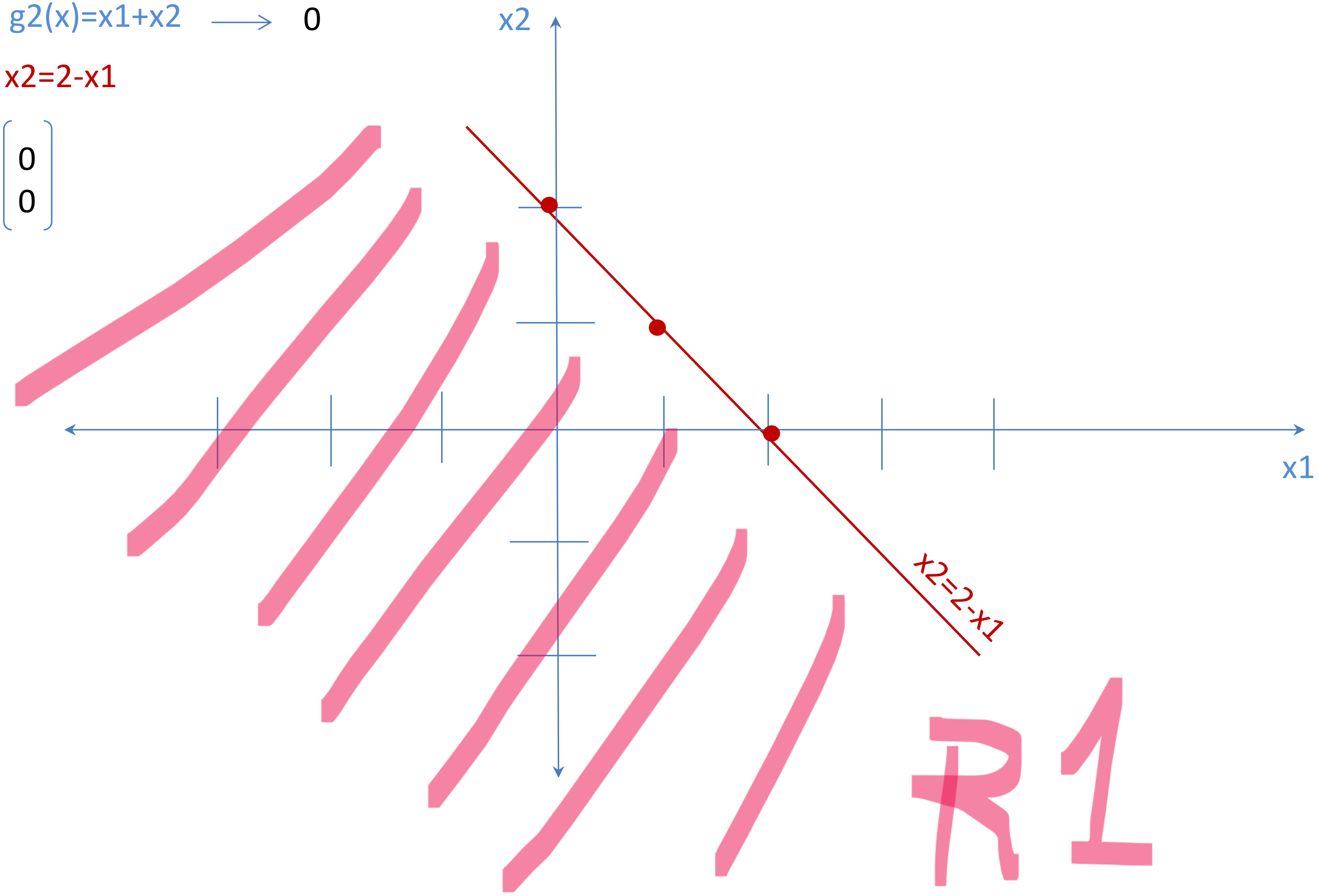
x2=2-x1

$g1(x)=2 \longrightarrow 2$

$g2(x)=x1+x2 \longrightarrow 0$

$x2=2-x1$

$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$

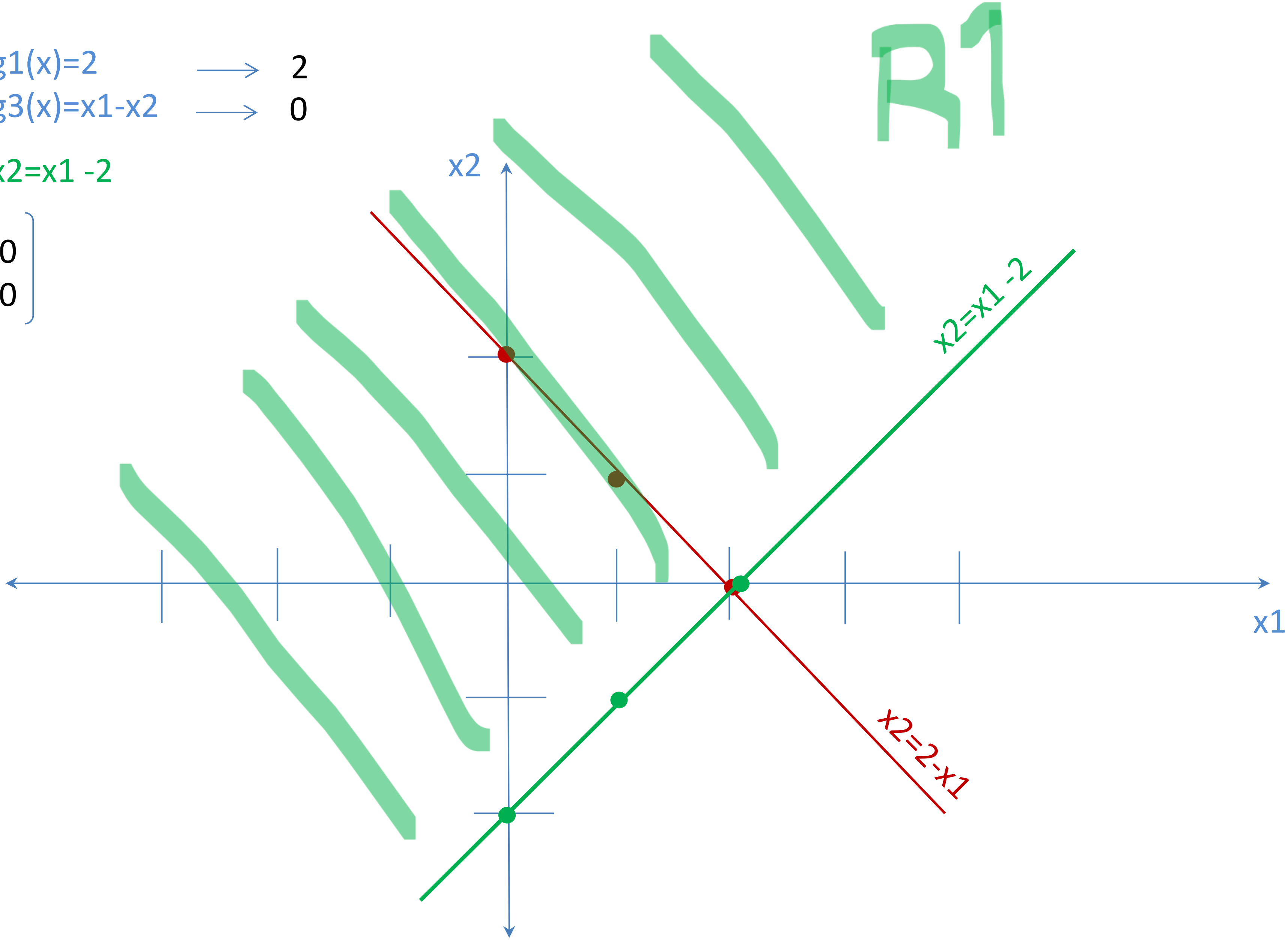


$g1(x)=2 \longrightarrow 2$

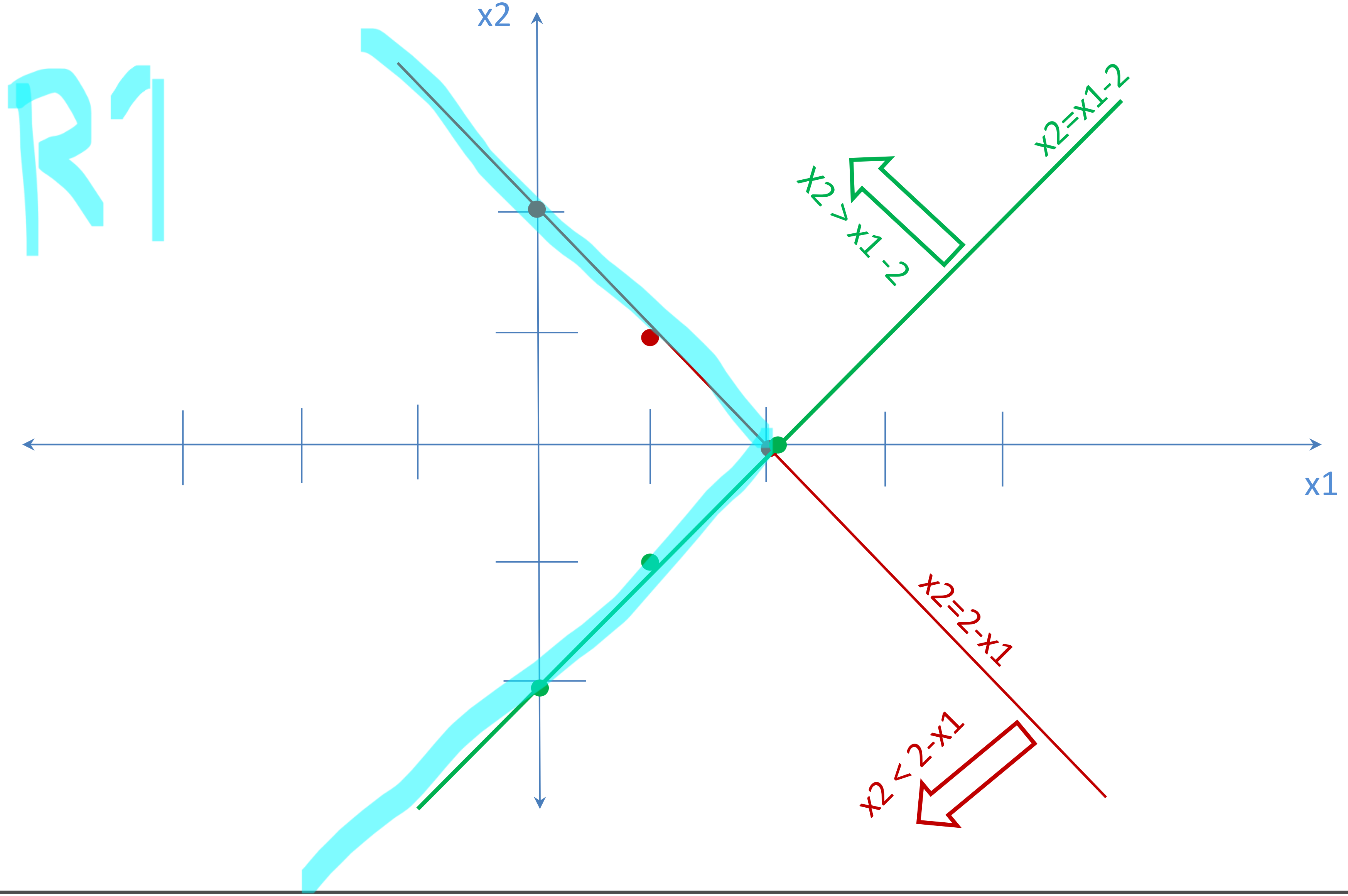
$g3(x)=x1-x2 \longrightarrow 0$

$x2=x1-2$

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$$



Respuesta: B



6. Perceptron

Dadas N muestras de aprendizaje $(\mathbf{x}_1, c_1)(\mathbf{x}_2, c_2) \dots (\mathbf{x}_1, c_N)$ pertenecientes a C clases, queremos encontrar C vectores de pesos $\mathbf{w}_j$, $1 \leq j \leq C$ que clasifiquen lo más correctamente (posible) las muestras de aprendizaje dadas; es decir:

$$\mathbf{w}_{c_1}^t \mathbf{x}_1 > \mathbf{w}_{c_j}^t \mathbf{x}_1, \forall j \neq c_1$$

$$\mathbf{w}_{c_2}^t \mathbf{x}_2 > \mathbf{w}_{c_j}^t \mathbf{x}_2, \forall j \neq c_2$$

$$\dots$$

$$\mathbf{w}_{c_N}^t \mathbf{x}_N > \mathbf{w}_{c_j}^t \mathbf{x}_N, \forall j \neq c_N$$

Una solución es ajustar iterativamente unos **pesos iniciales** mediante el **Algoritmo del Perceptron** introducido en 1957 por Frabk Rosenblatt.

Los algoritmos basados en el Perceptron pueden considerarse como los ejemplos más sencillos de redes neuronales.

Presentan problemas de convergencia con clases no linealmente separables: Solución: “margen”.

6. Perceptron

```
;; Input:  w_j, 1 ≤ j ≤ C, C vectores de pesos iniciales
;;        (x_1, c_1)(x_2, c_2) ... (x_1, c_N), N muestras de aprendizaje con sus clases
;;        α ∈ ℝ^{>0}   factor de aprendizaje
;;        b ∈ ℝ 'margen' para ajustar la convergencia
```

;; Input: $\mathbf{w}_j, 1 \leq j \leq C, C$ vectores de pesos iniciales

;; $(\mathbf{x}_1, c_1)(\mathbf{x}_2, c_2) \ldots (\mathbf{x}_1, c_N), N$ muestras de aprendizaje con sus clases

;; $\alpha \in \mathbb{R}^{>0}$ factor de aprendizaje

;; $b \in \mathbb{R}$ 'margen' para ajustar la convergencia

do {
  $m = 0$;; número de muestras mal clasificadas
  for ($n = 1; n \leq N; n$++){ ;; bucle muestras
    $i = c_n$;
    $g = \mathbf{w}_i^t \mathbf{x}_n$;
    error=false;
    for ($j = 1; j \leq C; j$++) ;; bucle clases
      if ($i \neq j$
        then if ($\mathbf{w}_j^t \mathbf{x}_n + b \geq g$)
          then { $\mathbf{w}_j = \mathbf{w}_j - \alpha \mathbf{x}_n$
            error = true }
    if (error)
      then { $\mathbf{w}_i = \mathbf{w}_i + \alpha \mathbf{x}_n$
        $m = m + 1$}
} while ($m \neq 0$)

6. Perceptron (ejercicio)

En un problema de clasificación en 2 clases, para objetos representados mediante vectores de características bidimensionales, se tienen dos muestras de entrenamiento: $\boldsymbol{y}_1 = (0,0)^t$, $\boldsymbol{y}_2 = (1,1)^t$ de clases $c_1 = 1$, $c_2 = 2$, respectivamente.

Mostrar una traza de ejecución del algoritmo Perceptrón, con *vectores de pesos iniciales* nulos, *factor de aprendizaje* $\alpha = 1$ y *margen* $b = 0{,}1$. La traza debe incluir las sucesivas actualizaciones de los vectores de pesos de las clases.

$y_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\quad$ $y_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\quad$ $w_1 = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$ $\quad$ $w_2 = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$

c=1 $\quad$ c=2

6. Perceptron (ejercicio)

$$\underbrace{\begin{pmatrix}0\\0\\0\end{pmatrix}}_{w_1} \quad \underbrace{\begin{pmatrix}0\\0\\0\end{pmatrix}}_{w_2}$$

Iteration 1 m=0 mal clasificadas

muestras: for n=1 i=1 (c=1)

$$g = \underbrace{\begin{pmatrix}0 & 0 & 0\end{pmatrix}}_{w_1} \underbrace{\begin{pmatrix}1\\0\\0\end{pmatrix}}_{x_1} = 0$$

error=false

clases: j=1 j≠i

$$\underbrace{\begin{pmatrix}0 & 0 & 0\end{pmatrix}}_{w_2} \underbrace{\begin{pmatrix}1\\0\\0\end{pmatrix}}_{x_1} + 0.1 = 0.1 >= g$$

$$w_2 = \underbrace{\begin{pmatrix}0\\0\\0\end{pmatrix}}_{w_2} - 1 \underbrace{\begin{pmatrix}1\\0\\0\end{pmatrix}}_{x_1} = \begin{pmatrix}-1\\0\\0\end{pmatrix}$$

error=true

6. Perceptron (ejercicio)

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$w_1 \quad w_2$$

$$w_1 = \underbrace{\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}}_{w_1} + 1 \underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}_{x_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

m=m+1= 1 muestra mal clasificada

6. Perceptron (ejercicio)

$$\underset{w_1}{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}} \quad \underset{w_2}{\begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}}$$

muestras: for n=2 i=2 (c=2)

$$g = \underset{w_2}{\begin{pmatrix} -1 & 0 & 0 \end{pmatrix}} \underset{x_2}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}} = -1$$

error=false

clases: j=1 j≠i

$$\underset{w_1}{\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}} \underset{x_2}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}} + 0.1 = 1.1 >= g$$

$$w_1 = \underset{w_1}{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}} - 1 \underset{x_2}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}} = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}$$

error=true

6. Perceptron (ejercicio)

$$w_2 = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

$$\quad w_2 \qquad x_2$$

m=m+1= 2 muestras mal clasificadas

6. Perceptron (ejercicio)

$$\underset{w_1}{\begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}} \quad \underset{w_2}{\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}}$$

Iteration 2 m=0 mal clasificadas

muestras: for n=1 i=1 (c=1)

$$g = \underset{w_1}{\begin{pmatrix} 0 & -1 & -1 \end{pmatrix}} \underset{x_1}{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}} = 0$$

error=false

clases: j=1 j≠i

$$\underset{w_2}{\begin{pmatrix} 0 & 1 & 1 \end{pmatrix}} \underset{x_1}{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}} + 0.1 = 0.1 \quad >= g$$

$$w_2 = \underset{w_2}{\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}} - 1 \underset{x_1}{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

error=true

6. Perceptron (ejercicio)

$$w_1 = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$$

$\quad w_1 \qquad x_1$

m=m+1= 1 muestra mal clasificada

6. Perceptron (ejercicio)

$$\underset{w_1}{\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}} \quad \underset{w_2}{\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}}$$

muestras: for n=2 i=2 (c=2)

$$g = \underset{w_2}{\begin{pmatrix} -1 & 1 & 1 \end{pmatrix}} \underset{x_2}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}} = 1$$

error=false

clases: j=1 j≠i

$$\underset{w_1}{\begin{pmatrix} 1 & -1 & -1 \end{pmatrix}} \underset{x_2}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}} + 0.1 = -0.9 < g$$

6. Perceptron (ejercicio)

$$\underset{w_1}{\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}} \quad \underset{w_2}{\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}}$$

Iteration 3 m=0 mal clasificadas

muestras: for n=1 i=1 (c=1)

$$g = \underset{w_1}{\begin{pmatrix} 1 & -1 & -1 \end{pmatrix}} \underset{x_1}{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}} = 1$$

error=false

clases: j=1 j≠i

$$\underset{w_2}{\begin{pmatrix} -1 & 1 & 1 \end{pmatrix}} \underset{x_1}{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}} + 0.1 = -0.9 < g$$

6. Perceptron (ejercicio)

$$\underset{w_1}{\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}} \quad \underset{w_2}{\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}}$$

muestras: for n=2 i=2 (c=2)

$$g = \underset{w_2}{\begin{pmatrix} -1 & 1 & 1 \end{pmatrix}} \underset{x_2}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}} = 1$$

error=false

classes: j=1 j≠i

$$\underset{w_1}{\begin{pmatrix} 1 & -1 & -1 \end{pmatrix}} \underset{x_2}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}} + 0.1 = -0.9 < g$$

$$m = 0$$

The END

6. Perceptron: convergencia y calidad de los resultados

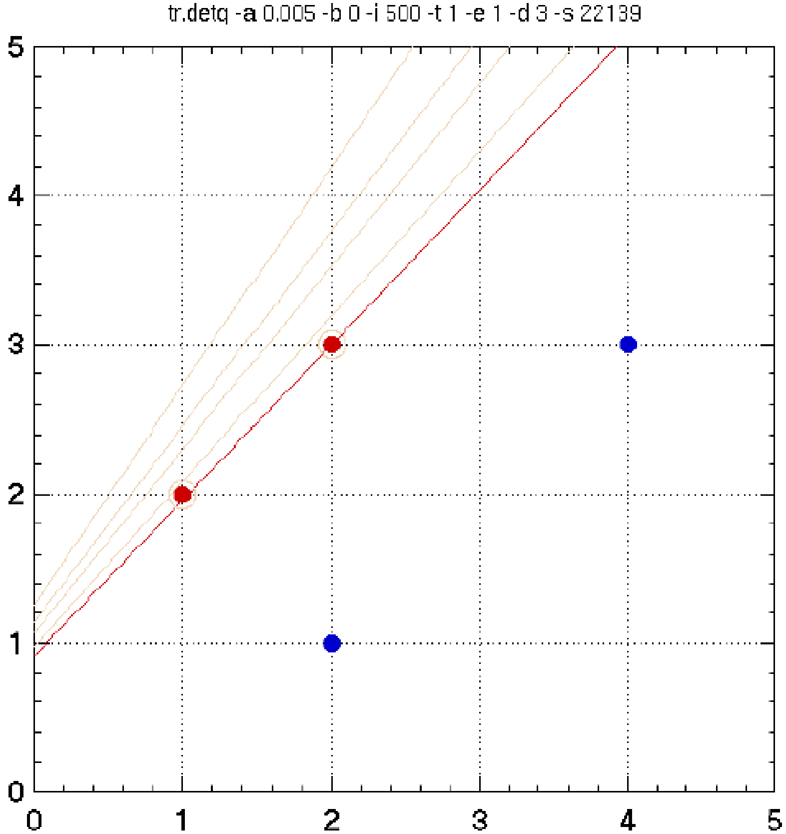
Tres parámetros controlan la convergencia y calidad de resultados:

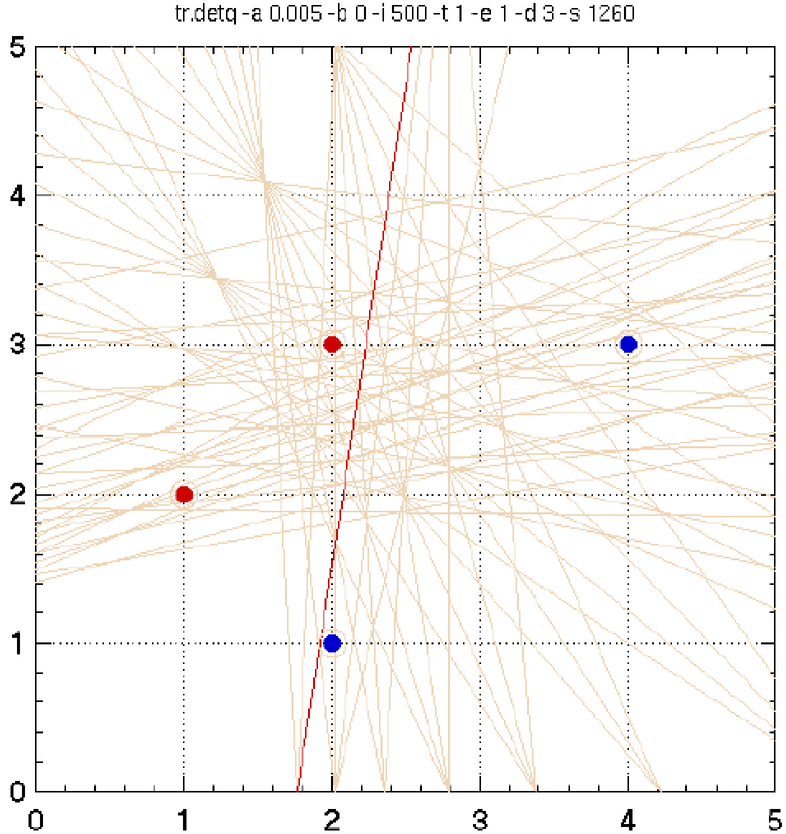
α, b *(margen)*, M *(máximo número de iteraciones)*

α determina el tamaño de las correcciones y por tanto la *velocidad de aprendizaje*. En general, $\alpha \ll \Rightarrow$ convergencia suave, pero con más iteraciones.

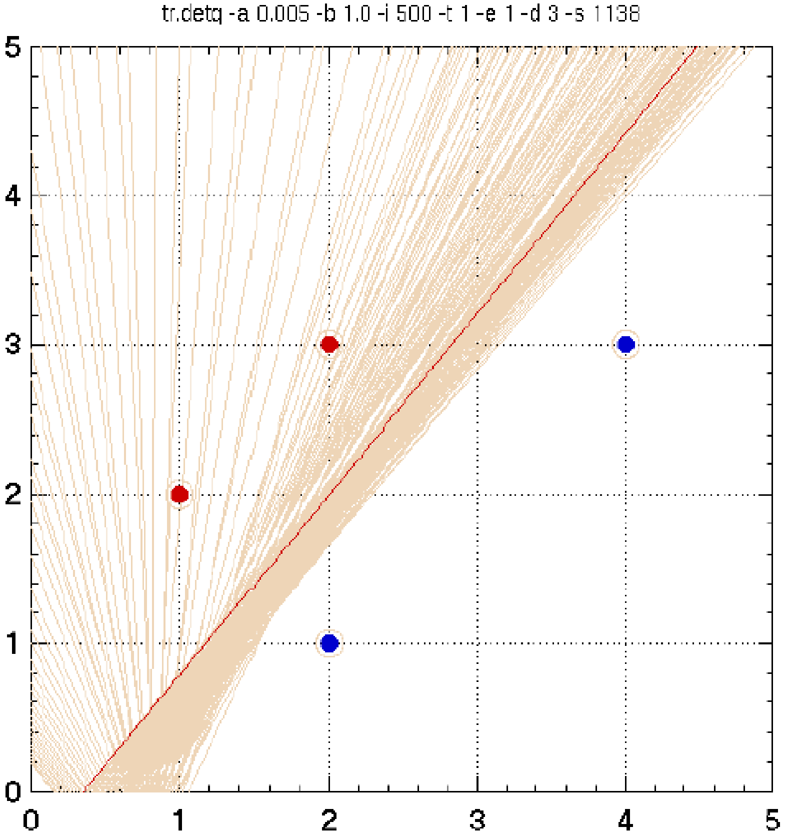
- *Comportamiento con conjuntos de entrenamiento linealmente separables:*
  - Converge en un número finito de iteraciones $\forall \alpha > 0$
  - $b = 0$: Las fronteras de decisión pueden tener poca "holgura"; es decir, pueden resultar demasiado cerca de algunos datos
  - $b > 0$: si b es bastante grande, se obtienen fronteras de decisión "centradas" entre las regiones de decisión (típicamente mucho mejores que con $b = 0$)
- *Comportamiento con conjuntos de entrenamiento* ***no*** *linealmente separables:*
  - $b = 0$: ninguna garantía de convergencia ni de calidad del resultado
  - $b > 0$: no hay convergencia pero, con b y M suficientemente grandes, se obtienen buenas fronteras de decisión; generalmente (casi-)*óptimas*, en el sentido de minimizar el error de clasificación del conjunto de entrenamiento

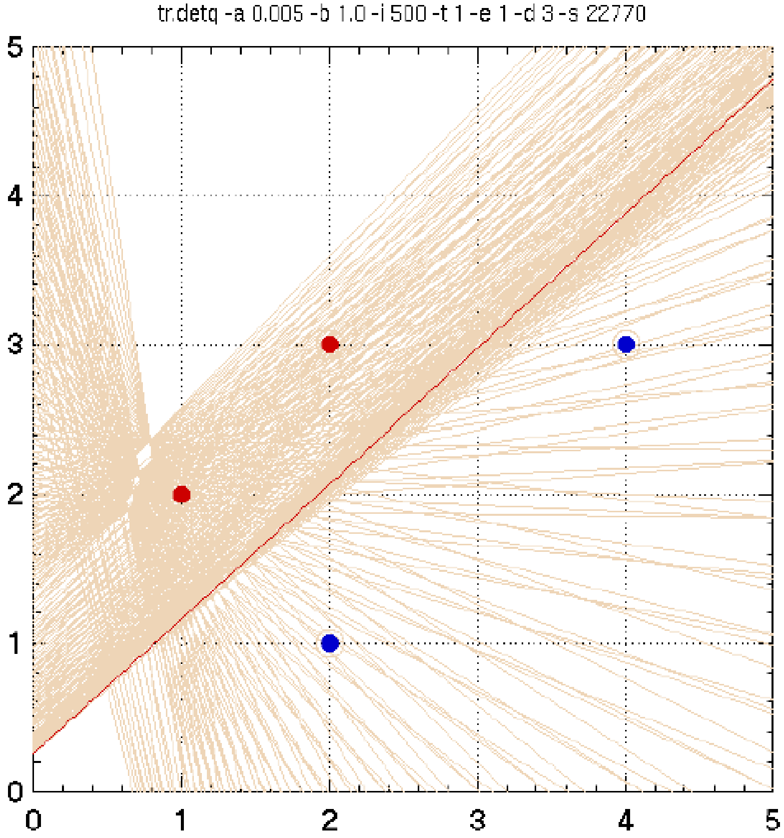
6. Perceptron: ejemplo simple separable con margen nulo



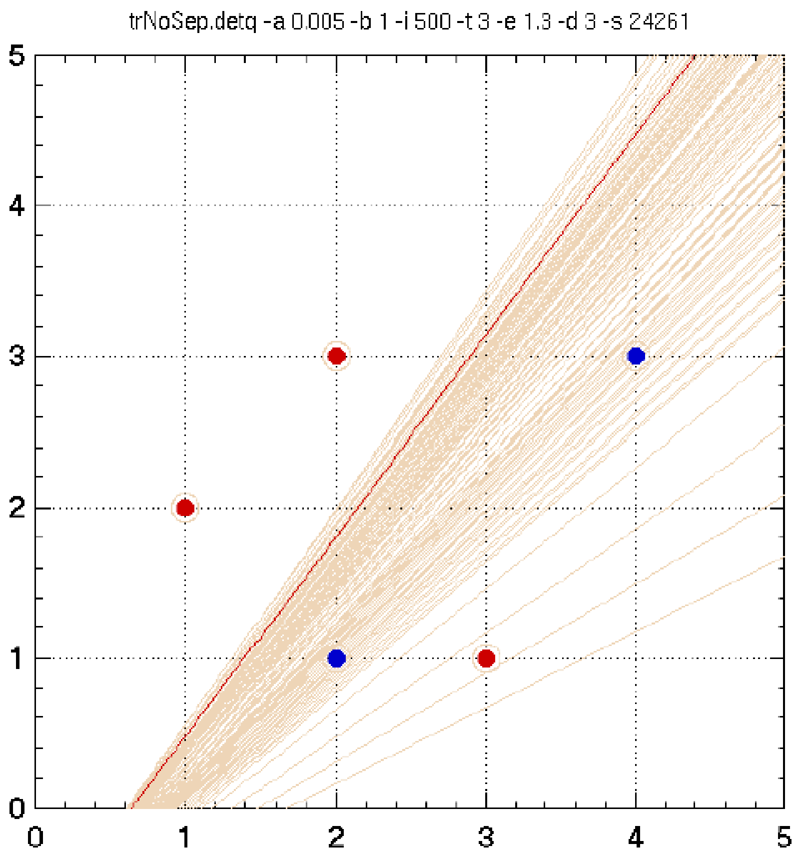


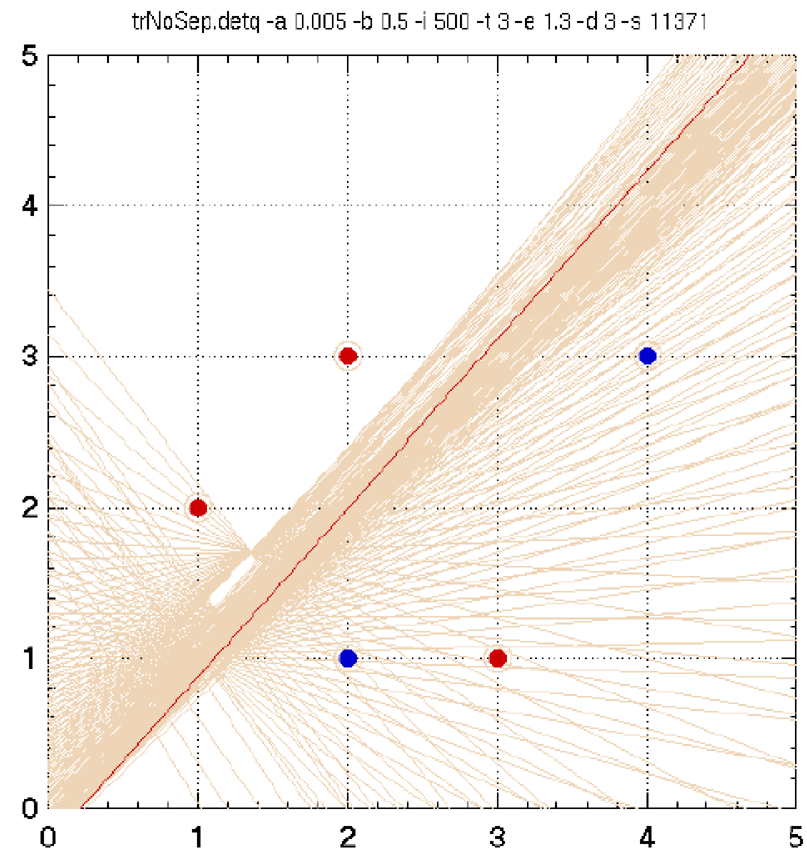
6. Perceptron: ejemplo simple separable con margen positivo





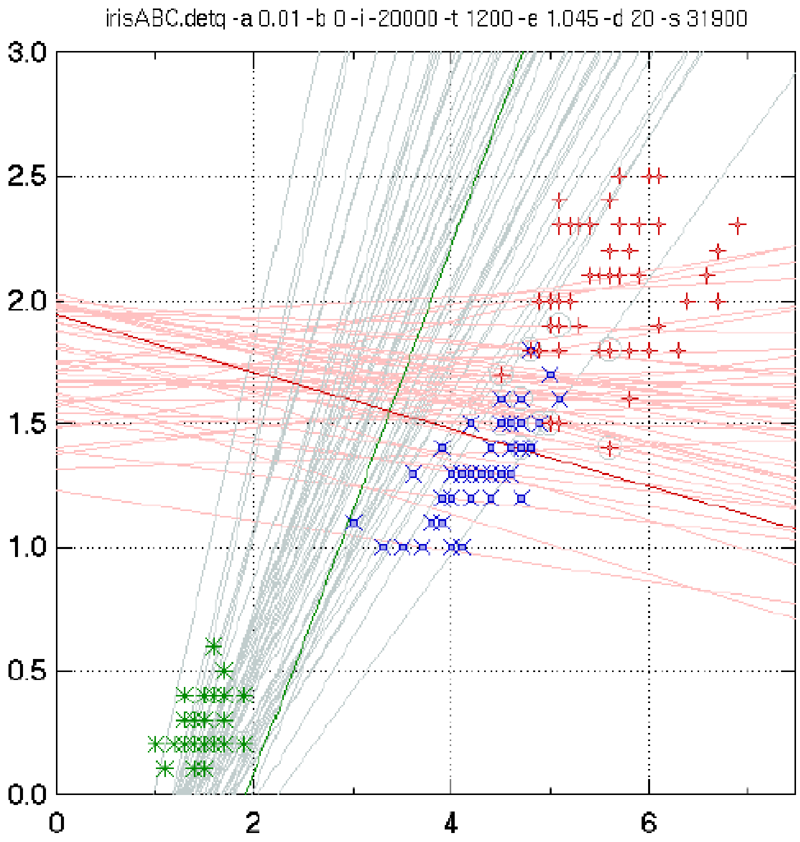
6. Perceptron: ejemplo simple no linealmente separable con margen positivo



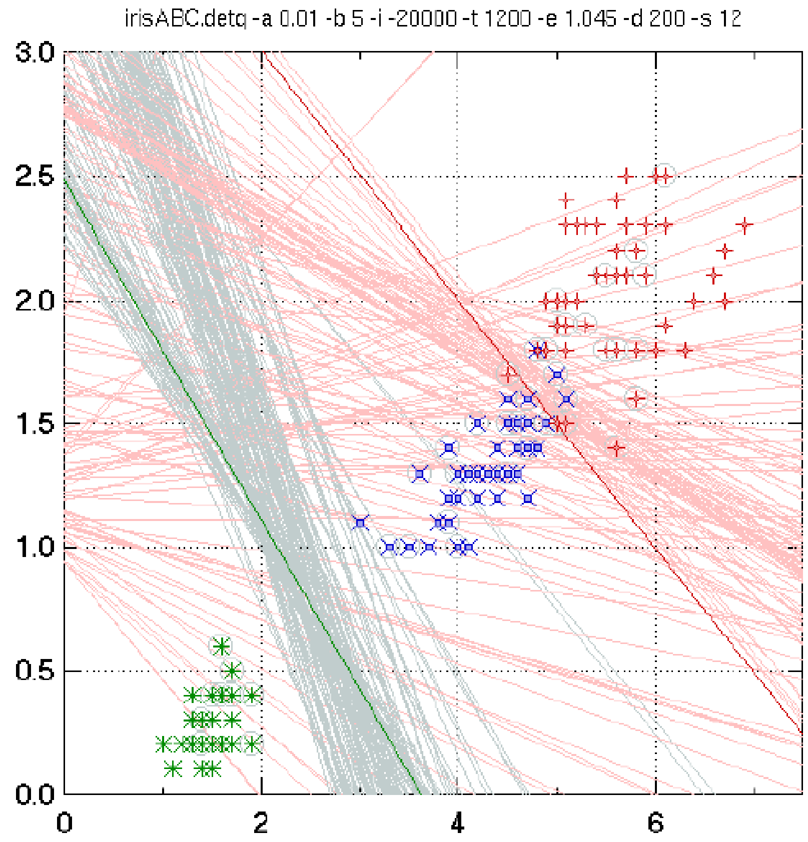


6. Perceptron: flores Iris2D (no separable)

margen nulo



margen positivo



7. Estimación empírica del error

¿Cómo es el **error teórico** (esperado) de un clasificador c(x) para toda muestra x?

$$\varepsilon = \mathrm{E}(\varepsilon(c(x))) = \sum_x P(x)\,\varepsilon(c(x))$$

error del clasificador para cada muestra x multiplicado por la probabilidad a priori de la muestra

$\varepsilon(c(x))$: probabilidad de error del clasificador c(x) para la muestra x

$$\varepsilon(c(x)) = 1 - \boxed{P(c = c(x) \mid x)}$$

- probabilidad de que el clasificador c(x) escoja la clase c para la muestra x
- si escoge la clase que maximiza la probabilidad entonces es mínima probabilidad de error o error de mínimo riesgo (error Bayes)

7. Estimación empírica del error

Ne: número de errores
N: número de muestras
$\hat{p}$: error estimado
p: probabilidad real de error

$$\Rightarrow \quad \hat{p} = Ne \,/\, N$$

Cómo de cerca está $\hat{p}$ de p? Cómo aproxima el error estimado $\hat{p}$ la verdadera probabilidad de error en un problema real?

El intervalo de confianza es un tipo de estimación de intervalo (de un parámetro poblacional) que se calcula a partir de los datos observados. Los intervalos de confianza consisten en un rango de valores (intervalo) que actúan como buenas estimaciones del parámetro poblacional desconocido.

7. Estimación empírica del error

El nivel de confianza deseado lo establece el investigador (no lo determinan los datos). Lo más habitual es utilizar el nivel de confianza del **95%.**

95 % confidence interval:

$$P(\hat{p} - \epsilon \leq p \leq \hat{p} + \epsilon) \; = \; 0.95; \qquad \epsilon \; = \; 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

Intervalo de confianza $\longrightarrow$ $I = [\hat{p} \pm \epsilon]$

Por ejemplo: observamos 50 decisiones equivocadas (errores) en un dataset de 1000 muestras

$$\hat{p} = Ne \,/\, N = 50/1000 = 0.05$$

Con una confianza del 95% podemos asegurar que el verdadero error es:

$$p \; = \; 0.05 \pm 1{,}96 \sqrt{\frac{0{,}05 \cdot 0{,}95}{1000}} \; = \; 0{,}05 \pm 0{,}014 \; (5\,\% \pm 1{,}4\,\%)$$

Intervalo de confianza

55 $\square$ La probabilidad de error de un clasificador se estima que es del 14 %. Determina cuál es el número mínimo de muestras de test necesario, M, para conseguir que el intervalo de confianza al 95 % de dicho error no supere el $\pm 1\,\%$; esto es, $I = [13\,\%, 15\,\%]$:

A) $M < 2000$.

B) $2000 \leq M < 3500$.

C) $3500 \leq M < 5000$.

D) $M \geq 5000$.

52 $\square$ Se tiene un problema de clasificación para el cual se ha aprendido un clasificador. Asimismo, se tiene un conjunto de $M = 100$ muestras de test con el cual se ha estimado:

- La probabilidad de error del clasificador aprendido, $\hat{p} = 0.10 = 10\,\%$.
- Un intervalo de confianza al 95 % para dicha probabilidad de error, $\hat{I} = [0.04, 0.16] = [4\,\%, 16\,\%]$.

Se considera que la probabilidad de error estimada es razonable y que la misma no variará significativamente aunque usemos muchas más muestras de test. Ahora bien, el intervalo de confianza (al 95 %) estimado, $\hat{I} = 10\,\% \pm 6\,\%$, nos parece un poco amplio y nos preguntamos si es posible reducir su amplitud mediante el uso de más de $M = 100$ muestras de test. Además, si ello fuera posible, nos preguntamos si sería posible reducir dicha amplitud a la mitad o menos; esto es, tal que $\hat{I} = 10\,\% \pm \hat{R}$ con $\hat{R} \leq 3\,\%$. En relación con estas cuestiones, indica cuál de las siguientes afirmaciones es correcta.

A) En general, no es posible reducir la amplitud de $\hat{I}$ pues $\hat{I}$ no depende significativamente de M.

B) No es posible reducir la amplitud de $\hat{I}$ ya que hemos considerado que $\hat{p}$ no variará significativamente y, siendo así, la amplitud de $\hat{I}$ tampoco puede variar significativamente.

C) Sí es posible reducir la amplitud de $\hat{I}$, a la mitad o menos, si doblamos M al menos ($M \geq 200$).

D) Sí es posible reducir la amplitud de $\hat{I}$, a la mitad o menos, si empleamos al menos cuatro veces más muestras de test aproximadamente ($M \geq 400$).

45 $\square$ Supóngase que se está aplicando el algoritmo Perceptrón con $b = 1.5$ y que los vectores de pesos actuales de las clases son los dados en la cuestión anterior. Asimismo, supóngase que el objeto $\mathbf{x} = (1, 2, 2)^t$ dado en la cuestión anterior es la siguiente muestra de entrenamiento a procesar, la cual suponemos perteneciente a la clase 3. Entonces:

A) Se modificarán los vectores de pesos $\mathbf{a}_2$, $\mathbf{a}_3$ y $\mathbf{a}_4$.
B) Se modificará sólo el vector de pesos $\mathbf{a}_3$.
C) No se modificará ningún vector de pesos.
D) Se modificarán todos los vectores de pesos.

$$\mathbf{a}_1 = (-2, 1, 2, 0)^t \qquad \mathbf{a}_2 = (0, 2, 2, 0)^t \qquad \mathbf{a}_3 = (1, 1, 1, 0)^t \qquad \mathbf{a}_4 = (3, 0, 0, 2)^t$$

Clase 3:

$$(1 \quad 1 \quad 1 \quad 0) \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix} = 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 2 + 0 \cdot 2 = 4$$

Clase 1:

$$(-2 \quad 1 \quad 2 \quad 0) \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix} + 1.5 = 3 + 1.5 = 4.5$$

45 ☐ Supóngase que se está aplicando el algoritmo Perceptrón con $b = 1.5$ y que los vectores de pesos actuales de las clases son los dados en la cuestión anterior. Asimismo, supóngase que el objeto $\mathbf{x} = (1, 2, 2)^t$ dado en la cuestión anterior es la siguiente muestra de entrenamiento a procesar, la cual suponemos perteneciente a la clase 3. Entonces:

A) Se modificarán los vectores de pesos $\mathbf{a}_2$, $\mathbf{a}_3$ y $\mathbf{a}_4$.
B) Se modificará sólo el vector de pesos $\mathbf{a}_3$.
C) No se modificará ningún vector de pesos.
D) Se modificarán todos los vectores de pesos.

Clase 2:

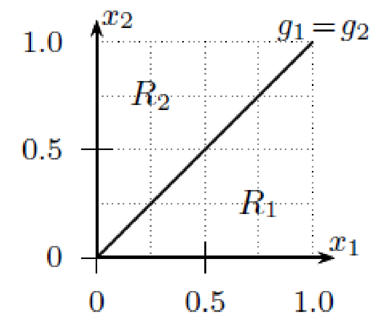
$$(0 \quad 2 \quad 2 \quad 0) \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix} + 1.5 = 6 + 1.5 = 7.5$$

Respuesta D

Clase 4:

$$(3 \quad 0 \quad 0 \quad 2) \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix} + 1.5 = 7 + 1.5 = 8.5$$

50 $\square$ Durante la aplicación del algoritmo Perceptrón ($\alpha = 1.0$ y $b = 0$) en un problema de clasificación en dos clases, se han obtenido los vectores de pesos $\mathbf{w}_1 = (-1, 1, 0)^t$ y $\mathbf{w}_2 = (1, 0, 1)^t$. Supón que el siguiente paso en la aplicación de Perceptrón consiste en procesar una cierta muestra de entrenamiento $\mathbf{x}$ de clase c. Indica cuál de las siguientes opciones daría como resultado un conjunto de pesos que define la frontera y regiones de decisión de la figura de la derecha.



A) $\mathbf{x} = (-1, 1)^t$ y $c = 2$.

B) $\mathbf{x} = (0, 0)^t$ y $c = 2$.

C) $\mathbf{x} = (-1, 1)^t$ y $c = 1$

D) $\mathbf{x} = (0, 0)^t$ y $c = 1$.

67 ☐ Supóngase que estamos aplicando el algoritmo Perceptrón, con factor de aprendizaje $\alpha = 1$ y margen $b = 0.1$, a un conjunto de 3 muestras bidimensionales de aprendizaje para un problema de 2 clases. Tras procesar las primeras 2 muestras se han obtenido los vectores de pesos $\mathbf{w}_1 = (0, 0, -2)^t$, $\mathbf{w}_2 = (0, 0, 2)^t$. A continuación, se procesa la última muestra $(\mathbf{x}_3, c_3)$ y se obtienen los vectores de pesos $\mathbf{w}_1 = (1, 1, -1)^t$, $\mathbf{w}_2 = (-1, -1, 1)^t$, ¿cuál de las siguientes es esa última muestra?

A) $((2, 3)^t, 1)$

B) $((1, 1)^t, 1)$

C) $((2, 1)^t, 2)$

D) $((2, 5)^t, 2)$

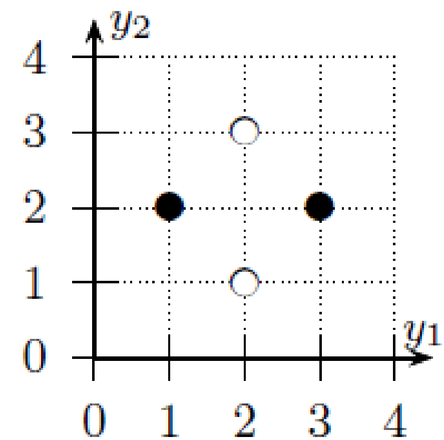
23 ☐ En la figura de la derecha se representan cuatro muestras de aprendizaje bidimensionales de 2 clases: $\circ$ y $\bullet$. A estas muestras se les aplica el algoritmo Perceptrón con pesos iniciales $\mathbf{a}_\circ = (0, 1, 0)^t$ y $\mathbf{a}_\bullet = (0, 0, 1)^t$, una constante de aprendizaje $\alpha > 0$ y un margen b. Indica cuál de las siguientes afirmaciones es correcta:

A) El algoritmo convergerá para algún $b > 0$.

B) El algoritmo solo puede converger si $b \leq 0$.

C) Si $b > 0$, no hay convergencia, pero se puede ajustar el valor de α tal que, tras un número finito de iteraciones, se obtengan buenas soluciones (con $25\,\%$ de error de resustitución).

D) El algoritmo no es aplicable a eastas muestras porque no son linealmente separables.



47 $\square$

Sea $S = \{(\mathbf{y}_1, c_1), \ldots, (\mathbf{y}_N, c_N)\}$, $1 \leq c_j \leq C, 1 \leq j \leq N$, un conjunto *linealmente separable* de muestras representativas de aprendizaje en notación homogénea. S se usa como entrada al algoritmo *Perceptrón*, inicializado con $\mathbf{a}'_j = \mathbf{0}$, $1 \leq j \leq C$. Tras un número suficientemente grande de iteraciones con un factor de aprendizaje $\alpha = 1$ y margen $b = 10$, el algoritmo termina y obtiene C vectores de pesos, en notación homogénea, $\mathbf{a}_j$, $1 \leq j \leq C$. Indica cuál de las siguienes afirmaciones es *correcta*.

A) Se cumplen las $N \cdot C$ inecuaciones:
$\mathbf{a}_i^t \mathbf{y}_i > \mathbf{a}_j^t \mathbf{y}_j \; 1 \leq i \leq N, \; 1 \leq j \leq C, \; i \neq j$

B) Las N muestras de S se clasifican correctamente; es decir, se cumplen las N inecuaciones:
$\mathbf{a}_{c_i}^t \mathbf{y}_i > \mathbf{a}_j^t \mathbf{y}_i, \; 1 \leq i \leq N, \; j \neq c_i$

C) Las N muestras de S se clasifican correctamente, es decir, se cumplen las $N \cdot (C-1)$ inecuaciones:
$\mathbf{a}_{c_i}^t \mathbf{y}_i > \mathbf{a}_j^t \mathbf{y}_i, \; 1 \leq i \leq N, \; 1 \leq j \leq C, \; j \neq c_i$

D) Aunque S es separable, como $b \gg \alpha > 0$, no se puede afirmar que todas las muestras de S se clasifiquen correctamente con los vectores de pesos $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_C$.