Facultad de Ciencias Exactas, Naturales y Ambientales
Ciencia de Datos • Aprendizaje Automático Inicial

Resumen no. 5: Evaluación de modelos de aprendizaje automático

Andrés Merino • Periodo 2025-2

1. Modelos de clasificación

Consideremos un problema de clasificación binaria $f\colon X \to \{0,1\}$ y un modelo de clasificación $\hat{f}\colon X \to \{0,1\}$ que intenta aproximar a f. Además, sea un conjunto de datos $D = \{(x_i, y_i)\}_{i=1}^{n}$ el conjunto de datos utilizado para evaluar el desempeño del modelo, donde $y_i = f(x_i)$. f

Definimos los siguientes términos:

Medida	**Definición**	**Igual a**
número de positivos	$\mathrm{Pos} = \sum_{x \in D} \mathrm{I}[f(x) = 1]$	
número de negativos	$\mathrm{Neg} = \sum_{x \in D} \mathrm{I}[f(x) = 0]$	
verdaderos positivos	$\mathrm{TP} = \sum_{x \in D} \mathrm{I}[\hat{f}(x) = f(x) = 1]$	
verdaderos negativos	$\mathrm{TN} = \sum_{x \in D} \mathrm{I}[\hat{f}(x) = f(x) = 0]$	
falsos positivos	$\mathrm{FP} = \sum_{x \in D} \mathrm{I}[\hat{f}(x) = 1, f(x) = 0]$	
falsos negativos	$\mathrm{FN} = \sum_{x \in D} \mathrm{I}[\hat{f}(x) = 0, f(x) = 1]$	
exactitud (accuracy)	$\mathrm{acc} = \frac{1}{\lvert Te \rvert} \sum_{x \in D} \mathrm{I}[\hat{f}(x) = f(x)]$	$\frac{\mathrm{TP} + \mathrm{TN}}{\lvert D \rvert}$
error	$\mathrm{err} = \frac{1}{\lvert Te \rvert} \sum_{x \in D} \mathrm{I}[\hat{f}(x) \neq f(x)]$	$\frac{\mathrm{FP} + \mathrm{FN}}{\lvert D \rvert}$
precisión, confianza	$\mathrm{pre} = \frac{\sum_{x \in D} \mathrm{I}[\hat{f}(x) = f(x) = 1]}{\sum_{x \in D} \mathrm{I}[\hat{f}(x) = 1]}$	$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$
tasa de verdaderos positivos, **sensibilidad** (recall)	$\mathrm{tpr} = \mathrm{rec} = \frac{\sum_{x \in D} \mathrm{I}[\hat{f}(x) = f(x) = 1]}{\sum_{x \in D} \mathrm{I}[f(x) = 1]}$	$\frac{\mathrm{TP}}{\mathrm{FN} + \mathrm{TP}}$
tasa de verdaderos negativos, **especificidad**	$\mathrm{tnr} = \frac{\sum_{x \in D} \mathrm{I}[\hat{f}(x) = f(x) = 0]}{\sum_{x \in D} \mathrm{I}[f(x) = 0]}$	
tasa de falsos positivos	$\mathrm{fpr} = \frac{\sum_{x \in D} \mathrm{I}[\hat{f}(x) = 1, f(x) = 0]}{\sum_{x \in D} \mathrm{I}[f(x) = 0]}$	
tasa de falsos negativos	$\mathrm{fnr} = \frac{\sum_{x \in D} \mathrm{I}[\hat{f}(x) = 0, f(x) = 1]}{\sum_{x \in D} \mathrm{I}[f(x) = 1]}$	
Puntuación F1	$\mathrm{F1} = 2 \cdot \frac{\mathrm{pre} \cdot \mathrm{rec}}{\mathrm{pre} + \mathrm{rec}}$	

Con esto, la matriz de confusión queda definida como:

	Predicción: 1	Predicción: 0
Real: 1	TP	FN
Real: 0	FP	TN

Precisión: Proporción de instancias clasificadas como positivas que son realmente positivas. Se prioriza cuando el costo de un Falso Positivo es alto.

Un ejemplo claro es un filtro de correo spam en un entorno corporativo. Aquí, clasificar un correo importante (legítimo) como spam (Falso Positivo) es grave, pues podría perderse información crítica. Por ello, se prefiere que el modelo sea muy conservador: solo debe marcar como spam si está totalmente seguro, incluso si eso significa dejar pasar algunos correos basura a la bandeja de entrada (baja sensibilidad).

	Predicción: Spam (1)	Predicción: No Spam (0)
Real: Spam (1)	30 (TP)	50 (FN)
Real: No Spam (0)	1 (FP)	919 (TN)

En este caso, minimizamos los Falsos Positivos para maximizar la precisión:

$$\text{pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{30}{30 + 1} = \frac{30}{31} \approx 0{,}968$$

Esto indica un 96.8 % de confianza. Aunque se filtraron pocos correos de spam del total real (muchos FN), garantizamos que lo que se capturó era efectivamente basura.

Sensibilidad (recall): Proporción de instancias positivas que fueron correctamente identificadas. Se prioriza cuando el costo de un Falso Negativo es alto.

El mejor ejemplo es el diagnóstico médico de una enfermedad grave y contagiosa. El objetivo es detectar a todos los individuos enfermos. Es preferible alarmar a un paciente sano para realizarle más pruebas (Falso Positivo) que enviar a casa a un paciente enfermo diciéndole que está sano (Falso Negativo), ya que esto tendría consecuencias fatales.

	Predicción: Enfermo (1)	Predicción: Sano (0)
Real: Enfermo (1)	98 (TP)	2 (FN)
Real: Sano (0)	60 (FP)	840 (TN)

Aquí buscamos minimizar los Falsos Negativos a toda costa:

$$\text{rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{98}{98 + 2} = \frac{98}{100} = 0{,}98$$

El modelo tiene una sensibilidad del 98 %. Detecta casi todos los casos reales, sacrificando la precisión (muchos falsos positivos) en favor de la seguridad.

1.1 Curva ROC

Dado un modelo de clasificación binaria que produce probabilidades, es decir,

$$\hat{f}\colon X \to [0, 1],$$

dado un umbral de decisión $\theta \in [0, 1]$, definimos la función de clasificación binaria asociada como

$$\hat{f}_\theta(x) = \begin{cases} 1 & \text{si } \hat{f}(x) \geqslant \theta, \\ 0 & \text{si } \hat{f}(x) < \theta. \end{cases}$$

Para cada valor de θ, podemos calcular las métricas del modelo, por ejemplo, la tasa de verdaderos positivos (tpr_θ) y la tasa de falsos positivos (fpr_θ).

DEFINICIÓN 1: Curva ROC.
La **curva ROC** (Receiver Operating Characteristic) es la representación gráfica de los pares $(\mathrm{fpr}_\theta, \mathrm{tpr}_\theta)$ al variar el umbral θ en el intervalo $[0, 1]$.

DEFINICIÓN 2: Área bajo la curva ROC (AUC).
El **AUC** (Area Under the Curve) es el área bajo la curva ROC. Un AUC de 1 indica un modelo perfecto, mientras que un AUC de 0.5 indica un modelo sin capacidad discriminativa (equivalente a una clasificación aleatoria).

Podemos establecer los siguientes rangos para interpretar el AUC:

- $[0{,}97, 1{,}00)$: Excelente
- $[0{,}90, 0{,}97)$: Muy bueno
- $[0{,}75, 0{,}90)$: Bueno
- $[0{,}60, 0{,}75)$: Regular
- $[0{,}50, 0{,}60)$: Malo

2. MODELOS DE REGRESIÓN

Consideremos un problema de regresión $f\colon X \to \mathbb{R}$ y un modelo de regresión $\hat{f}\colon X \to \mathbb{R}$ que intenta aproximar a f. Además, sea un conjunto de datos $D = \{(x_i, y_i)\}_{i=1}^n$ el conjunto de datos utilizado para evaluar el desempeño del modelo, donde $y_i = f(x_i)$.

Definimos las siguientes métricas de evaluación:

Medida	**Definición**
Error absoluto medio (MAE)	$\mathrm{MAE} = \frac{1}{\lvert D\rvert}\sum_{i=1}^{n} \lvert\hat{f}(x_i) - f(x_i)\rvert$
Error cuadrático medio (MSE)	$\mathrm{MSE} = \frac{1}{\lvert D\rvert}\sum_{i=1}^{n} (\hat{f}(x_i) - f(x_i))^2$
Raíz del error cuadrático medio (RMSE)	$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{\lvert D\rvert}\sum_{i=1}^{n} (\hat{f}(x_i) - f(x_i))^2}$
Coeficiente de determinación (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{f}(x_i) - f(x_i))^2}{\sum_{i=1}^{n} (f(x_i) - \bar{y})^2}$, donde $\bar{y} = \frac{1}{\lvert D\rvert}\sum_{i=1}^{n} f(x_i)$

3. Modelos de agrupamiento

Dado un agrupamiento de datos $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$, donde cada C_i es un conjunto de datos que representa un clúster, tenemos las siguientes definiciones

- **Centroide del clúster:** Es el punto medio del clúster, calculado como el promedio de todos los puntos en el clúster

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x.$$

- **Diámetro del clúster:** Mide la distancia máxima entre dos puntos dentro del mismo clúster

$$diam(C_i) = \max_{x,y \in C_i} d(x, y).$$

- **Inercia:** Mide la cohesión interna de los clústeres, definida como la suma de las distancias cuadráticas entre cada punto y el centroide de su clúster

$$I_i = \sum_{x \in C_i} d(x, \mu_i)^2.$$

- **Distancia promedio dentro del clúster:** Mide la distancia promedio entre los puntos y el centroide del clúster

$$S_i = \frac{1}{|C_i|} \sum_{x \in C_i} d(x, \mu_i).$$

- **Cohesión de Silueta:** Mide la distancia promedio entre un punto y todos los demás puntos en el mismo clúster

$$a(x_i) = \frac{1}{|C_i| - 1} \sum_{\substack{x_j \in C_k \\ j \neq i}} d(x_i, x_j).$$

- **Distancia entre clústeres:** Mide la distancia mínima entre puntos de dos clústeres diferentes

$$d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y).$$

- **Distancia entre centroides:** Mide la distancia entre los centroides de dos clústeres

$$d_c(C_i, C_j) = d(\mu_i, \mu_j).$$

- **Separación de Silueta:** Mide la distancia promedio entre un punto y todos los puntos en el clúster más cercano al que no pertenece

$$b(x_i) = \min_{j \neq i} \frac{1}{|C_j|} \sum_{x_k \in C_j} d(x_i, x_k).$$

Con esto, definimos las siguientes métricas de evaluación:

Medida	**Definición**
Inercia total	$I_{total} = \sum_{i=1}^{K} I_i = \sum_{i=1}^{K} \sum_{x \in C_i} d(x, \mu_i)^2$
Índice de Davies-Bouldin	$DB = \frac{1}{K} \sum_{i=1}^{K} \max_{j \neq i} \left(\frac{S_i + S_j}{d_c(C_i, C_j)} \right)$
Índice de Dunn	$Dunn = \frac{\min_{i \neq j} d(C_i, C_j)}{\max_{1 \leqslant k \leqslant K} diam(C_k)}$
Índice de Silueta	$S = \frac{1}{\lvert D \rvert} \sum_{i=1}^{\lvert D \rvert} \frac{b(x_i) - a(x_i)}{\max\{a(x_i), b(x_i)\}}$