TEMA 4: ALGORITMOS CLASIFICACIÓN Y REGRESIÓN

↳ SUPERVISADOS:

CLASIFICACIÓN: $X_{NOBS \times NVAR}$ matriz datos y $w_{NOBS \times 1}$ vector etiquetas clase — categórica/discreta

Ej: iris – CLASIFICACIÓN

4 atributos
150 observ.

$X_{150 \times 4}$ y $w_{150 \times 1} = \begin{array}{c c} \begin{bmatrix} \text{'vi'} \\ \text{'ve'} \\ \text{'vi'} \\ \text{'s'} \\ \vdots \end{bmatrix} & \begin{array}{c} 1 \\ \\ \\ \\ 150 \end{array} \end{array}$

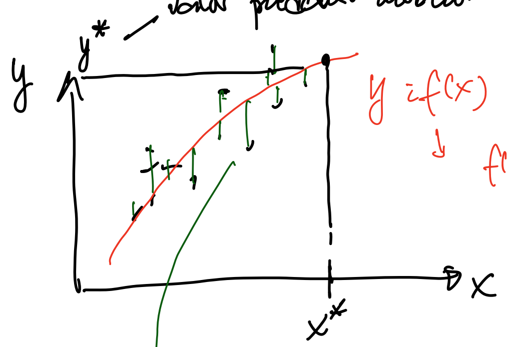
$w_i \in \{$'setosa', 'versicolor', 'virginica'$\}$

$\left\{ \begin{matrix} X \\ w \end{matrix} \right\} \rightarrow \boxed{\text{MODELO}} \rightarrow \hat{w}$ predicción de clase

REGRESIÓN: $X_{NOBS \times NVAR}$ y y — continua, reales

↓ vector de la variable a predecir

valor predicho modelo



$y = f(x)$
↓
$f(x) = \underline{a}x^2 + \underline{b}x + \underline{c}$

parámetros modelo regresión

residuos $= \hat{y} - y$

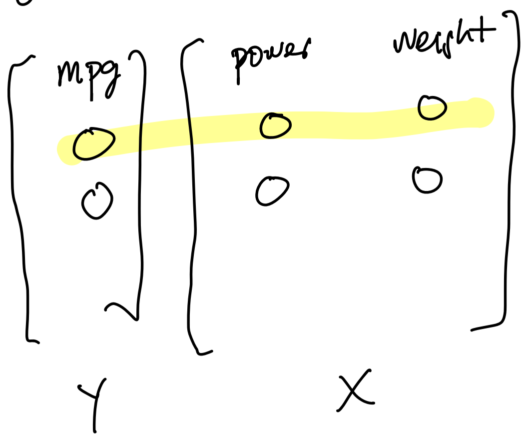
$\hat{y}$ ↓ predicción modelo

y ↘ valores reales (observaciones) → MSE $\xrightarrow{\text{OLS}}$ $\hat{a}, \hat{b}, \hat{c}$

$x^* \rightarrow \boxed{\text{modelo } f(x, \hat{a}, \hat{b}, \hat{c})} \rightarrow y^*$

$\{x, y\} \rightarrow \boxed{\text{MODELO regresión}} \rightarrow y^*$ predicción

Ej: cars database



$x^* = (\text{power}^*, \text{weight}^*) \rightarrow \text{MODELO} \rightarrow \text{mpg}^*$

METODOLOGÍA ENTRENAR / INFERENCIA PARÁMETROS / AJUSTAR

① train / test

$\text{DATOS} = \{ X_{NOBS \times NVAR}, \; w_{NOBS \times 1} \}$

~70% → subconjunto de entrenamiento (construir el modelo)

↓

$\{X_{train}, w_{train}\}$

↓ entrenar

MODELO (θ^*)

~30% división aleatoria → subconjunto test (validar el modelo)

↓

$\{X_{test}, w_{test}\}$

↓

MODELO (θ^*)

↓

$\hat{w}_{test}$: predicción para los datos test parámetros

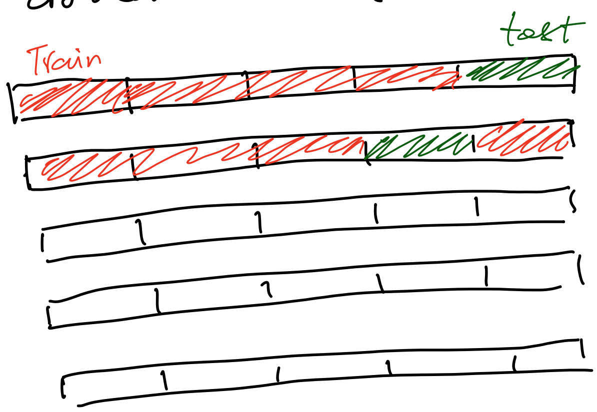
entrenar → modelo: probabilístico
estimar parámetros
geométrico
distancias.
....

COMPARAR ($\hat{\omega}_{test}$, ω_{test})
modelo   real

⇓

MEDIDAS DE RENDIMIENTO
- matriz confusión
- errores tipo I, II
- medidas: Accuracy, Recall, ...

② Validación cruzada k-veces (k-fold cross-validation) -(bootstrapping?)



→ θ_1^* → MODELO$_1$ → $\hat{\omega}_{test}$

→ θ_2^* ...

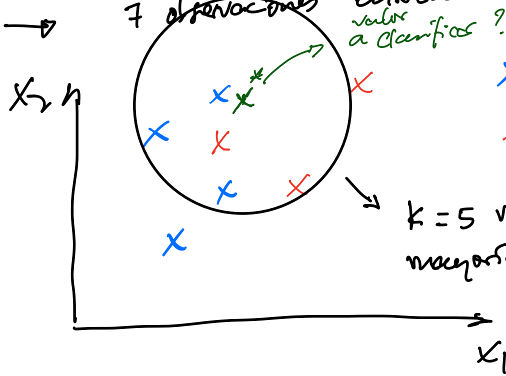
⋮

rendimiento promedio

① K-primeros vecinos (KNN)
② Análisis de discriminantes (probabilístico)
③ Árboles de decisión
④ Redes Neuronales

① KNN (K-primeros vecinos)

Ex: NVAR = 2
NOBS = 10 → 7 observaciones entrenamiento



X - ($\omega = 0$)
X - ($\omega = 1$)

$k = 5$ vecinos
mayoría azules →
$x^* \to \omega = 0$

② Análisis de discriminantes:

Fórmula de Bayes (clasificación)

$$\underbrace{p(\omega|\vec{x})}_{\text{densidad posterior (predictiva de } \omega \text{ en función de } \vec{x})} = \frac{p(\vec{x}|\omega) \cdot p(\omega)}{p(\vec{x})}$$

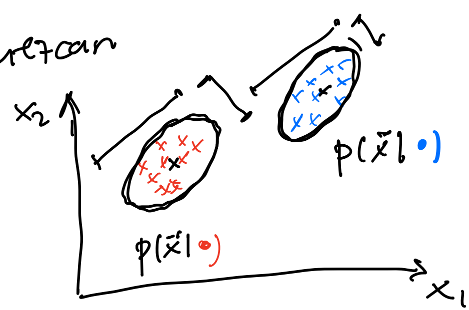
ω: etiqueta; $\vec{x}$: observación

$p(\vec{x}|\omega)$: función de verosimilitud (Likelihood) MODELO

$p(\omega)$: APRIORI (Prior)

$p(\vec{x})$ → constante

$p(\vec{x}|\omega)$: describe cómo se distribuyen los valores de $\vec{x}$ en función de la clase a la que pertenezcan



$p(\vec{x}|\bullet) = \underline{N(\mu_\bullet, \Sigma_\bullet)}$

modelo likelihood normal multivariada

$p(\omega)$: densidad de probabilidad de las clases (sin datos)

↳ experto

Análisis Discriminantes

$$g_i(\vec{x}) = \log \underbrace{p(\omega_i|\vec{x})}_{\text{posterior}} \qquad i = 1 \dots NCLASES$$

$g_i(\vec{x})$ ↓ función discriminante

$\vec{x}$ se asigna a la clase que tenga mayor $g_i(\vec{x})$

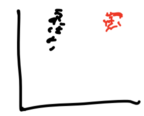
$$\hat{\omega} = \underset{1 \dots NCLASES}{\operatorname{argmax}} \; g_i(\vec{x})$$

$\vec{x}$? $\longrightarrow$ $p(w=0|\vec{x}) = 0.8$ $\longrightarrow$ $\vec{x}$ es de $w=0$ 

$p(w=1|\vec{x}) = 0.2$

$$p(w|\vec{x}) = \underbrace{p(\vec{x}|w)}_{\text{likelihood}} \cdot p(w)$$

likelihood :



① $p(\vec{x}|w_i) = N(\vec{x}|\vec{\mu}_i, \Sigma_i)$ (QDA)

a peor

② $p(\vec{x}|w_i) = N(\vec{x}|\vec{\mu}_i, \Sigma)$ (LDA)

Σ $\downarrow$ iguales en todas las clases

③ (GAUSSIAN NAÏVE BAYES) (GNB)

las variables son independientes $\rightarrow$ hipótesis naive

$\downarrow$

$$p(\vec{x}|w_i) = \prod_{j=1}^{NVAR} \underbrace{p(x_j|w_i)}_{N(x_j|\mu_i,\sigma_i)}$$

Gaussiana univariada

... $\underbrace{p(\vec{x}|w)}$ ...