

# EVALUACIÓN DE MODELOS. Técnicas y métricas

Disponemos de distintas métricas para evaluar nuestros modelos de aprendizaje automático. En el caso de AA supervisado:

## 1. Modelos de Regresión

Métricas en los modelos de regresión:

- **Error absoluto medio (MAE):** Media del valor absoluto de los errores
- **Error cuadrático medio (MSE):** Media del cuadrado de los errores
- **Raíz del error cuadrático medio:** Raíz de MSE
- **R<sup>2</sup>, coeficiente de determinación:** explica si las predicciones del modelo coinciden con los valores reales.

## 2. Modelos de Clasificación

Vamos a tener en cuenta los siguientes valores:

- Verdaderos positivos (TP): Casos correctamente predichos como positivos.
- Verdaderos negativos (TN): Casos correctamente predichos como negativos.
- Falsos positivos (FP): Casos incorrectamente predichos como positivos.
- Falsos negativos (FN): Casos incorrectamente predichos como negativos.

Nuestras métricas serán:

- **Exactitud (accuracy):** indica el porcentaje de predicciones correctas realizadas por un modelo.

$$\text{Accuracy} = \frac{\text{Verdaderos positivos} + \text{Verdaderos negativos}}{\text{Total de muestras}}$$

Aunque esta medida es sencilla, no es suficiente así que tenemos otras métricas también a tener en cuenta:

- **Precisión (Precision):** Indica cuántas de las predicciones positivas fueron realmente correctas.

$$\text{Precisión} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos positivos}}$$

- **Sensibilidad (Recall):** Mide cuántos de los casos positivos reales fueron correctamente identificados por el modelo.

$$\text{Recall} = \frac{\text{Verdaderos positivos (TP)}}{\text{Verdaderos positivos (TP)} + \text{Falsos negativos (FN)}}$$

- **F1-score:** Es la media armónica entre precisión y recall y proporciona una medida más equilibrada en problemas de clases desbalanceadas.

$$\text{F1-score} = 2 \times \frac{\text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}}$$

F1-score varía entre 0 y 1, siendo 1 el mejor valor posible y 0 el peor. Es una métrica útil cuando se quiere un equilibrio entre no tener falsos positivos (precisión) y no tener falsos negativos (recall), especialmente en casos con clases desbalanceadas.

Esas 3 últimas medidas las obtenemos desde un punto de vista de TP si lo hacemos desde el TN tendríamos:

- *Especificidad:* equivalente a precisión pero con clase negativa
- *Valor predictivo negativo:* equivalente a recall, pero con la clase negativa
- *F1-score negativo:* Equivalente a F1 pero con valores negativos.

## 2.1. Visualización de resultados

### 2.1.1. Matriz de confusión

Para visualizar los resultados del algoritmo clasificador, se suele construir una *matriz de confusión*, que es una tabla que resume las predicciones del modelo frente a los resultados reales. En un clasificador binario, la matriz de confusión tiene la siguiente estructura:

Actual value	Positive	TP	FN
	Negative	FP	TN
		Positive	Negative
		Predicted value	

<https://www.ibm.com/es-es/topics/confusion-matrix#:~:text=En%20una%20matriz%20de%20confusi%C3%B3n,tambi%C3%A9n%20aparece%20en%20la%20investigaci%C3%B3n.>

Fíjate siempre donde aparecen los valores de predicción y los valores reales ya que puedes encontrar otras matrices de confusión con otro orden (por ejemplo: <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>)

Para problemas de clasificación multiclase tendríamos una matriz similar a:

Actual value	Walleye	TP			
	Largemouth Bass		TP		
	Bluegill			TP	
	Rainbow Trout				TP
		Walleye	Largemouth Bass	Bluegill	Rainbow Trout
		Predicted value			

en la que los cuadros diagonales son los que nos indican los verdaderos positivos previstos. Los otros representan el resto de valores.

### 2.1.2. Curva ROC

Otra herramienta gráfica para los modelos de clasificación es la *curva ROC* (Receiver Operating Characteristic Curve), sirve para evaluar el rendimiento especialmente cuando se trata de un problema binario. Muestra la relación entre la tasa de verdaderos positivos (TPR, True Positive Rate) y la tasa de falsos positivos (FPR, False Positive Rate)

**TPR** es realmente el Recall que comentamos antes:

$$TPR = \frac{TP}{TP + FN}$$

**FPR** representa el porcentaje de negativos reales que fueron incorrectamente clasificados como positivos

$$FPR = \frac{FP}{FP + TN}$$

Funcionamiento de la curva ROC:

- En el eje **x** de la curva ROC se representa la **FPR** (tasa de falsos positivos).
- En el eje **y** se representa el **TPR** (tasa de verdaderos positivos).
- Un modelo perfecto tendría una **FPR** de 0 y una **TPR** de 1, lo que significa que no hay falsos positivos y se identifican correctamente todos los positivos (esquinas superiores a la izquierda).
- Un modelo aleatorio generaría una línea diagonal desde la esquina inferior izquierda a la esquina superior derecha, ya que predice de manera aleatoria entre positivo y negativo.

Curva ROC ideal:

La **curva ROC** ideal se acerca a la esquina superior izquierda del gráfico, donde **TPR** es 1 (sin falsos negativos) y **FPR** es 0 (sin falsos positivos).

### ¿Qué es AUC (Area Under the Curve)?

Un **AUC de 1** indica un modelo perfecto (capaz de clasificar correctamente todas las muestras), mientras que un **AUC de 0.5** indica que el modelo no tiene capacidad de separar y está funcionando aleatoriamente. Cuanto más cerca de 1 esté el valor de AUC, mejor será el modelo..

La curva ROC nos permite comparar distintos métodos de clasificación. Un modelo de clasificación es mejor cuanto más alta es la curva. Por tanto, cuanto mayor sea el área bajo la curva, mejor será el clasificador. Y exactamente esta área se refleja en el valor AUC.

