

# 05 MLAR-Aprendizaje supervisado

## Tema 2 – Validación y evaluación

Curso Abril 23/24



**Universidad**  
Internacional  
de Valencia

# ¿Dónde estudiar el Tema 2?

**Manual de la Asignatura – Tema 2**

**Scripts del Tema 2**

**03. Materiales del profesor > Recursos adicionales > Esquemas > Validación cruzada anidada  
> Pipeline Machine Learning  
> Metodología ML**

**03. Materiales del profesor > Recursos adicionales > Otros > Ejemplo learning curves**

# Índice

2.0. Introducción

2.1. Validación hold-out

2.2. Validación cruzada

2.3. Ajuste de parámetros y validación anidada

2.4. Evaluación en regresión

2.5. Evaluación en clasificación

# Índice

## **2.0. Introducción**

2.1. Validación hold-out

2.2. Validación cruzada

2.3. Ajuste de parámetros y validación anidada

2.4. Evaluación en regresión

2.5. Evaluación en clasificación

## 2.0. Introducción (I)

“A partir de este capítulo en adelante, en esta asignatura nos centraremos únicamente en **problemas de inferencia o predicción de una clase** (siempre dentro del aprendizaje **supervisado**).”

“Dado un mismo conjunto de datos, no todos los algoritmos de inferencia producen las mismas predicciones ... **necesario comparar los resultados de predicción entre diferentes algoritmos** aplicados sobre el mismo conjunto de datos ... durante la fase de **explotación**, se suele utilizar **solo el mejor modelo** conseguido en la fase de modelado.”

- **Conjunto de entrenamiento:** aprendizaje, generar el modelo de conocimiento
- **Conjunto de test:** evaluar la bondad del model

“... **el conjunto de test no debe contener ninguna instancia del conjunto de entrenamiento**, sino que deben ser ejemplos diferentes...”

## 2.0. Introducción (II)

Problemas en el aprendizaje:

- **Sobreajuste (overfitting):** el modelo **no** extrae conocimiento **generalizable**, sino que **memoriza** los datos.
- **Subajuste (underfitting):** el modelo no es capaz de discriminar bien en el conjunto de entrenamiento.

¿Cómo evaluar dichos fenómenos?

- **Error de entrenamiento**
- **Error de generalización**

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$Exactitud = \frac{100}{n} \cdot \sum_{i=1}^n y_i = \hat{y}_i$$

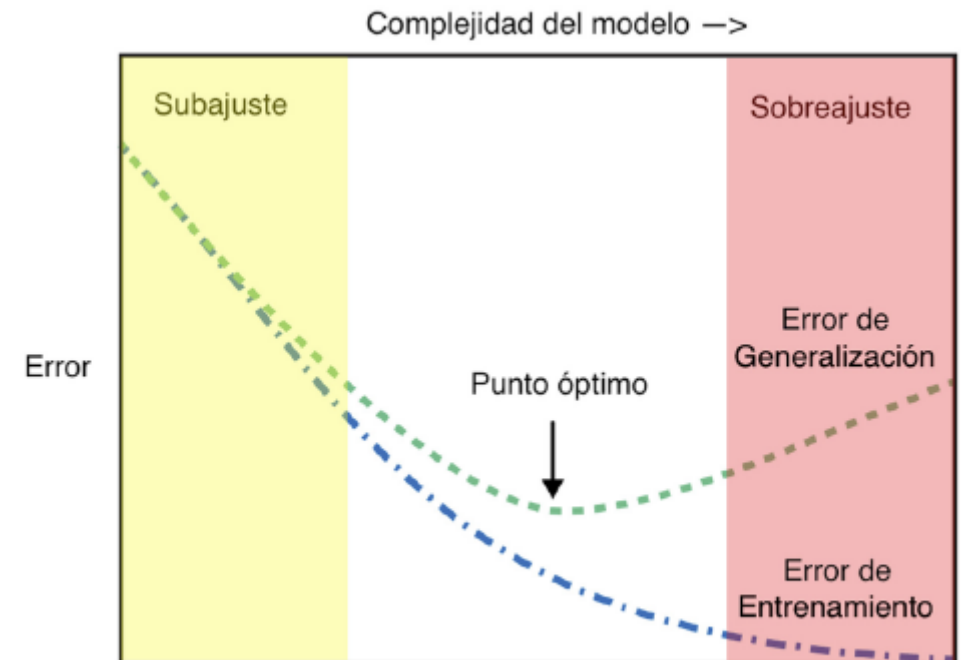


Figura 3. Subajuste y sobreajuste en aprendizaje supervisado.

¿Cómo obtenemos ambos errores? En distintos subconjuntos de datos -> **Validación**

# Índice

2.0. Introducción

**2.1. Validación hold-out**

2.2. Validación cruzada

2.3. Ajuste de parámetros y validación anidada

2.4. Evaluación en regresión

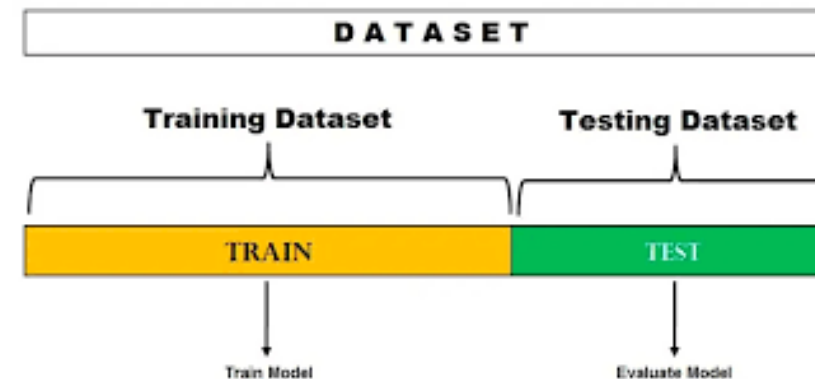
2.5. Evaluación en clasificación

## 2.1. Validación hold out

“La primera forma de validación que estudiaremos será la **validación hold-out** (mantener fuera). En ella, **el conjunto de datos del problema se divide en dos partes: entrenamiento y test**. Para definir esta división, se utiliza habitualmente un **porcentaje** que indica la proporción de muestras que se destinarán al conjunto de entrenamiento, siendo el resto de las muestras las destinadas para el conjunto de test”

Típicamente: **70% Train y 30% Test**

- Aleatorio
- Secuencial
- Dirigida
- Estratificada



**Ventajas:** sencilla de realizar, económica en términos computacionales.

**Limitación:** **escasa representatividad** de los resultados que se obtienen, sesgos.



# Índice

2.0. Introducción

2.1. Validación hold-out

**2.2. Validación cruzada**

2.3. Ajuste de parámetros y validación anidada

2.4. Evaluación en regresión

2.5. Evaluación en clasificación

## 2.2. Validación cruzada

“La validación cruzada consiste en partir, en primer lugar, el conjunto de datos original en  **$K$  subconjuntos** (también llamados *bolsas*, *folds* en inglés) **de igual tamaño**”

“Una vez partido el conjunto de datos en  $K$  bolsas, se realizan  **$K$  validaciones de tipo hold-out** empleando, en cada validación, como test una bolsa distinta y como entrenamiento el resto de ejemplos de las demás bolsas.”

Tabla 3

Validación cruzada con  $K = 5$  bolsas

Iteraciones	Bolsas de instancias					
		Bolsa 1	Bolsa 2	Bolsa 3	Bolsa 4	Bolsa 5
	1	Test	Train	Train	Train	Train
	2	Train	Test	Train	Train	Train
	3	Train	Train	Test	Train	Train
	4	Train	Train	Train	Test	Train
	5	Train	Train	Train	Train	Test

- ¿Cómo combinar los distintos folds en una métrica de bondad? -> Promedio
- Caso especial: **leave-one-out**. Sustituír promedio por concatenación.

# Índice

2.0. Introducción

2.1. Validación hold-out

2.2. Validación cruzada

**2.3. Ajuste de parámetros y validación anidada**

2.4. Evaluación en regresión

2.5. Evaluación en clasificación

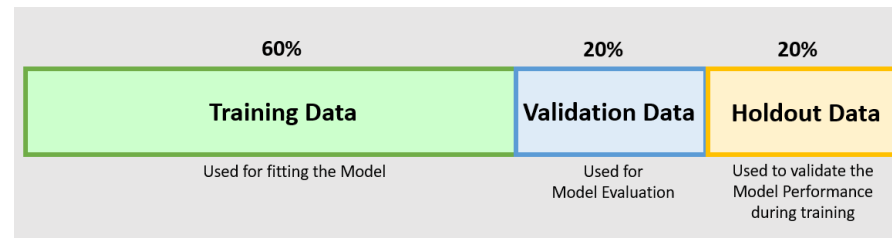
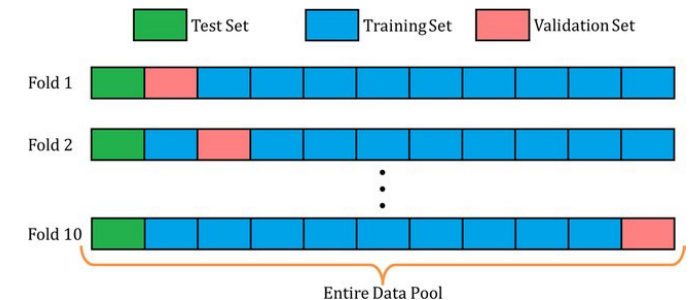
## 2.3. Ajuste de parámetros y validación anidada

“...los **algoritmos de aprendizaje supervisado** pueden poseer uno o varios **parámetros de configuración**, que **regulan su funcionamiento** y permiten adaptar el algoritmo a la naturaleza de los datos de entrenamiento. Los valores que se utilicen para los parámetros de los algoritmos afectan a la bondad de los resultados obtenidos, con lo cual estos deben ser **cuidadosamente seleccionados**. ”



Figura 4. Esquema de validación anidada con tests de validación. Adaptado de “Evaluation” por P. L. Lanzi, 2005, en *Machine Learning and Data Mining* (diapositivas en línea). Recuperado de <https://www.slideshare.net/pierluca.lanzi/machine-learning-and-data-mining-14-evaluation-and-credibility>

### • Validación anidada



# Índice

2.0. Introducción

2.1. Validación hold-out

2.2. Validación cruzada

2.3. Ajuste de parámetros y validación anidada

**2.4. Evaluación en regresión**

2.5. Evaluación en clasificación

## 2.4. Evaluación en regresión

“... MAE y RMSE, pertenecen al conjunto de **métricas de evaluación absolutas**, pues las desviaciones entre los valores reales y predichos tienen la misma unidad de magnitud que la variable de clase ... “

$$\text{MAE} = \frac{1}{n} \cdot \sum_{i=1}^n |y_i - \hat{y}_i| \qquad \text{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

“... conocer la magnitud de los errores cometidos de forma relativa al valor real que se predice... nos revelen la proporción .... **métricas de evaluación relativas**. Entre las más utilizadas, se encuentra la métrica MAPE ... “

$$\text{MAPE} = \frac{100}{n} \cdot \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

# Índice

2.0. Introducción

2.1. Validación hold-out

2.2. Validación cruzada

2.3. Ajuste de parámetros y validación anidada

2.4. Evaluación en regresión

**2.5. Evaluación en clasificación**

## 2.5. Evaluación en clasificación

“A diferencia de la regresión, una **predicción en clasificación generalmente se evalúa teniendo en cuenta si se acierta o no se acierta la clase real**, no con un grado de acierto.”

Tabla 5

Matriz de confusión con 4 clases

		Clase predicha			
		C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>
Clase real	C <sub>1</sub>	P <sub>1,1</sub>	P <sub>1,2</sub>	P <sub>1,3</sub>	P <sub>1,4</sub>
	C <sub>2</sub>	P <sub>2,1</sub>	P <sub>2,2</sub>	P <sub>2,3</sub>	P <sub>2,4</sub>
	C <sub>3</sub>	P <sub>3,1</sub>	P <sub>3,2</sub>	P <sub>3,3</sub>	P <sub>3,4</sub>
	C <sub>4</sub>	P <sub>4,1</sub>	P <sub>4,2</sub>	P <sub>4,3</sub>	P <sub>4,4</sub>

Tabla 6

Matriz de confusión binaria (2 clases)

		Clase predicha	
		Positiva	Negativa
Clase real	Positiva	TP	FN
	Negativa	FP	TN

$$\text{Exactitud}(C_i) = 100 \cdot \frac{P_{i,i}}{\sum_{j=1}^m P_{i,j}}$$

$$\text{Sensibilidad} = \frac{TP}{TP + FN}$$

$$\text{Precision} = 100 \cdot \frac{TP}{TP + FP}$$

$$F1 = 100 \cdot \frac{2 \cdot \text{Sensibilidad} \cdot \text{Precision}}{\text{Sensibilidad} + \text{Precision}}$$





viu

**Universidad**  
Internacional  
de Valencia