

Flujo de trabajo y métricas de evaluación para clasificación

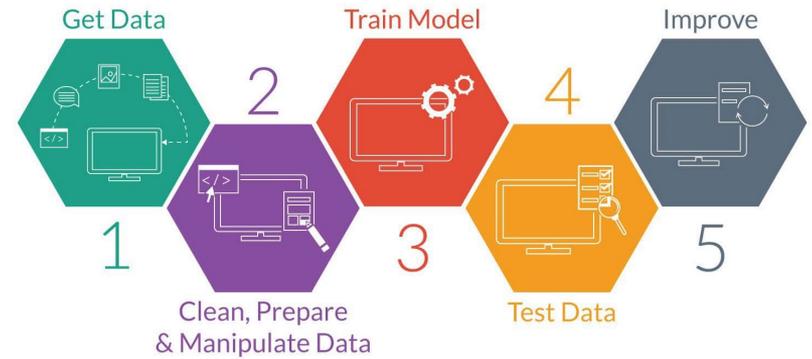
Germán Rosati

IDAES/UNSAM - CONICET - PIMSA

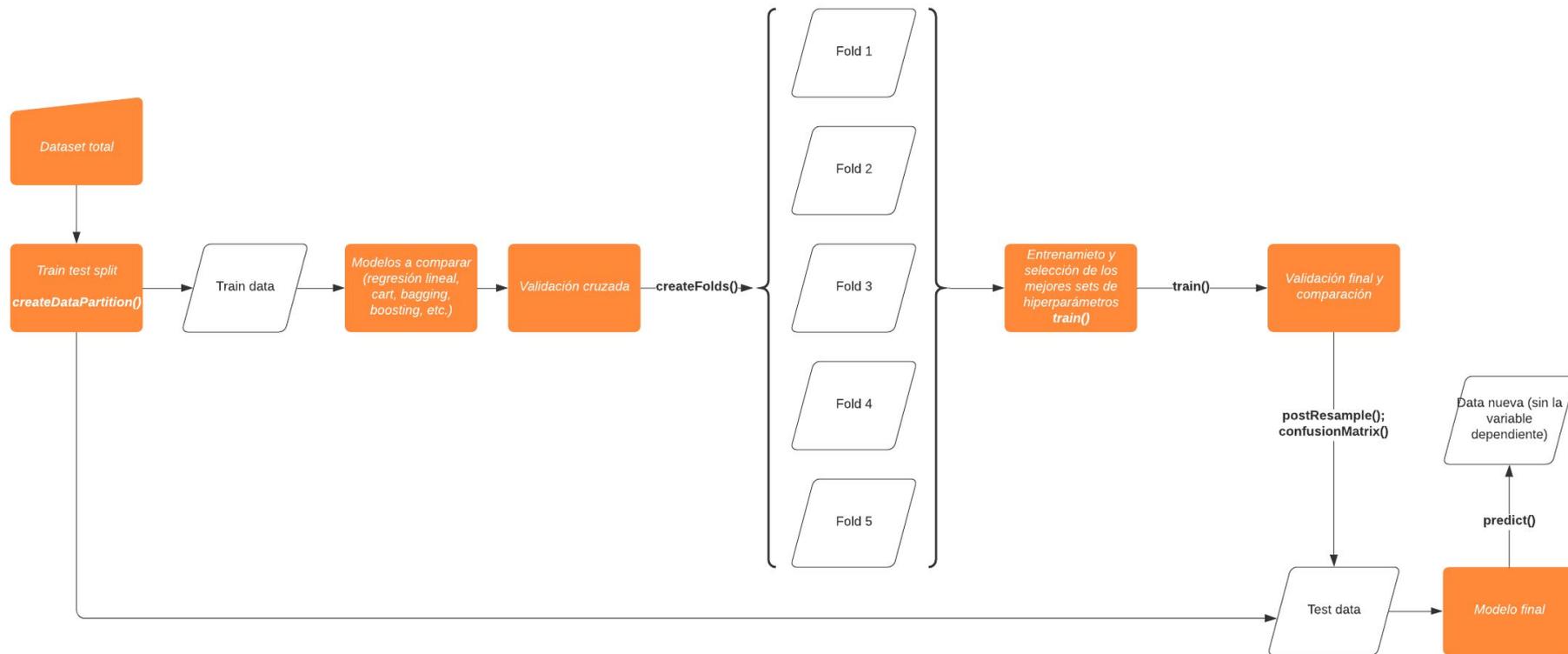
12 de Noviembre de 2020

Introducción

- Momento de pasar en limpio
- ¿Cuáles son las etapas básicas del flujo de trabajo de un proceso de entrenamiento de un modelo?
- El flujo que vamos a ver es uno de los múltiples posibles



Esquema general



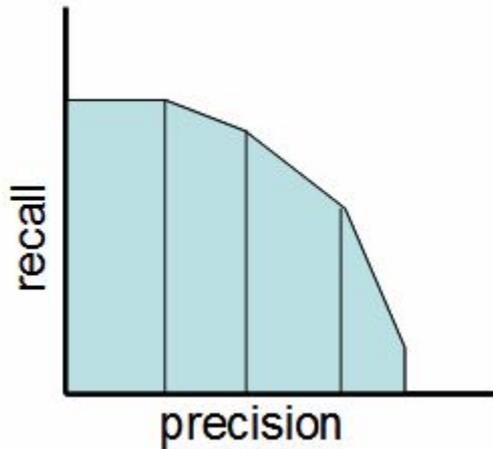
Evaluando modelos de clasificación

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

También llamada **recall**

Evaluando modelos de clasificación

$$F_1 = 2 * \frac{\textit{precision} * \textit{recall}}{\textit{precision} + \textit{recall}}$$



- Es la media armónica entre precision y recall
- Balancea ambas métricas
- Si precision es baja => F1 es baja
- Si recall es baja => F1 es baja

Evaluando modelos de clasificación

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

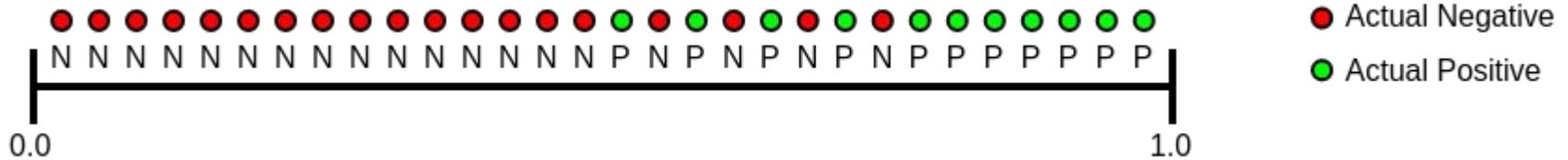
- F1 da igual peso a precision y recall
- F_β permite balancear ese peso, mediante el parámetro beta
- Beta mide cuánta importancia le damos a precision en relación al recall

Evaluando modelos de clasificación

- Ahora... ¿cómo decide un método si la predicción es positiva o negativa?
- Regresión logística (pero vale para otros)
 - La predicción “nativa” de una regresión logística es una probabilidad
 - Se aplica una regla de decisión para asignarle un valor positivo o negativo
 - Si $p(y | X) > 0.5 \Rightarrow$ el caso es positivo
 - Si $p(y | X) \leq 0.5 \Rightarrow$ el caso es negativo

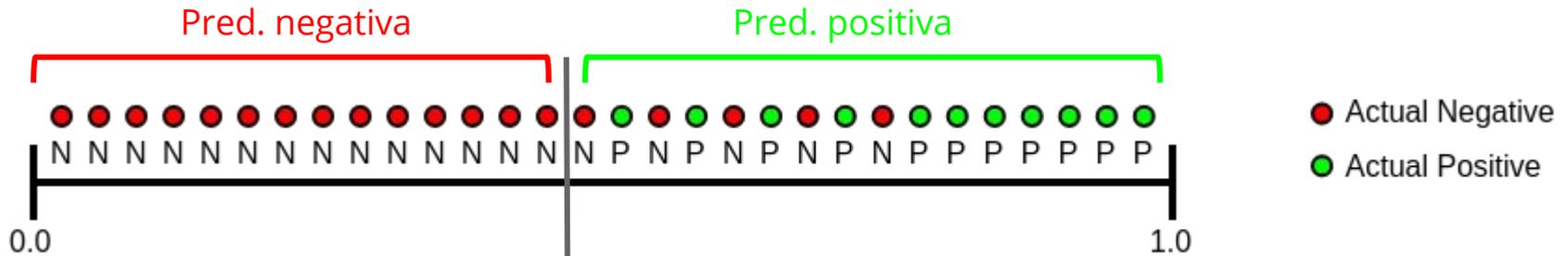
Evaluando modelos de clasificación

- Variando ese umbral (0.5), puedo hacer hacer variar las métricas basadas en la matriz de confusión



Evaluando modelos de clasificación

- Variando ese umbral (0.5), puedo hacer hacer variar las métricas basadas en la matriz de confusión

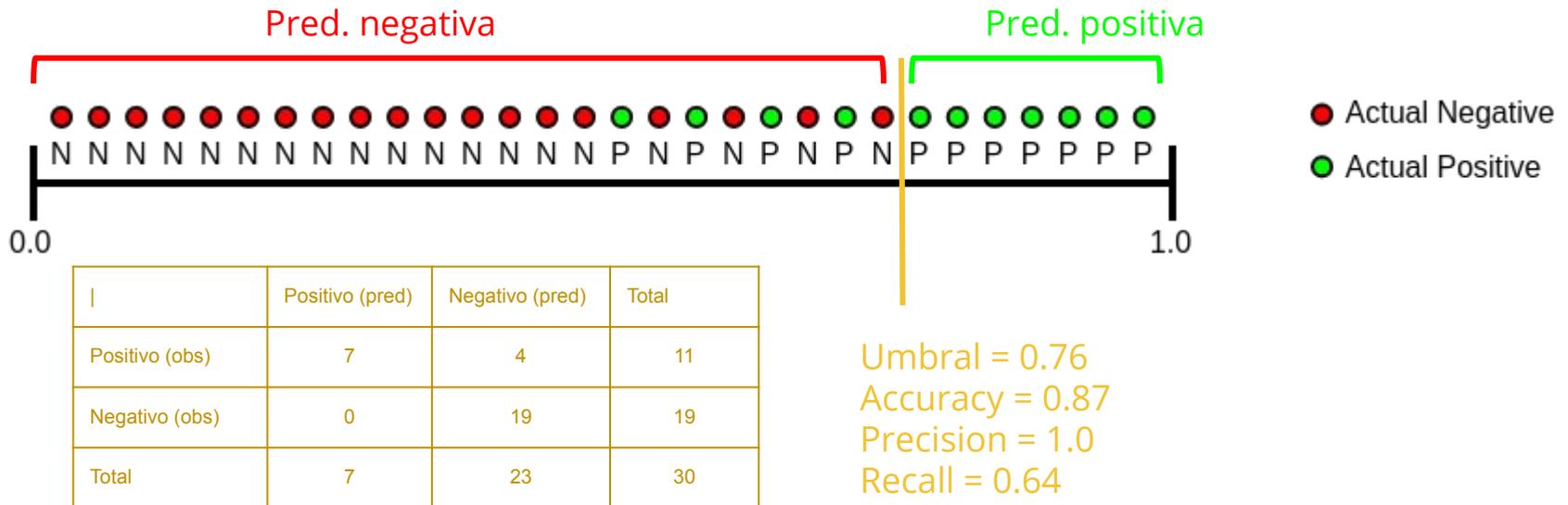


Umbral = 0.5
Accuracy = 0.83
Precision = 0.69
Recall = 1.0

	Positivo (pred)	Negativo (pred)	Total
Positivo (obs)	11	0	11
Negativo (obs)	5	14	19
Total	16	14	30

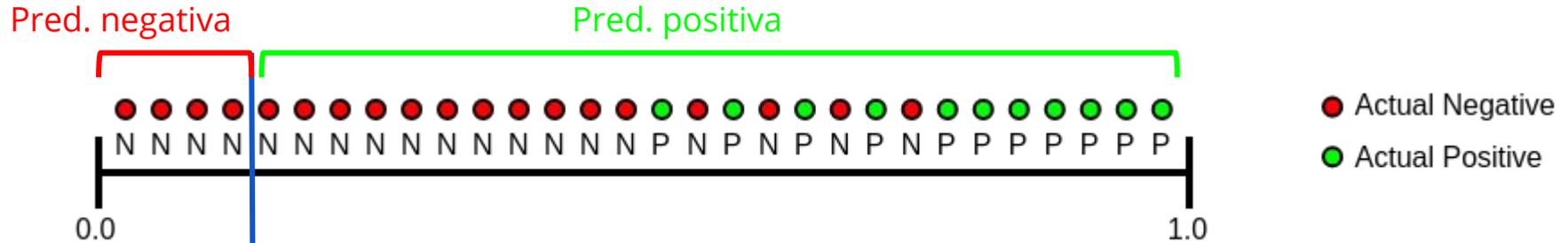
Evaluando modelos de clasificación

- Variando ese umbral (0.5), puedo hacer hacer variar las métricas basadas en la matriz de confusión



Evaluando modelos de clasificación

- Variando ese umbral (0.5), puedo hacer hacer variar las métricas basadas en la matriz de confusión

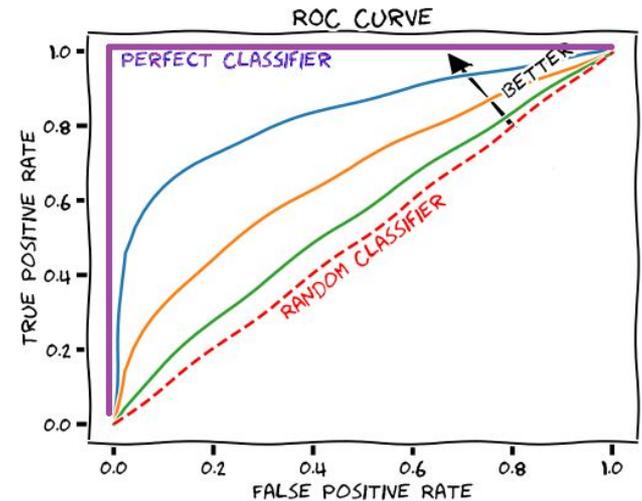


Umbral = 0.13
Accuracy = ?
Precision = ?
Recall = ?

	Positivo (pred)	Negativo (pred)	Total
Positivo (obs)	11	0	11
Negativo (obs)	15	4	19
Total	26	4	30

Evaluando modelos de clasificación

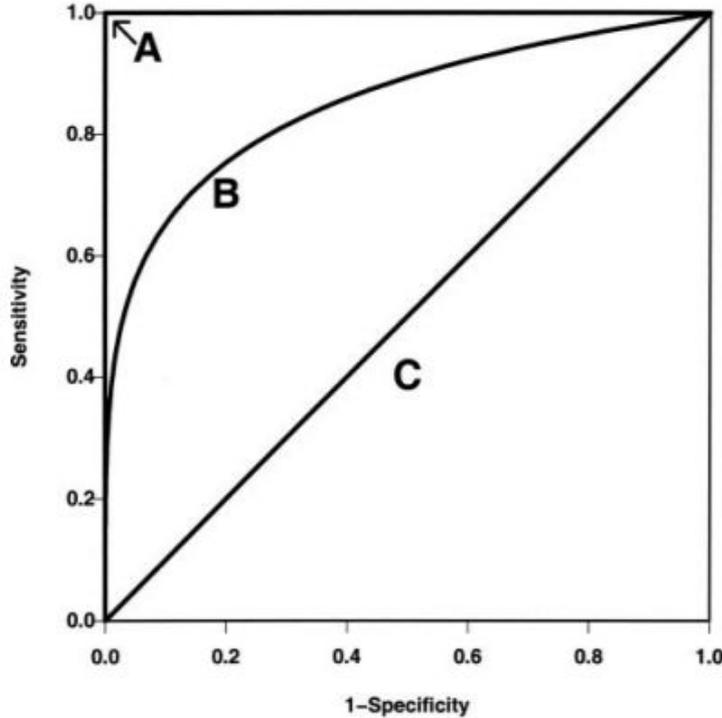
- Curva ROC y métrica AUC: cuantificar la performance agregada de diferentes clasificadores para todos los umbrales posibles
- Eje X: tasa de falsos positivos (1 - especificidad)
- Eje Y: tasa de verdaderos positivos (sensibilidad)
- Querriámos clasificadores que scoreen alto en Y y bajo en X



Evaluando modelos de clasificación

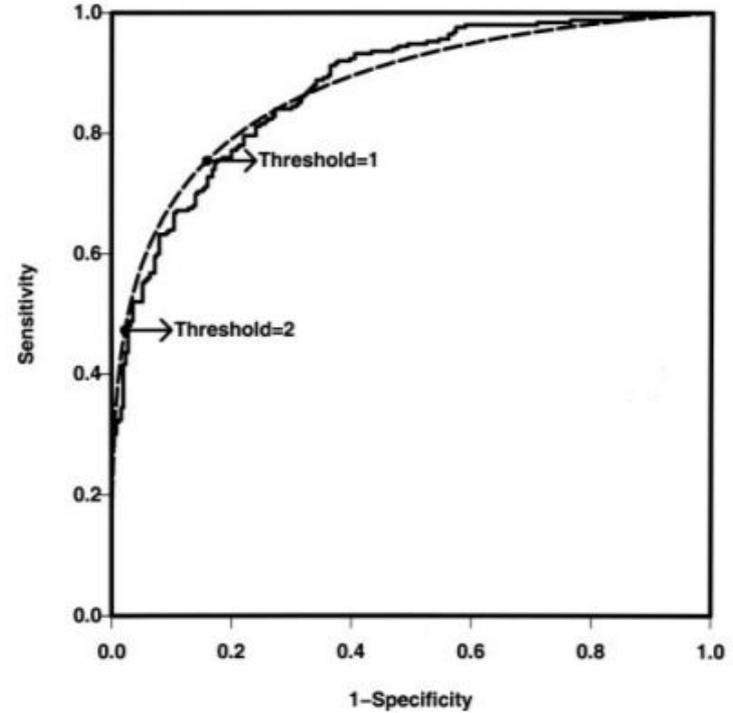
True positive rate (o recall)

$$\frac{tp}{tp + fn}$$

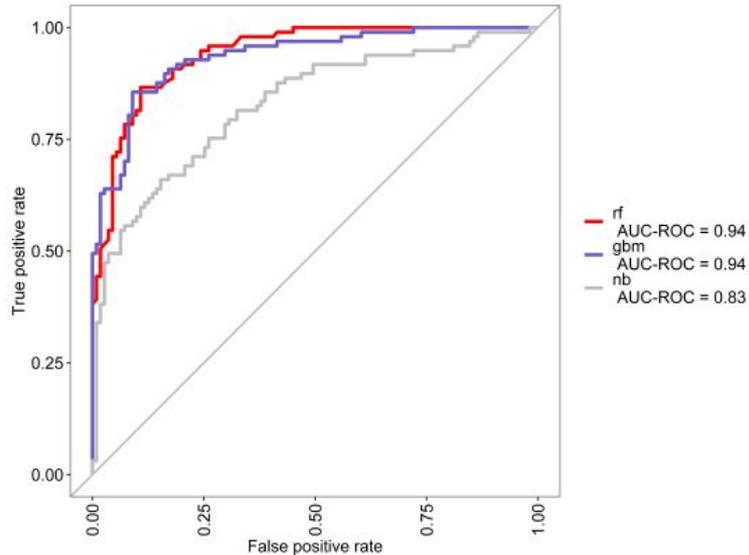


False positive rate

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



Evaluando modelos de clasificación



- En este ejemplo tenemos ploteados las curvas ROC de tres modelos de clasificación
- ¿Cuál es mejor? ¿Por qué?