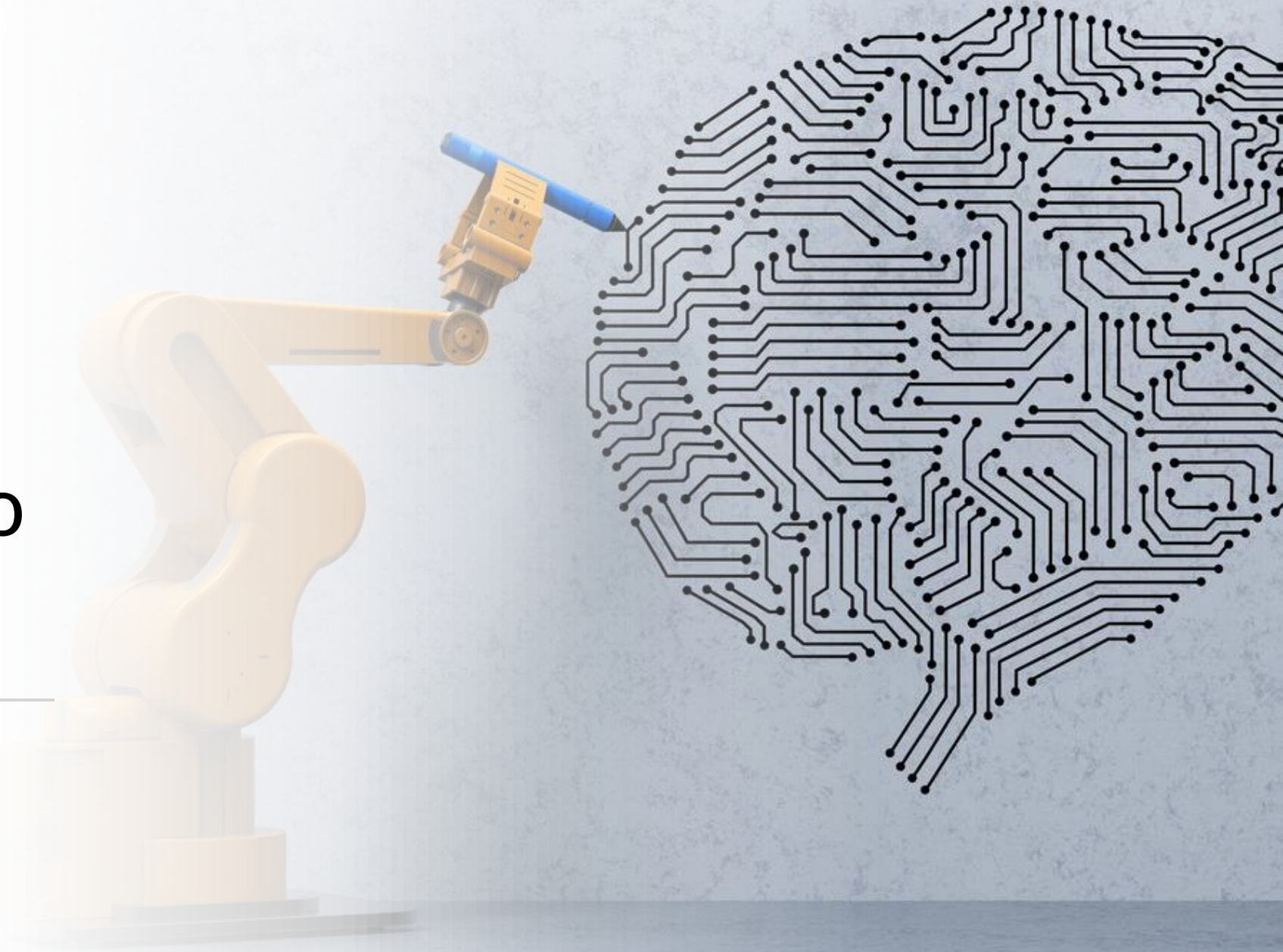




Reconocimiento de Patrones

Clase 8

Funciones de decisión

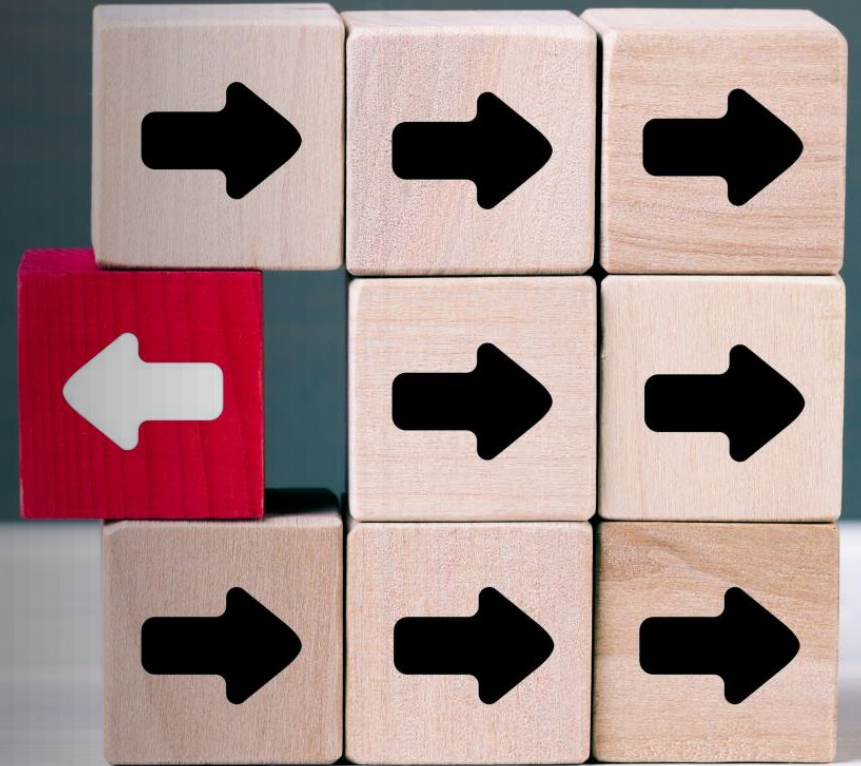


Antes de empezar

- ¿Dudas de la clase anterior?



Funciones de decisión



¿Qué es una función de decisión?

Definición: Una función de decisión $d(x)$ es una función matemática que toma un patrón x como entrada y produce un valor que determina a qué clase pertenece.

$d(x) > 0 \rightarrow \text{Clase } C_1$

$d(x) < 0 \rightarrow \text{Clase } C_2$

Entrada

Vector de características
 $x \in \mathbb{R}^n$ que representa el
patrón a clasificar

Proceso

Aplicación de la función:
evalúa $d(x)$ calculando
una combinación de las
características

Salida

Valor escalar cuyo signo o
magnitud determina la
clase asignada al patrón

Tipos de Funciones de Decisión

- **Lineal**

$$d(x) = w_1x_1 + w_2x_2 + w_0 = w^T x + w_0$$

✓ Simple, rápida, interpretable

✗ Solo funciona si las clases son separables por una línea/plano

- **Cuadrática**

$$d(x) = x^T A x + b^T x + c$$

✓ Más flexible que la lineal

✗ Más parámetros a estimar, riesgo de sobreajuste

- **No lineal**

$$d(x) = f(x) \text{ [forma arbitraria]}$$

✓ Máxima flexibilidad

✗ Más difícil de entrenar e interpretar

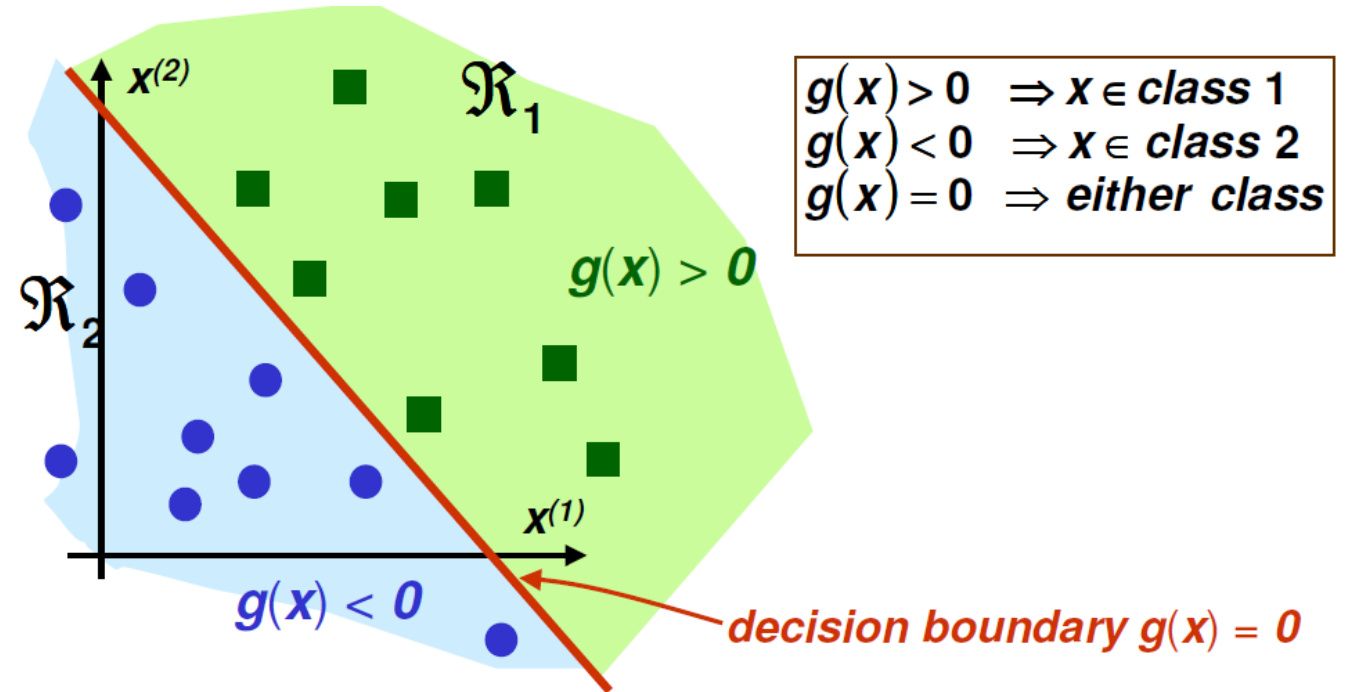
Lineales

- De forma general:

$$d(x) = w_1x_1 + w_2x_2 + \dots + w_nx_n + w_0$$

$$d(x) = \mathbf{w}^T \mathbf{x} + w_0$$

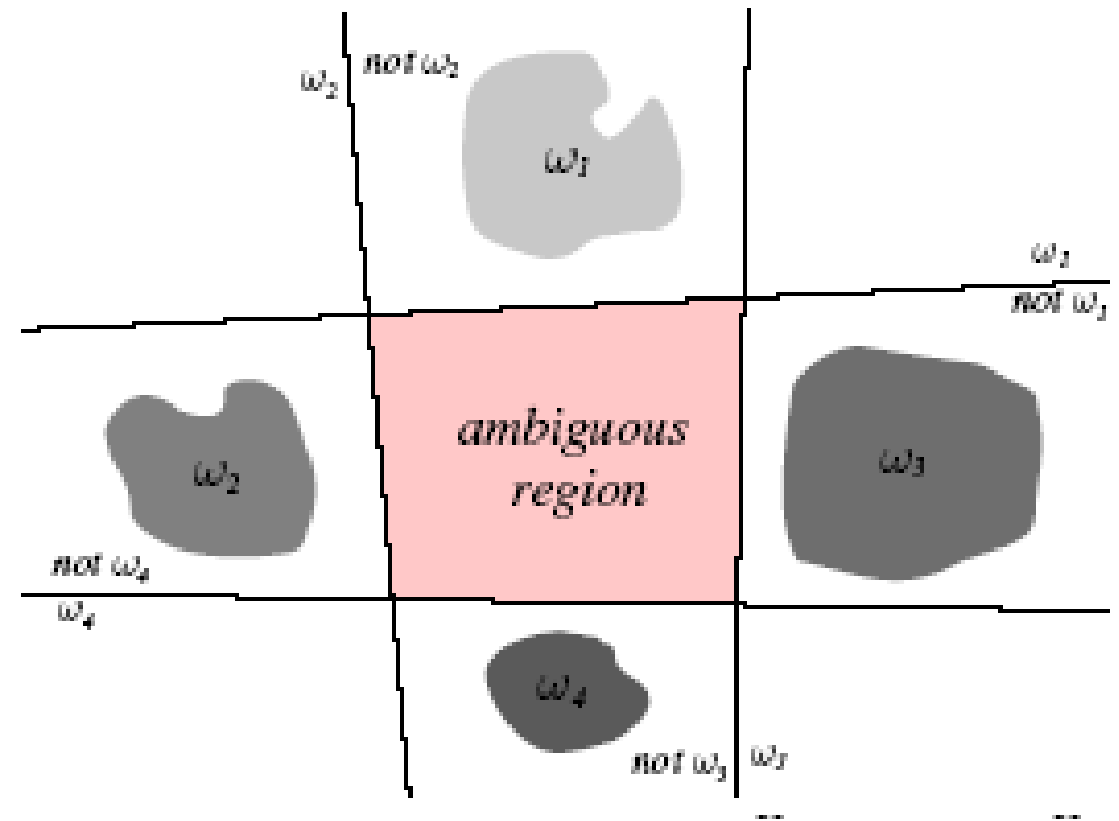
el signo de $d(x)$ determina la clase



¿Qué pasa si tenemos más de dos clases?

- Uno vs el resto
- Cada clase de patrón es separable de las otras clases por una superficie de decisión simple
- Hay M funciones de decisión con la propiedad:

$$d_i(x) = \begin{cases} > 0 & \text{si } x \in w_i \\ < 0 & \text{en otro caso} \end{cases}$$

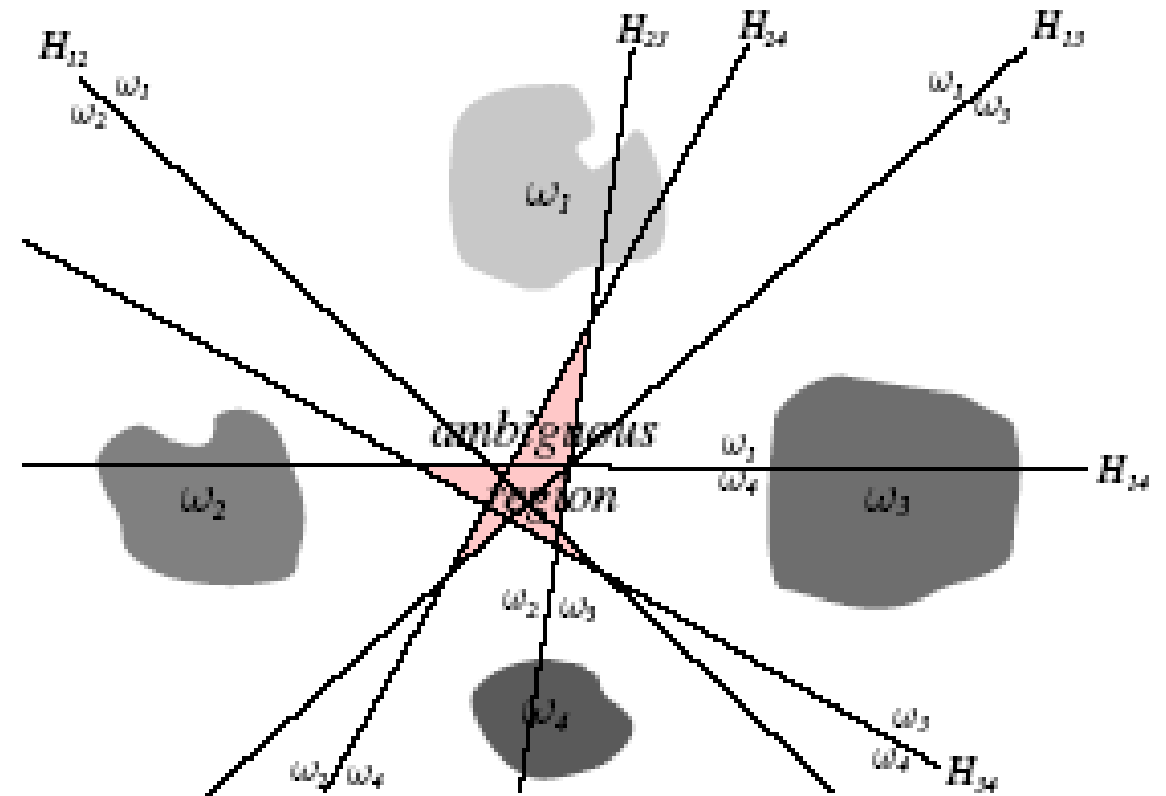


- Uno vs otro

Se define una función de decisión que separa la clase ω_i de cada una de las otras clase de forma individual. Las funciones de decisión tienen la forma:

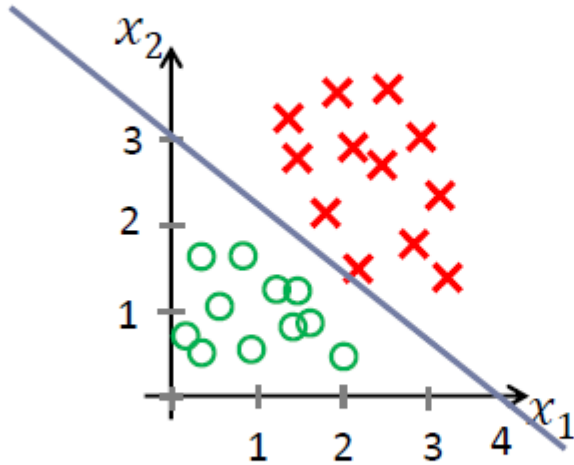
$$d_{ij}(x) = W'_{ij}X$$

Donde $d_{ij}(x) > 0$ para todo $i \neq j$



Ejemplo

$$3 + \frac{3}{4}x_1 + x_2 = 0$$



$$\mathbf{x} = (4, 3)$$

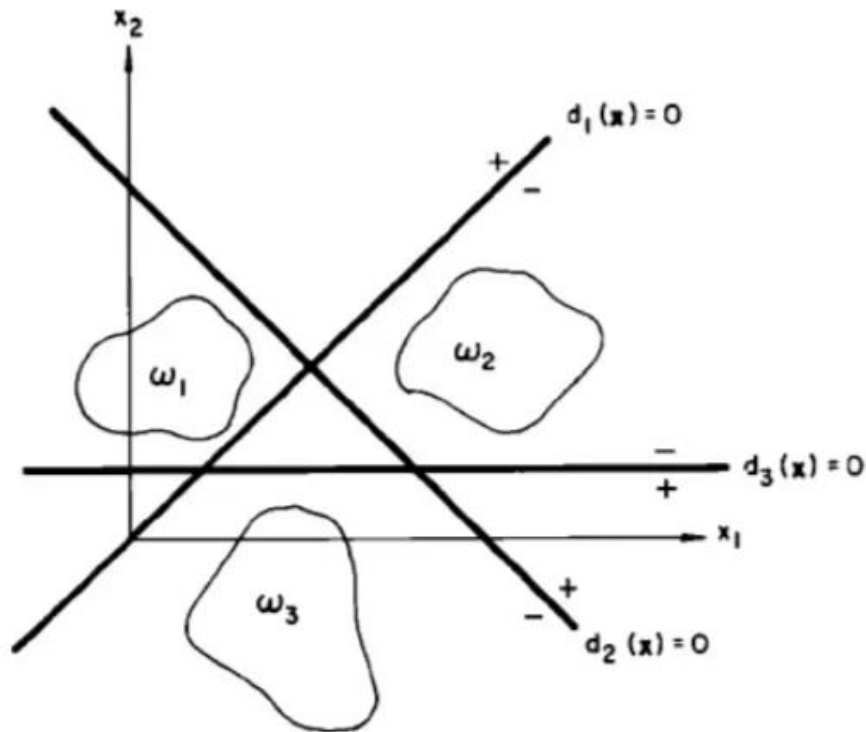
$$\mathbf{x} = (0, 0)$$

if $\mathbf{w}^T \mathbf{x} \geq 0$ ω_1
else ω_2

$$\mathbf{w} = [3, 0.75, 1]$$

$$\mathbf{x} = [1, x_1, x_2]$$

Ejemplo. Supón la siguiente región:



$$d_1(x) = -x_1 + x_2$$

$$d_2(x) = x_1 + x_2 - 5$$

$$d_3(x) = -x_2 + 1$$

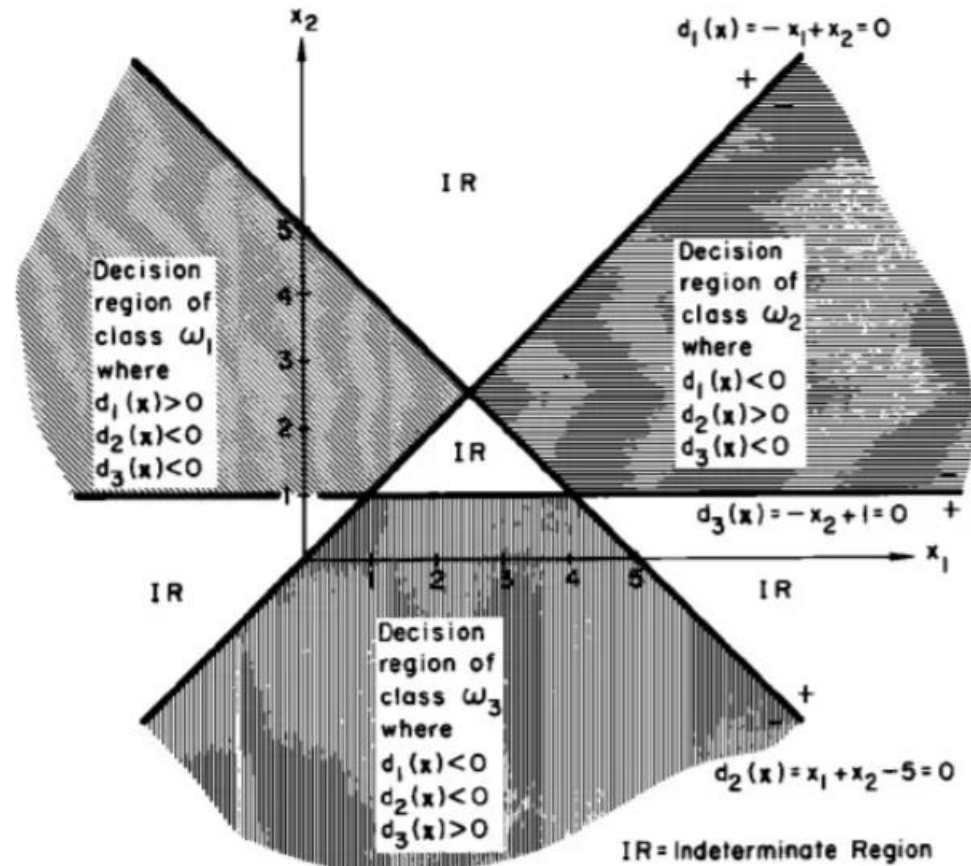
$$X = (6, 5)$$

$$d_1(x) = -1, d_2(x) = 6, d_3(x) = -4$$

$$x = (0, 2)$$

$$x = (6, 4)$$

$$x = (2, -2)$$



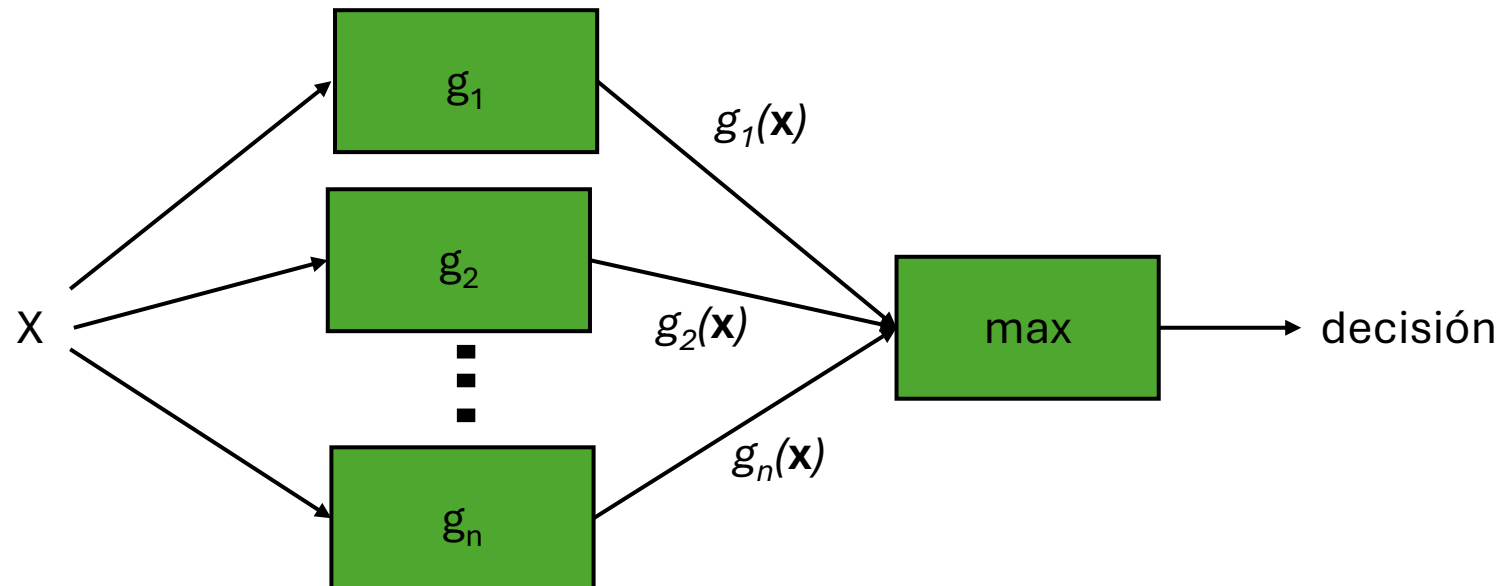
$$x = (2, 6)$$

$$x = (10, -2)$$

$$x = (-6, -2)$$

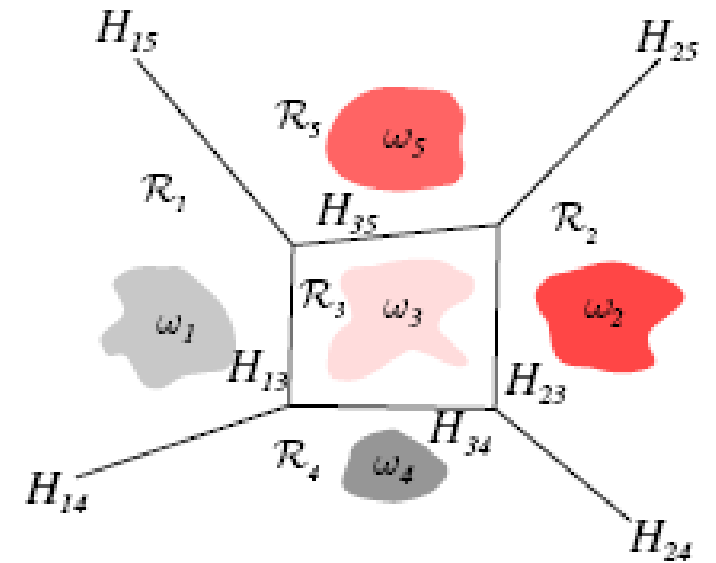
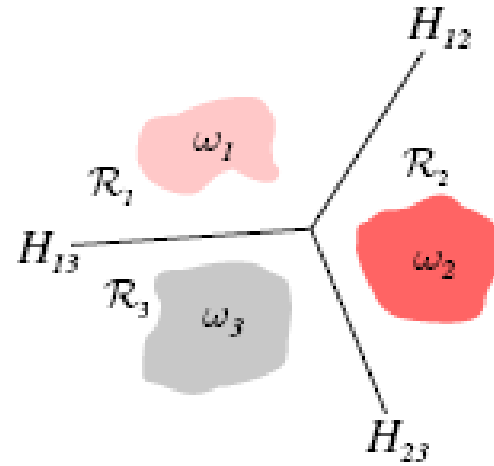
Linear Machine

- Para solventar el problema de regiones ambiguas:
 - Definir c funciones discriminantes lineales
 - Asignar x a ω_i si $g_i(\mathbf{x}) > g_j(\mathbf{x})$ para todo $j \neq i$



- Un maquina lineal divide el espacio de características en c regiones de decisión convexas

- Si $\mathbf{x}$ está en la región R_i , la $g_i(\mathbf{x})$ es la más grande.



Funciones de decisión generalizadas

- Una forma generalizada de considerar las funciones de decisión es:

$$d(x) = w_1 f_1(x) + w_2 f_2(x) + \dots + w_k f_k(x) + w_{k+1} = \sum_{i=1}^{k+1} w_i f_i(x)$$

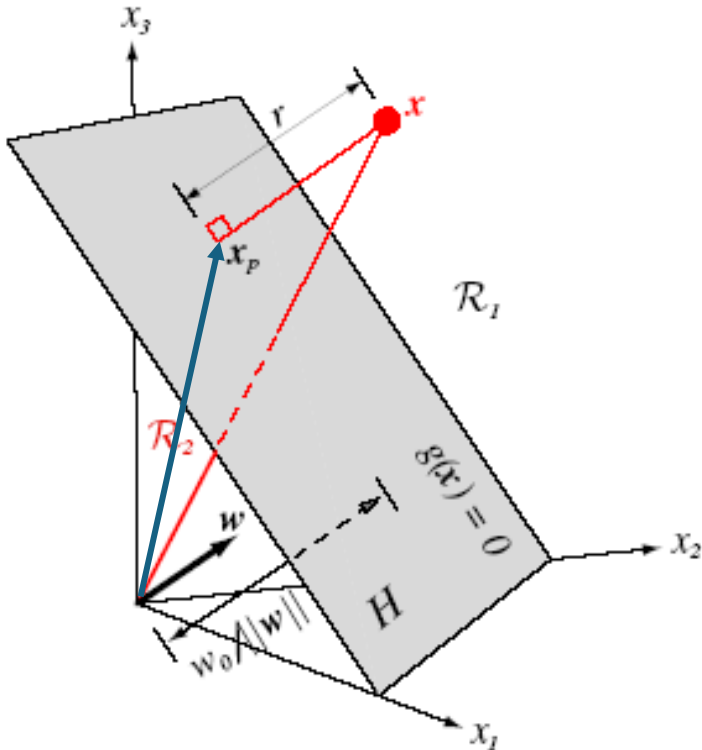
$$g(x) = \sum w_i \phi_i(x)$$

Puede expresarse de forma lineal como:

$$X^* = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_k(x) \\ 1 \end{pmatrix} \quad d(x) = W' X^*$$

$$W = (w_1, w_2, \dots, w_k, w_{k+1})$$

Interpretación geométrica



- $g(x)$ proporciona una medida algebraica de la distancia de x desde el hiperplano.
- x se puede expresar como:

$$\mathbf{x} = \mathbf{x}_p + r \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

- Sustituyendo x en:

$$g(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0$$

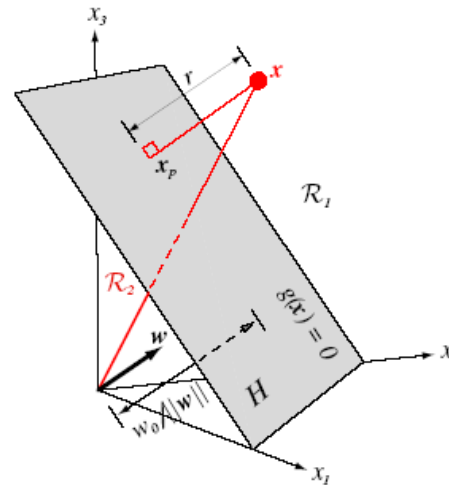
$$g(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0 = \mathbf{w}^t \left(\mathbf{x}_p + r \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) + w_0 =$$

$$= \mathbf{w}^t \mathbf{x}_p + r \frac{\mathbf{w}^t \mathbf{w}}{\|\mathbf{w}\|} + w_0 = r \|\mathbf{w}\|$$

Como: $\mathbf{w}^t \mathbf{w} = \|\mathbf{w}\|^2$ y $\mathbf{w}^t \mathbf{x}_p + w_0 = 0$

$$g(\mathbf{x}) = r \|\mathbf{w}\|$$

La distancia de $\mathbf{x}$ desde el hiperplano



$$r = \frac{g(\mathbf{x})}{\|\mathbf{w}\|}$$

¿Qué es el Espacio de Características?

Es el espacio matemático $\mathbb{R}^n$ en el que cada patrón está representado como un **vector de n características** medibles del objeto a clasificar.

Ejemplo: clasificar frutas

Característica	x_1	x_2	x_3
	Peso (g)	Color (0-1)	Diámetro (cm)
Manzana	182	0.85	7.5
Naranja	210	0.60	8.2
Limón	95	0.72	5.1

$$x = [\text{peso}, \text{color}, \text{diámetro}]^T \in \mathbb{R}^3$$

Dimensión del espacio

Igual al número de características (n). A más características, mayor dimensión.

Punto en el espacio

Cada patrón es un punto. Patrones similares aparecen agrupados en zonas.

Relevancia de características

No todas las características son igualmente útiles para distinguir clases.

Concepto clave: un patrón = un punto en $\mathbb{R}^n$

¿Qué es el Espacio de Pesos?

Es el espacio formado por los **parámetros** $w = [w_0, w_1, \dots, w_n]$ de la función de decisión. Cada punto en este espacio representa un **clasificador diferente**.

Espacio de Características

- Ejes: características $x_1, x_2, \dots, x_n$
- Puntos: patrones (ejemplos)
- Grupos: clases de patrones
- Pregunta: ¿Dónde está cada patrón?

Espacio de Pesos

- Ejes: parámetros $w_1, w_2, \dots, w_n$
- Puntos: clasificadores posibles
- Región: clasificadores válidos (región factible)
- Pregunta: ¿Qué clasificador separa las clases?

Clase, Región de decisión y Frontera de decisión

Clase (C_i)

C

Conjunto de patrones que comparten una categoría.
Ejemplo: C_1 = 'perro', C_2 = 'gato'.

Se busca asignar cada patrón x a la clase correcta

$$x \in C_i \quad \text{si} \quad d_i(x) > d_j(x) \quad \forall j \neq i$$

Región de Decisión (R_i)

R

Subespacio del espacio de características donde todos los patrones son asignados a la clase C_i .

Las regiones particionan todo el espacio sin solapamiento

$$R_i = \{x : d_i(x) > d_j(x) \quad \forall j \neq i\}$$

Frontera de Decisión

F

Superficie que separa dos regiones de decisión. Es donde $d(x) = 0$ o $d_i(x) = d_j(x)$.

Su forma depende del tipo de función: línea (lineal), curva (cuadrática), etc.

$$\text{Frontera: } \{x : d(x) = 0\}$$

Clasificación, Agrupamiento y Reconocimiento

Clasificación

Supervisada

Asignar un patrón a una de las clases predefinidas. Las clases se conocen de antemano y se usan ejemplos etiquetados para entrenar.

Entrada: Patrón x + clases conocidas

Salida: Etiqueta de clase: $C_1, C_2, \dots, C_k$

Ej: Reconocer si un email es spam o no spam

Agrupamiento

No supervisado

Encontrar grupos naturales en los datos SIN clases predefinidas. El algoritmo descubre la estructura por sí mismo.

Entrada: Solo patrones x (sin etiquetas)

Salida: Grupos o clusters (sin nombre)

Ej: Segmentar clientes por comportamiento de compra

Reconocimiento

General

Proceso más amplio que incluye detectar, localizar e identificar patrones. Puede combinar técnicas supervisadas y no supervisadas.

Entrada: Señales, imágenes, texto crudo

Salida: Identidad o descripción del patrón

Ej: Reconocimiento de voz, facial, escritura

Instrumentación de funciones de decisión

1. Inicialización

Se parte de $w = 0$ o valores aleatorios pequeños.

2. Predicción

Se evalúa la función con los pesos actuales.

3. Error

Se compara la predicción con la etiqueta real.

4. Actualización

Si hay error, se ajustan los pesos en la dirección correcta.

5. Convergencia

Se repite hasta que todos los patrones sean bien clasificados.

Pseudocódigo

```
// ENTRENAMIENTO
ENTRADA:  $\{(x_1, y_1), \dots, (x_m, y_m)\}$ ,  $\alpha$ 
INICIALIZAR:  $w \leftarrow$  vector de ceros

REPETIR hasta convergencia:
  PARA cada patrón  $(x_i, y_i)$ :
    calcular  $\hat{y} = \text{signo}(w^T x_i)$ 
    SI  $\hat{y} \neq y_i$ :
       $w \leftarrow w + \alpha \cdot y_i \cdot x_i$  // actualizar

// CLASIFICACIÓN
ENTRADA: nuevo patrón  $x^*$ 
 $val \leftarrow w^T x^*$ 
SI  $val > 0 \rightarrow$  clase  $C_1$ 
SI  $val < 0 \rightarrow$  clase  $C_2$ 
RETORNAR clase asignada
```

α = tasa de aprendizaje (hiperparámetro)

Sistemas de Funciones Ortogonales

Idea central: Representar la función de decisión en una **base ortonormal** $\{\phi_k\}$ donde cada componente captura información **independiente** del patrón.

$$\langle \phi_i, \phi_j \rangle = 0 \quad (i \neq j) \quad \rightarrow \quad \text{Ortogonalidad}$$



Sin redundancia

Cada función base captura información diferente del patrón



Cálculo desacoplado

Los coeficientes w_k se calculan de forma independiente



Compresión eficiente

Pocos coeficientes representan bien el patrón



Estabilidad numérica

Sistema bien condicionado al entrenar el clasificador

Polinomios de Legendre

Dominio: $x \in [-1, 1]$

Peso ortogonal: $w(x) = 1$

Relación de recurrencia: $(n + 1)P_{n+1}(x) = (2n + 1)x P_n(x) - P_{n-1}(x)$

Uso en patrones: Características normalizadas, contornos

Primeros Polinomios

$$P_0(x) = 1$$

$$P_1(x) = x$$

$$P_2(x) = (3x^2 - 1) / 2$$

$$P_3(x) = (5x^3 - 3x) / 2$$

$$P_4(x) = (35x^4 - 30x^2 + 3) / 8$$

• Aplicaciones en Reconocimiento

- Clasificación de formas geométricas normalizadas.
- Descriptores de contorno en visión computacional
- Análisis de perfiles de intensidad en imágenes
- Cualquier señal acotada en un intervalo finito

Polinomios de Laguerre

Dominio: $x \in [0, +\infty)$

Peso ortogonal: $w(x) = e^{-x}$

Relación de recurrencia: $(n+1)L_{n+1}(x) = (2n+1-x)L_n(x) - nL_{n-1}(x)$

Uso en patrones: Señales positivas con decaimiento exponencial

Primeros Polinomios

$$L_0(x) = 1$$

$$L_1(x) = 1 - x$$

$$L_2(x) = (x^2 - 4x + 2) / 2$$

$$L_3(x) = (-x^3 + 9x^2 - 18x + 6) / 6$$

• Aplicaciones en Reconocimiento

- Clasificación de señales de radar (decaimiento)
- Reconocimiento en señales biomédicas (ECG, EEG)
- Señales con distribución exponencial o gamma
- Tiempos de vida y duraciones entre eventos

Polinomios de Hermite

Dominio: $x \in (-\infty, +\infty)$

Peso ortogonal: $w(x) = e^{-x^2}$

Condición: $H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$

Uso en patrones: Datos de distribución gaussiana (muy frecuente)

Primeros Polinomios

$$H_0(x) = 1$$

$$H_1(x) = 2x$$

$$H_2(x) = 4x^2 - 2$$

$$H_3(x) = 8x^3 - 12x$$

$$H_4(x) = 16x^4 - 48x^2 + 12$$

• Aplicaciones en Reconocimiento

- Reconocimiento de voz y procesamiento de audio
- Visión computacional
- Clasificación con datos de distribución normal

Comparación: Legendre, Laguerre y Hermite

Característica	Legendre	Laguerre	Hermite
Dominio	$[-1, 1]$	$[0, +\infty)$	$(-\infty, +\infty)$
Peso $w(x)$	1	e^{-x}	e^{-x^2}
Tipo de datos	Acotados/norm.	Positivos	Cualquier valor
Distribución ideal	Uniforme	Exponencial/gamma	Gaussiana (normal)
Aplicación típica	Formas, contornos	Señales de decaimiento	Voz, imágenes, audio
Ventaja clave	Estabilidad num.	Modela decaimiento	Compatible con normal

Regla de selección: elige el polinomio cuyo dominio y función de peso se ajusten a la distribución natural de tus datos.

Para la próxima clase...

Clasificación de patrones por
medio de funciones de distancia.

