IIC 2433 Minería de Datos

https://github.com/marcelomendoza/IIC2433

- CLASIFICACIÓN -

Clasificación

¿Cuáles de estos ejemplos son árboles?








Clasificación

$input\ \mathbf{x} \in \mathbb{R}^d = \mathcal{X}.$

$output\ y \in \{-1, +1\} = \mathcal{Y}.$ (Clasificación binaria)

$target\ function\ f : \mathcal{X} \mapsto \mathcal{Y}.$ ← f desconocida

$data\ set\ \mathcal{D} = (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N).$ ← Escenario supervisado

Objetivo de aprendizaje:

Queremos **aprender** f usando $\mathcal{D}$.

Clasificación

El algoritmo de aprendizaje

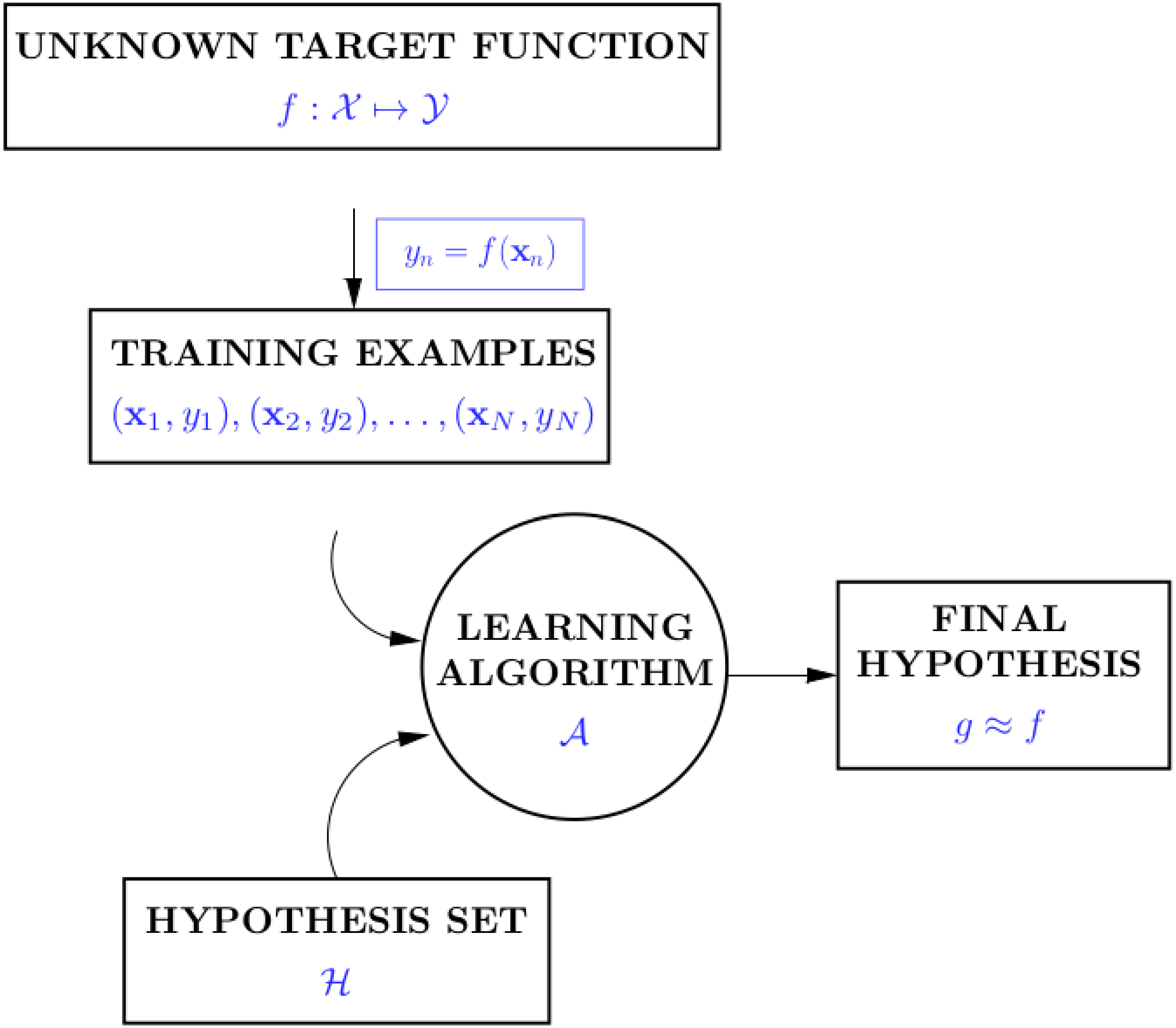
- Comenzamos con un conjunto de hipótesis candidatas $\mathcal{H}$ que son factibles de representar f.

$$\mathcal{H} = \{h_1, h_2, \ldots, \}$$ ← Conjunto de hipótesis

- Un algoritmo $\mathcal{A}$ selecciona una hipótesis g desde $\mathcal{H}$. El algoritmo opera sobre datos etiquetados disponibles para seleccionar g (datos de entrenamiento).

- Para nuevos datos no etiquetados (i.e., y desconocido), usamos g para inferir y.

Clasificación



- PERCEPTRÓN LINEAL -

Un modelo simple (lineal)

age	32 years
gender	male
salary	40,000
debt	26,000
years in job	1 year
years at home	3 years
...	...

¿Sujeto de crédito?

- Vector de características: $\mathbf{x} = [x_1, \ldots, x_d]^{\mathrm{T}}$.

- Aprenderemos distintos 'pesos' para las variables de entrada:

$$\text{“Credit Score”} = \sum_{i=1}^{d} w_i x_i.$$

- Usaremos un **umbral** para decidir si aprobamos el crédito:

Aprobamos si: $\sum_{i=1}^{d} w_i x_i > \text{threshold},$ (score de crédito suficiente)

Rechazamos si: $\sum_{i=1}^{d} w_i x_i < \text{threshold}.$ (score de crédito insuficiente)

Variable útil para mejorar el score: $w_i > 0$
Variable útil para disminuir el score: $w_i < 0$

Un modelo simple (lineal)

La hipótesis puede escribirse formalmente:

$$h(\mathbf{x}) = \text{sign}\left(\left(\sum_{i=1}^{d} w_i x_i\right) + w_0\right)$$

Usamos el signo para tomar la decisión

Bias (el umbral)

- El conjunto de hipótesis es:

$$\mathcal{H} = \{h(\mathbf{x}) = \text{sign}(\mathbf{w}^{\mathrm{T}}\mathbf{x})\}$$

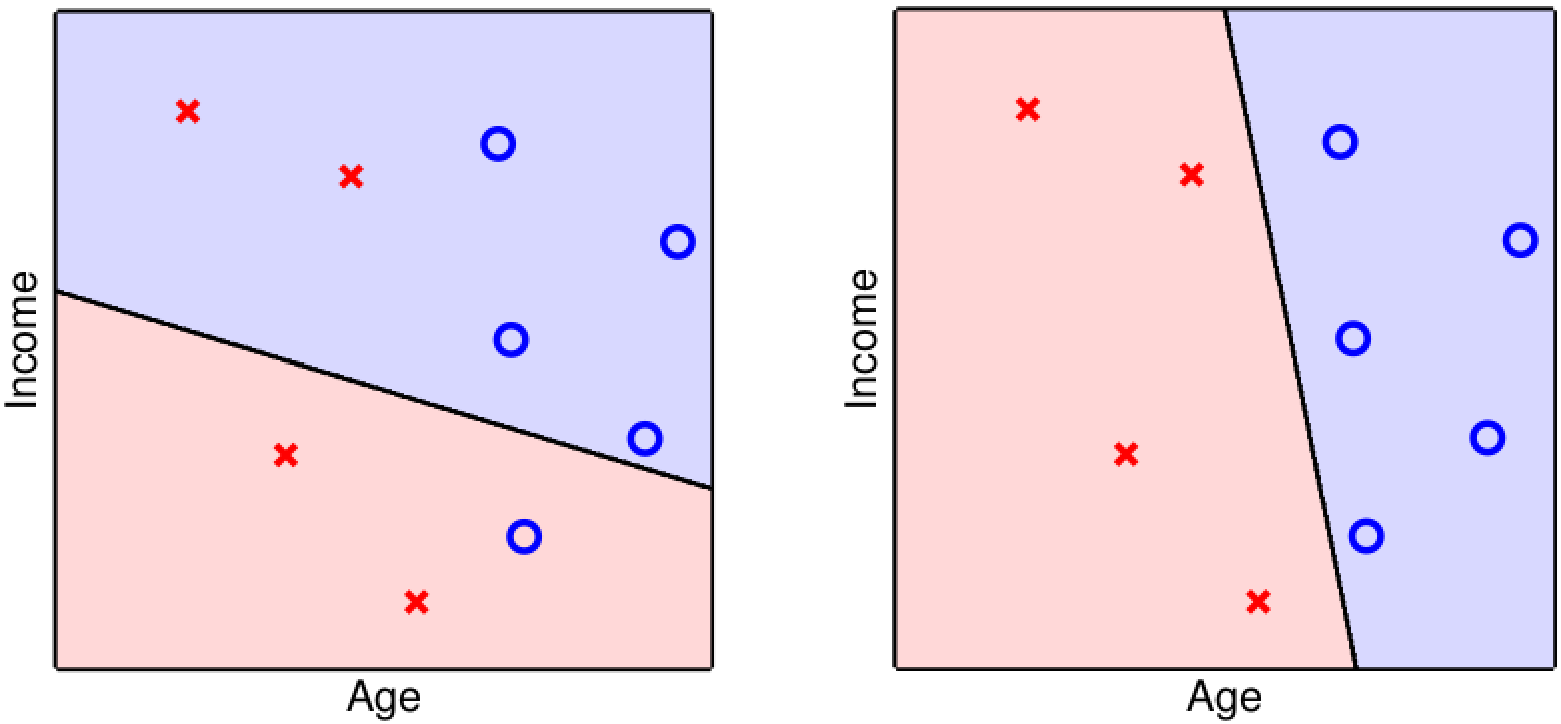
Infinito si los pesos están en $\mathbb{R}$

- Este modelo se llama **Perceptrón**:

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \in \mathbb{R}^{d+1}, \qquad \mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} \in \{1\} \times \mathbb{R}^d.$$

El Perceptrón

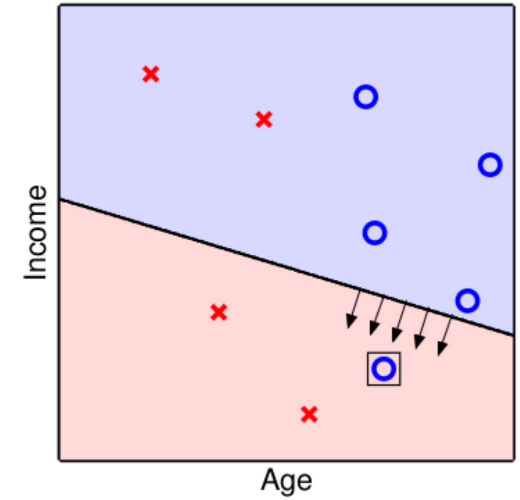
- El Perceptrón usa los datos para encontrar una separación lineal según el atributo de clase y



- Como los datos están en un espacio de representación *d*-dimensional, el separador es un hiperplano.

El Perceptrón

Clasificación binaria: $\mathcal{Y} = \{+1, -1\}$



Enfoque iterativo

$\mathbf{w}(1) = \mathbf{0}$

for iteration $t = 1, 2, 3, \ldots$

the weight vector is $\mathbf{w}(t)$. Random

From $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$ pick any misclassified example.

Call the misclassified example $(\mathbf{x}_*, y_*)$,

$$\text{sign}\left(\mathbf{w}(t) \bullet \mathbf{x}_*\right) \neq y_*.$$

Update the weight:

$$\mathbf{w}(t+1) = \mathbf{w}(t) + y_* \mathbf{x}_*.$$

$t \leftarrow t + 1$

Update rule

El Perceptrón

Clasificación binaria: $\mathcal{Y} = \{+1, -1\}$

- Update rule (en ejemplos mal clasificados):

$$\mathbf{w}(t+1) = \mathbf{w}(t) + y(t)\mathbf{x}(t).$$

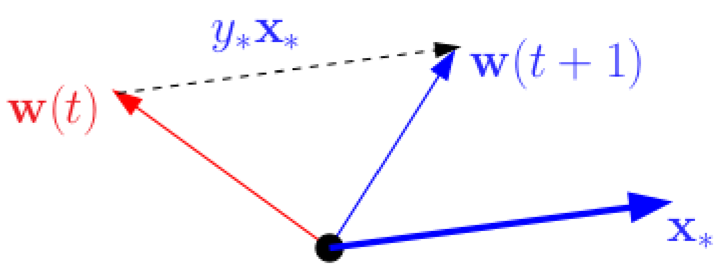
El Perceptrón

Clasificación binaria: $\mathcal{Y} = \{+1, -1\}$

- Update rule (en ejemplos mal clasificados):

$$\mathbf{w}(t+1) = \mathbf{w}(t) + y(t)\mathbf{x}(t).$$

Ground truth $\rightarrow$ $\boxed{y_* = +1}$



Orienta $\mathbf{w}$ hacia la dirección de $\mathbf{x}$

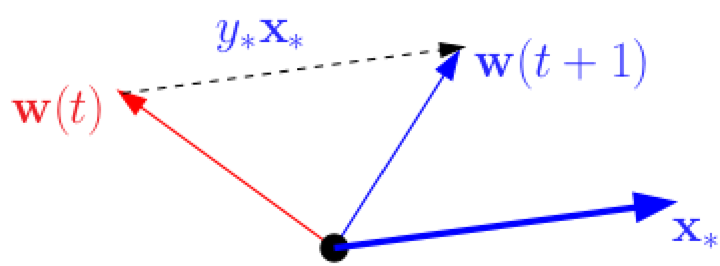
El Perceptrón

Clasificación binaria: $\mathcal{Y} = \{+1, -1\}$

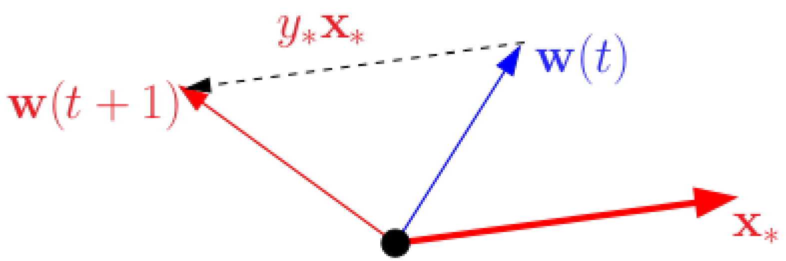
- Update rule (en ejemplos mal clasificados):

$$\mathbf{w}(t+1) = \mathbf{w}(t) + y(t)\mathbf{x}(t).$$

Ground truth $\boxed{y_* = +1}$



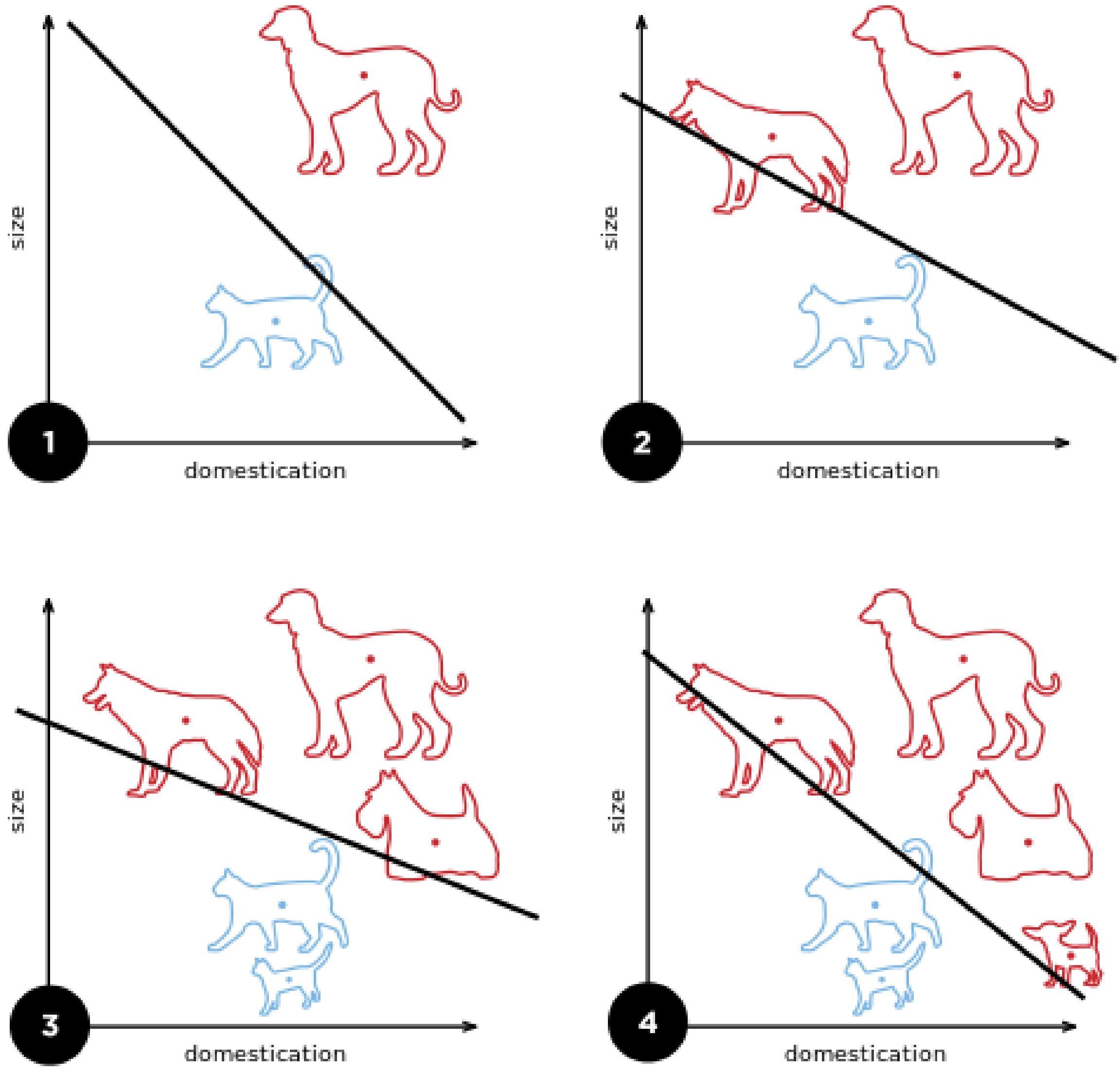
$\boxed{y_* = -1}$ Ground truth



Orienta $\mathbf{w}$ hacia la dirección de $\mathbf{x}$

Orienta $\mathbf{w}$ en la dirección contraria de $\mathbf{x}$

El Perceptrón



- REGRESIÓN LOGÍSTICA -

Modelos lineales

entrada

$$s = \mathbf{w}^{\mathrm{T}} \mathbf{x}$$

parámetros

Modelos lineales

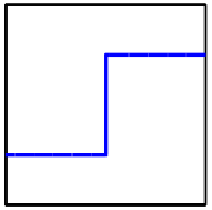
entrada

$$s = \mathbf{w}^{\mathrm{T}}\mathbf{x}$$

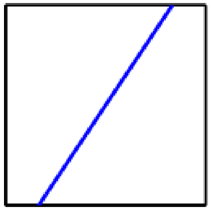
parámetros

$$y = \theta(s)$$

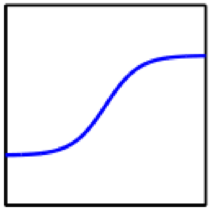
$$s = \mathbf{w}^{\mathrm{T}}\mathbf{x} \longrightarrow$$



$\rightarrow \mathrm{sign}(\mathbf{w}^{\mathrm{T}}\mathbf{x})$ $\{-1, +1\}$



$\rightarrow \mathbf{w}^{\mathrm{T}}\mathbf{x}$ $\mathbb{R}$



$\rightarrow \theta(\mathbf{w}^{\mathrm{T}}\mathbf{x})$ $[0, 1]$

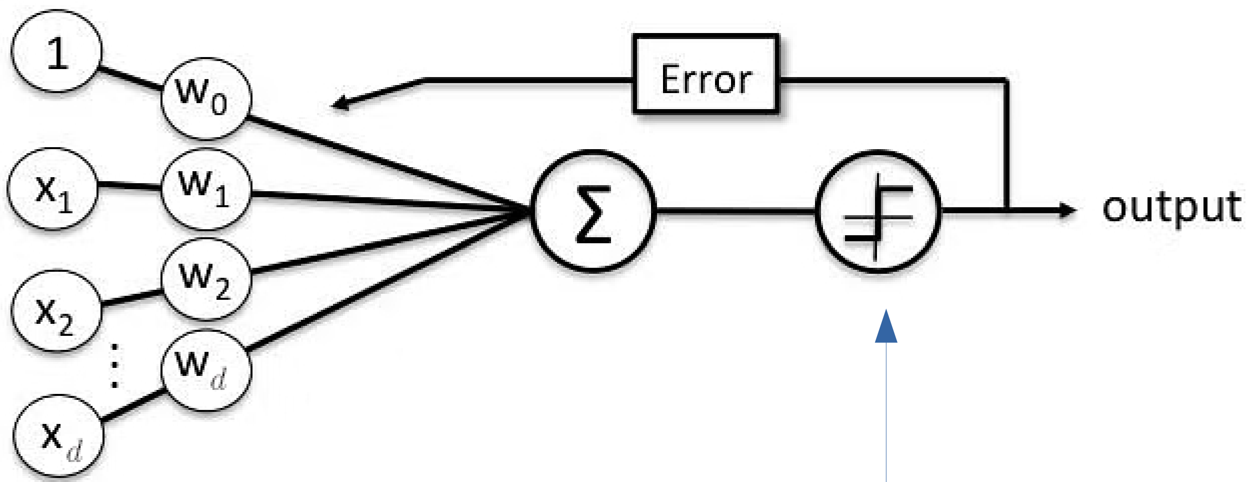
entrada

Perceptrón lineal

$$\mathcal{H}_{\text{lin}} = \{h(\mathbf{x}) = \text{sign}(\mathbf{w}^{\mathrm{T}}\mathbf{x})\}$$

$$s = \mathbf{w}^{\mathrm{T}}\mathbf{x}$$

parámetros



Función signo

Regresión logística

Objetivo:

$$f(\mathbf{x}) = \mathbb{P}[y = +1 \mid \mathbf{x}].$$

Regresión logística

Objetivo:

$$f(\mathbf{x}) = \mathbb{P}[y = +1 \mid \mathbf{x}].$$

Modelo:

$$h(\mathbf{x}) = \theta\left(\sum_{i=0}^{d} w_i x_i\right) = \theta(\mathbf{w}^{\mathrm{T}}\mathbf{x})$$

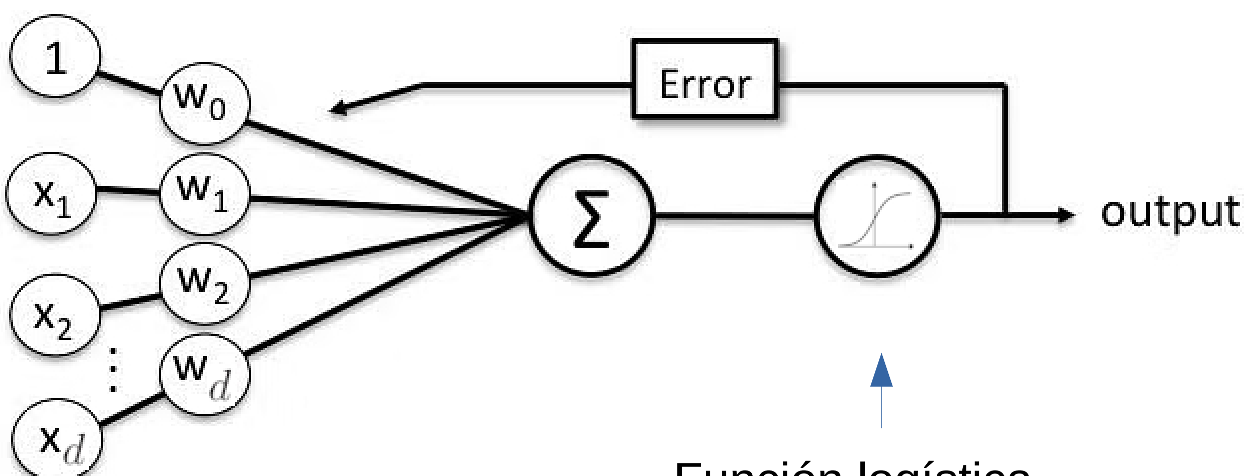
Regresión logística

Objetivo:

$$f(\mathbf{x}) = \mathbb{P}[y = +1 \mid \mathbf{x}].$$

Modelo:

$$h(\mathbf{x}) = \theta\left(\sum_{i=0}^{d} w_i x_i\right) = \theta(\mathbf{w}^{\mathrm{T}}\mathbf{x})$$



Función logística

Regresión logística

Objetivo:

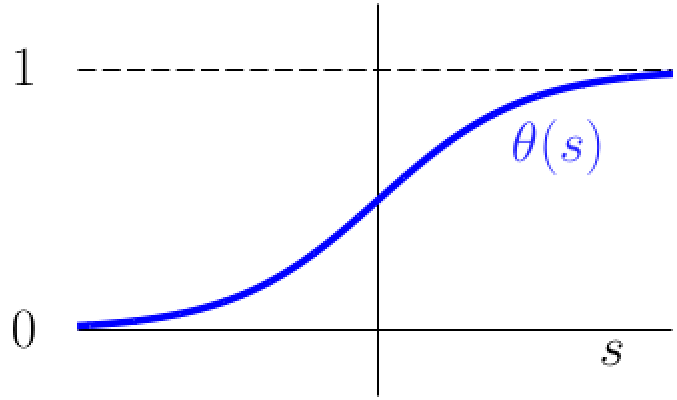
$$f(\mathbf{x}) = \mathbb{P}[y = +1 \mid \mathbf{x}].$$

Modelo:

$$h(\mathbf{x}) = \theta\left(\sum_{i=0}^{d} w_i x_i\right) = \theta(\mathbf{w}^{\mathrm{T}}\mathbf{x})$$

Función logística:

$$y \in [0, 1]$$



$$\theta(s) = \frac{e^s}{1+e^s} = \frac{1}{1+e^{-s}}.$$

$$\theta(-s) = \frac{e^{-s}}{1+e^{-s}} = \frac{1}{1+e^{s}} = 1 - \theta(s).$$

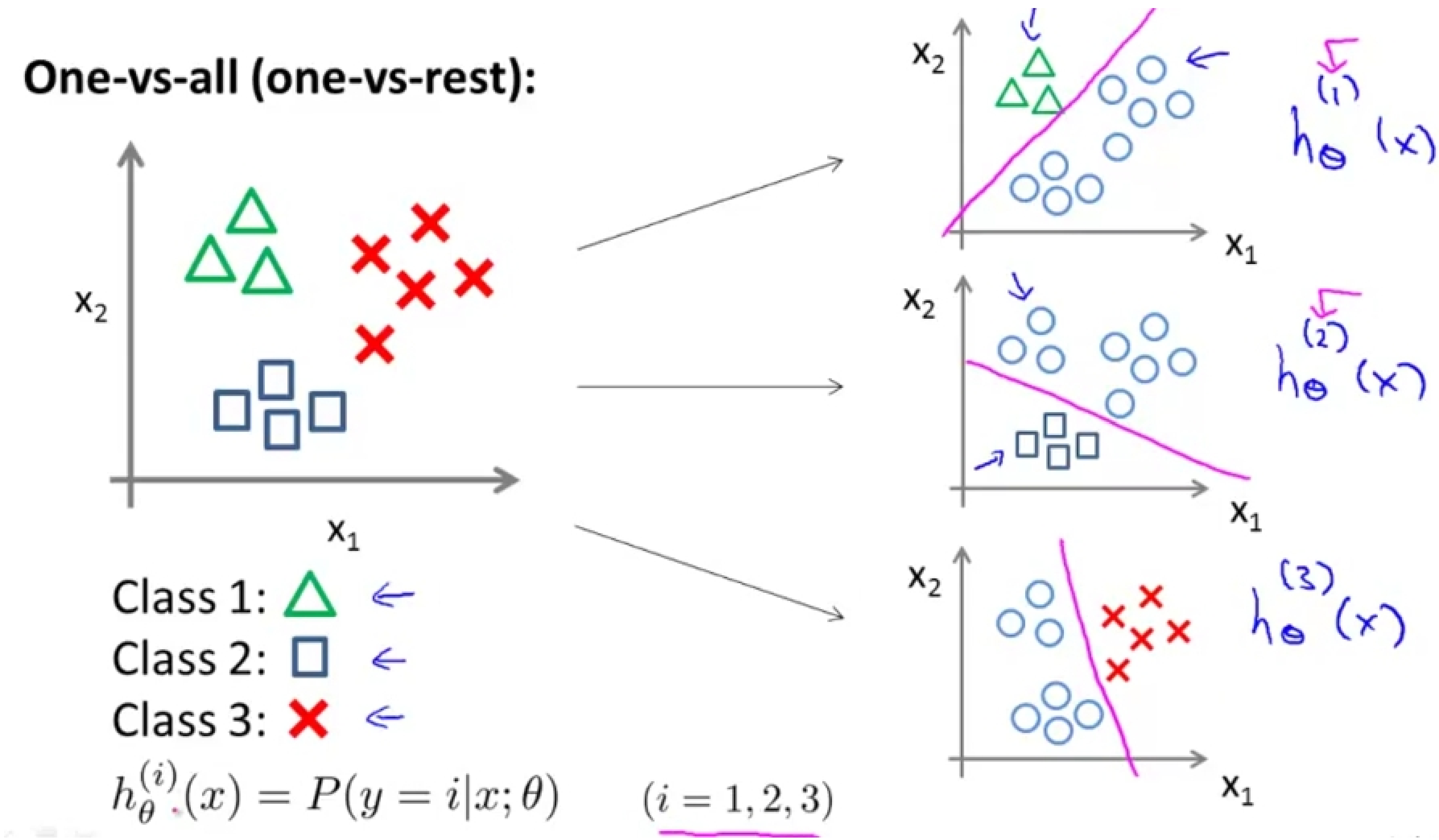
Regresión logística

Notar que: $\mathcal{D} = (\mathbf{x}_1, y_1 = \pm 1), \cdots, (\mathbf{x}_N, y_N = \pm 1)$

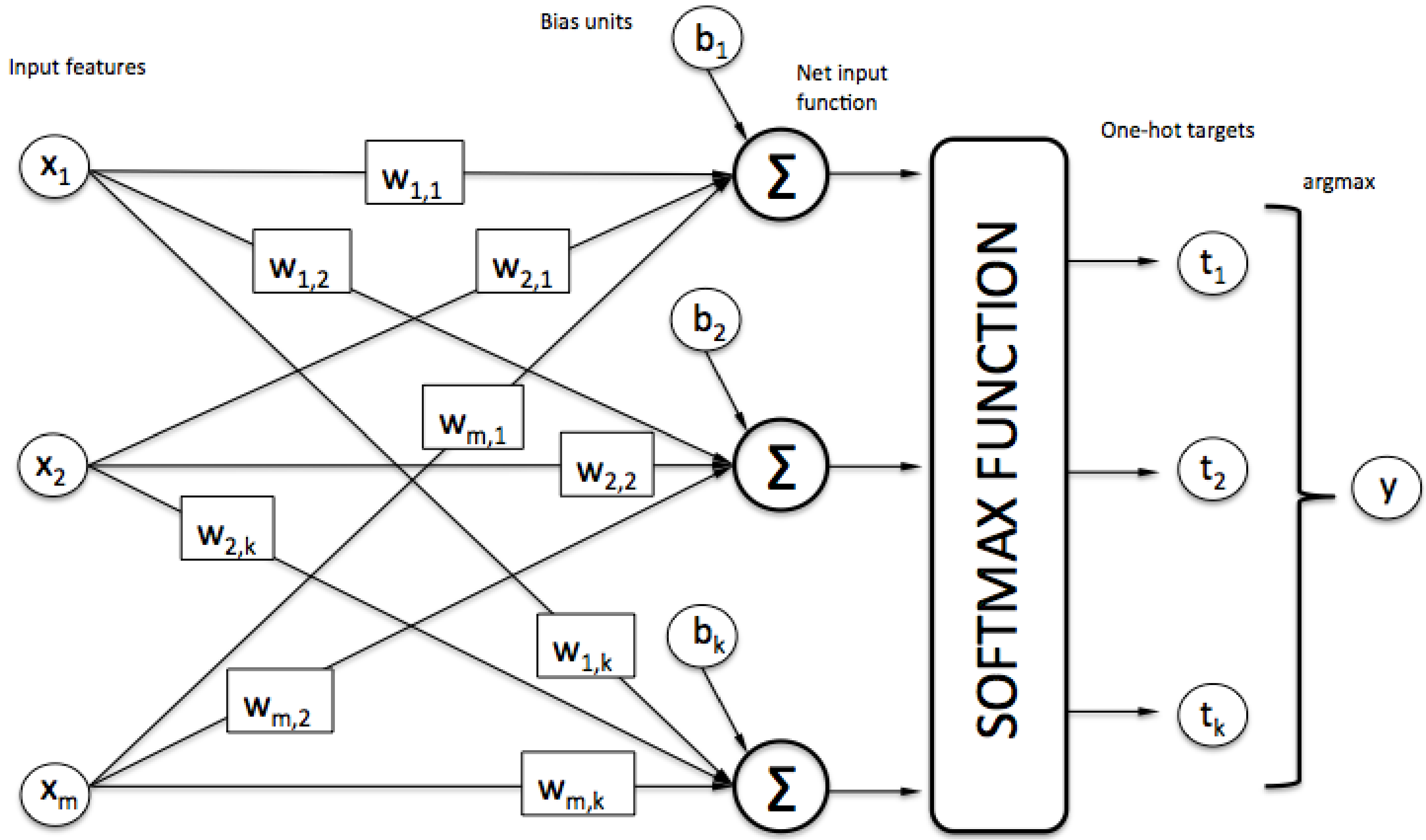
Un buen modelo logra lo siguiente:

$$\begin{cases} h(\mathbf{x}_n) \approx 1 & \text{si } y_n = +1; \\ h(\mathbf{x}_n) \approx 0 & \text{si } y_n = -1. \end{cases}$$

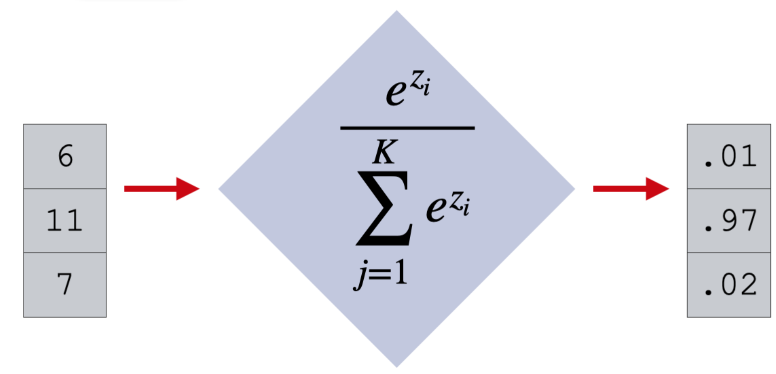
Clasificación multiclase



Clasificación multiclase



$$P(y = j \mid z^{(i)}) = \phi_{softmax}(z^{(i)})$$



- MÉTRICAS DE RENDIMIENTO -

Métricas

+1 -1

+1

-1

Métricas

verdaderos positivos (TP)	falsos positivos (FP)
falsos negativos (FN)	verdaderos negativos (TN)

$$\text{accuracy} = (TP + TN)/(TP + FP + FN + TN)$$

$$\begin{aligned} P &= TP/(TP + FP) \\ R &= TP/(TP + FN) \end{aligned}$$

$$F = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

F-measure balanceada con $\beta = 1$