

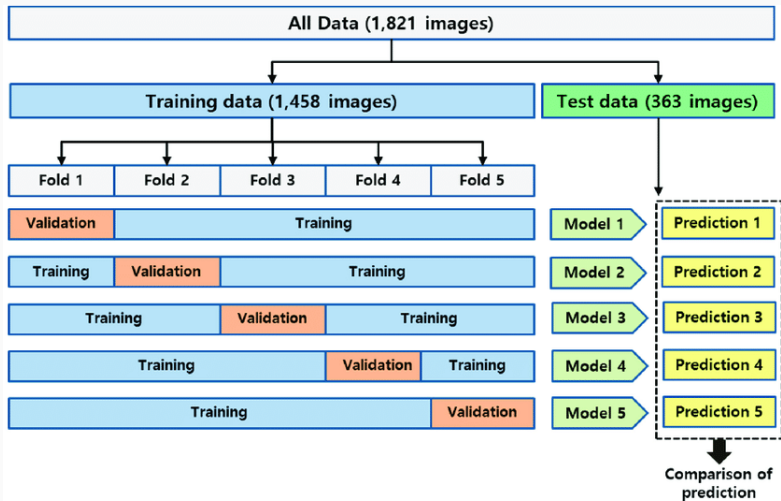
Métricas de evaluación

Blanca Vázquez

22 de noviembre de 2023

Validación cruzada (cross validation)

Es un método estadístico utilizado para evaluar el rendimiento de un modelo; garantizando la independencia entre los datos de entrenamiento y de prueba.



¿Por qué es importante la validación cruzada?

- El uso de múltiples divisiones de entrenamiento-prueba proporciona estimaciones más estables y confiables.
- Cada división se usa para entrenar y evaluar un modelo separado.
- La exactitud de un método de aprendizaje supervisado puede variar, dependiendo de qué muestras terminan en el conjunto de entrenamiento.
- Los resultados se promedian en múltiples conjuntos de entrenamiento diferentes en lugar de depender de un único modelo entrenado en un conjunto de entrenamiento particular.

Matriz de confusión

Una matriz de confusión es una representación matricial que describe el rendimiento de un modelo de clasificación binaria.

		Clase predicha	
		Positivo	Negativo
Clase real	Positivo	Verdadero positivo	Falso negativo
	Negativo	Falso positivo	Verdadero negativo

Exactitud ()

La exactitud es la fracción de todas las instancias en la que la predicción del clasificador es correcta.

		Clase predicha	
		Positivo	Negativo
Clase real	Positivo	Verdadero positivo	Falso negativo
	Negativo	Falso positivo	Verdadero negativo

$$accuracy = \frac{VP + VN}{VP + VN + FP + FN}$$

Exhaustividad (recall o true positive rate)

La exhaustividad es la fracción de las instancias positivas que el clasificador identifica correctamente como positivas.

$$recall = \frac{VP}{VP + FN}$$

		Clase predicha	
		Positivo	Negativo
Clase real	Positivo	Verdadero positivo	Falso negativo
	Negativo	Falso positivo	Verdadero negativo

Esta métrica nos permite intuir que tan bueno es el clasificador para encontrar todas las clases positivas.

Precisión

La precisión es la fracción de las predicciones positivas que son correctas.

		Clase predicha	
		Positivo	Negativo
Clase real	Positivo	Verdadero positivo	Falso negativo
	Negativo	Falso positivo	Verdadero negativo

$$precision = \frac{VP}{VP + FP}$$

Mide los patrones que son correctamente predichos a partir del total de patrones predichos en una clase positiva.

Especificidad

La especificidad es la fracción de las instancias negativas que el clasificador identifica correctamente como negativas.

		Clase predicha	
		Positivo	Negativo
Clase real	Positivo	Verdadero positivo	Falso negativo
	Negativo	Falso positivo	Verdadero negativo

$$especificidad = \frac{VN}{VN + FP}$$

F1 (equilibrio entre precisión y recall)

La métrica de F1 se interpreta como la media armónica entre la precisión y el *recall*.

Escenario 1

$$precision = \frac{VP}{VP + FP} = 1.0$$

$$recall = \frac{VP}{VP + FN} = 0.47$$

Alta precisión y bajo *recall*

Escenario 2

$$precision = \frac{VP}{VP + FP} = 0.36$$

$$recall = \frac{VP}{VP + FN} = 1.0$$

Baja precisión y alto *recall*

F1 combina la precisión y el *recall* en un solo valor.

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

Curva ROC (Receiver Operating Characteristic)

- La curva ROC es una herramienta estadística que se utiliza para evaluar la capacidad discriminativa de un clasificador binario.
- Es una representación gráfica que permite visualizar el equilibrio entre la tasa de verdaderos positivos (eje Y) y la tasa falsos positivos (eje X).
- Esto significa que la esquina superior izquierda de la gráfica es el punto ideal: una tasa de falsos positivos = 0 y una tasa de verdaderos positivos = 1.

Curva ROC (Receiver Operating Characteristic)

- Tasa de verdaderos positivos (TPR)

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

- Tasa de falsos positivos (FPR)

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

Área bajo la curva ROC

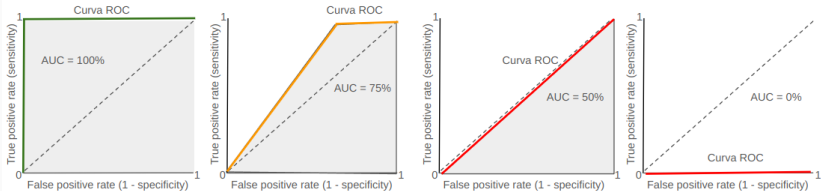
El AUC (por las siglas en inglés de *area under curve ROC*) mide toda el área bidimensional por debajo de la curva ROC completa. Esta métrica refleja el rendimiento global de un clasificador.

$$AUC = \frac{S_p - n_p(n_n + 1)/2}{n_p n_n}$$

donde S_p es la suma de todas las instancias positivas, mientras n_p y n_n indican el número de las muestras positivas y negativas respectivamente.

- Algunos teóricos indican que el AUC es la métrica ideal para medir el rendimiento y discriminación de un modelo de clasificación, en comparación con la exactitud.
- Sin embargo, esta métrica es altamente costosa en problemas multiclase.

Análisis de la curva ROC y el AUC



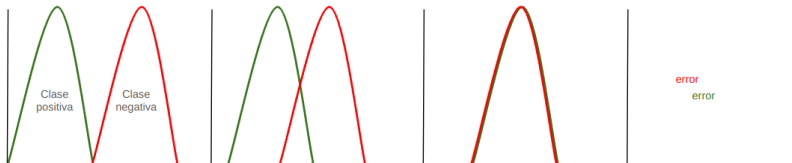
Excelente

Bueno

No separables

!!

Traslape: ¿qué tan bueno es el modelo para separar las clases positivas y las negativas?



Raíz del error cuadrático medio (RMSE)

El RMSE (por las siglas en inglés de *Root Mean Square Error*) mide la diferencia entre los valores predichos por un modelo y los valores reales.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N}}$$

- Rango: $[0, \text{infinito})$
- Entre más pequeño el valor, mejor

Error absoluto medio (MAE)

MAE (por las siglas en inglés de *Mean Absolute Error*) es el promedio de la diferencia absoluta entre el valor real y el predicho.

$$\text{MAE} = \frac{\sum_{i=0}^{N-1} |y_i - \hat{y}_i|}{N}$$

- Rango: $[0, \text{infinito})$
- Entre más pequeño el valor, mejor

Coeficiente de similitud de Dice

Es una de las métricas más comunes para comparar la similitud de dos muestras. La medida cuantifica la coincidencia de dos conjuntos A y B normalizando el tamaño de su intersección sobre el promedio de sus tamaños:

$$DICE = \frac{|A \cap B|}{\frac{1}{2}(|A| + |B|)} = \frac{2|A \cap B|}{|A| + |B|}$$

- Frecuentemente se utiliza para evaluar segmentación de imágenes.
- Devuelve un escalar numérico con un intervalo $[0, 1]$.
- Un valor igual a 1 significa que las segmentaciones en las dos imágenes coinciden perfectamente.

Coeficiente de similitud de Jaccard

Esta métrica mide el grado de similitud entre dos conjuntos, sea cual sea el tipo de elementos.

Para calcular este coeficiente se utiliza la cardinalidad de la intersección de ambos conjuntos dividida por la cardinalidad de su unión.

$$Jaccard = \frac{|A \cap B|}{|A \cup B|}$$

- Devuelve un escalar numérico con un intervalo $[0, 1]$.