

Interpretación de la Matriz de Confusión de sklearn

UNMSM – Inteligencia Artificial - Juan Gamarra Moreno

Reporte de clasificación de sklearn

- El reporte de clasificación se utiliza para medir la calidad de las predicciones de un algoritmo de clasificación.
- Cuántas predicciones son verdaderas y cuántas falsas.
- Más específicamente, los verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN) se utilizan para predecir las métricas de un informe de clasificación.

Reporte de clasificación de sklearn

El código usado para generar el reporte es:

```
from sklearn.metrics import classification_report  
y_true = [0, 1, 2, 2, 2]  
y_pred = [0, 0, 2, 2, 1]  
target_names = ['class 0', 'class 1', 'class 2']  
print(classification_report(y_true, y_pred, target_names=target_names))
```

Reporte de clasificación de sklearn

Salida:

	precision	recall	f1-score	support
class 0	0.50	1.00	0.67	1
class 1	0.00	0.00	0.00	1
class 2	1.00	0.67	0.80	3
avg / total	0.70	0.60	0.61	5

Precisión

La precisión es la capacidad de un clasificador de no etiquetar una instancia como positiva que en realidad es negativa. Para cada clase, se define como la relación entre verdaderos positivos y la suma de verdaderos y falsos positivos.

- TP: verdaderos positivos
- FP - Falsos positivos

Precisión: precisión de las predicciones positivas.

- $\text{Precisión} = \text{TP} / (\text{TP} + \text{FP})$

Recall (Recuerdo): ¿Qué porcentaje de casos positivos captó?

Recall es la capacidad de un clasificador de encontrar todas las instancias positivas. Para cada clase, se define como la proporción de verdaderos positivos con la suma de verdaderos positivos y falsos negativos.

- FN - Falsos negativos

Recall: Fracción de positivos que se identificaron correctamente.

- $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$

Puntuación F1: ¿Qué porcentaje de predicciones positivas fueron correctas?

La puntuación F1 es una media armónica ponderada de precisión y recuerdo de manera que la mejor puntuación es 1,0 y la peor es 0,0. En términos generales, los puntajes F1 son más bajos que las medidas de precisión, ya que incorporan precisión y recuerdo en su cálculo. Como regla general, el promedio ponderado de F1 debe usarse para comparar modelos de clasificación, no la precisión global.

- $\text{F1 score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$

¿Cuándo es más
importante la precisión que
el recall?

UNMSM – UPG-FISI-Machine Learning y Big Data

Nemónico para entenderlo

"PREcision is to PREGnancy tests as reCALL is to CALL center"

- Con una prueba de embarazo, el fabricante de la prueba debe asegurarse de que un resultado positivo signifique que la mujer está realmente embarazada. Las personas podrían reaccionar a una prueba positiva casándose repentinamente o comprando una casa (si muchos consumidores obtuvieran falsos positivos y sufrieran enormes costos sin motivo, el fabricante de la prueba se quedaría sin clientes).

Nemónico para entenderlo (Continua)

- Ahora imagine un centro de llamadas para reclamos de seguros. La mayoría de los reclamos fraudulentos se realizan por teléfono los lunes, después de que los estafadores se conecten con los colaboradores y elaboren sus historias inventadas ("digamos que robaron el automóvil") durante el fin de semana. ¿Qué es lo mejor que puede hacer una compañía de seguros los lunes? Tal vez deberían favorecer el recall sobre la precisión. Es mucho mejor marcar más reclamos como positivos (probable fraude) para una mayor investigación que pasar por alto algunos de los fraudes y pagar efectivo que nunca debería haberse pagado. Un falso positivo (marcado para un escrutinio adicional como posible fraude, pero la pérdida del cliente fue real) probablemente se puede aclarar asignando un ajustador experimentado, que puede insistir en un informe policial, solicitar un video de seguridad del edificio, etc. Un falso negativo (aceptar el reclamo falso de un defraudador y el pago en efectivo) es una pérdida pura para la compañía de seguros y fomenta más fraudes.

"Precision" más importante que "Recall"

- Imagine que queremos asegurarnos de que nuestro bloqueador de sitios web para nuestro hijo solo permita que se muestren sitios web "seguros". En este caso, un sitio web "seguro" es la clase positiva. Aquí, queremos que el bloqueador esté absolutamente seguro de que el sitio web es seguro, incluso si se prevé que algunos sitios web seguros sean parte de la clase negativa o insegura y, en consecuencia, se bloquen. Es decir, queremos una alta precisión a expensas del recall.

"Precision" más importante que "Recall"

		Predicción	
		Seg	No Seg
Verdad	Seg	180	120
	No Seg	10	290
		190	410

$$\text{Prec(Seg)} = \frac{180}{190} = 0.95$$

$$\text{Recall(Seg)} = \frac{180}{300} = 0.60$$

"Precision" más importante que "Recall"

		Predicción	
		Seg	No Seg
Verdad	Seg	180	120
	No Seg	0	300
		180	420

$$\text{Prec(Seg)} = \frac{180}{180} = 1.00$$

$$\text{Recall(Seg)} = \frac{180}{300} = 0.60$$

"Recall" más importante que "Precision"

- En el caso de la seguridad aeroportuaria, donde un riesgo de seguridad es la clase positiva, queremos asegurarnos de que se investigue cada riesgo potencial de seguridad. En este caso, tendremos un alto recall a expensas de la precisión (se investigarán muchas bolsas en las que no hay riesgos de seguridad).

"Recall" más importante que "Precision"

		Predicciones		
		Riesgo	No Riesgo	
Real	Riesgo	290	10	300
	No Riesgo	200	100	300
		490	110	600

$\text{prec}(\text{riesgo}) = \frac{290}{490} = 0.59$

$\text{recall}(\text{riesgo}) = \frac{290}{300} = 0.96$

"Recall" más importante que "Precision"

		Predicciones		
		Riesgo	No Riesgo	
Real	Riesgo	300	0	300
	No Riesgo	200	100	300
		500	100	600

$$\text{prec(riesgo)} = \frac{300}{500} = 0.60$$

$$\text{recall(riesgo)} = \frac{300}{300} = 1.00$$

Ejemplo del Centro de Llamadas

- Se tienen miles de clientes gratuitos registrándose en nuestro sitio web cada semana. El equipo del centro de llamadas quiere llamarlos a todos, pero es imposible, por lo que me piden que seleccione a los potenciales compradores. No nos importa llamar a un tipo que no va a comprar (así que la precisión no es importante) pero para nosotros es muy importante que los potenciales compradores estén siempre en mi selección, para que no se vayan sin comprar. Eso significa que mi modelo debe tener un alto recuerdo, sin importar demasiado la precisión.

`y_verdadero = [0, 1, 2, 2, 2]`

`y_predicho = [0, 0, 2, 2, 1]`

$c\emptyset$ = clase \emptyset
 $c1$ = clase 1
 $c2$ = clase 2

		Predicciones		
		$c\emptyset$	$c1$	$c2$
Verdaderos	$c\emptyset$	[1, 0, 0]	= 1	
	$c1$	[1, 0, 0]	= 1	
	$c2$	[0, 1, 2]	= 3	
		2	1	2
				$\sqrt{5}$

Predichos				precision	recall	f1-score	support
Verd.	C0	C1	C2				
C0	[1, 0, 0]	1		0.50	1.00	0.67	1
C1	[1, 0, 0]	1		0.00	0.00	0.00	1
C2	[0, 1, 2]	3		1.00	0.67	0.80	3
	2 1 2	15					
				accuracy	0.50	0.60	0.60
				macro avg	0.70	0.56	0.49
				weighted avg		0.60	0.61

precision(C0):

$$\frac{1}{2} = 0.5$$

precision(C1):

$$\frac{0}{1} = 0$$

precision(C2):

$$\frac{2}{2} = 1.0$$

precision promedio:

$$(0.5 + 0 + 1.0) / 3 = 0.5$$

precision ponderada

$$(0.5 \times 1 + 0 \times 1 + 1 \times 3) / 5 = 0.70$$

recall ponderado

$$(1 \times 1 + 0 \times 1 + 0.67 \times 3) / 5 = 0.60$$

clase 0
clase 1
clase 2

accuracy
macro avg
weighted avg

$$f_1(C0) = 2(1 \times 0.5) / (1+0.5) = 0.67$$

$$f_1(C1) = 2(0 \times 0) / (0+0) \approx 0.00$$

$$f_1(C2) = 2(2/3 \times 1) / (2/3 + 1) = 0.80$$

$$f_1 \text{ promedio} = (0.67 + 0 + 0.80) / 3 = 0.49$$

$$f_1 \text{ ponderado} = (0.67 \times 1 + 0 \times 1 + 0.80 \times 3) / 5 = 0.61$$

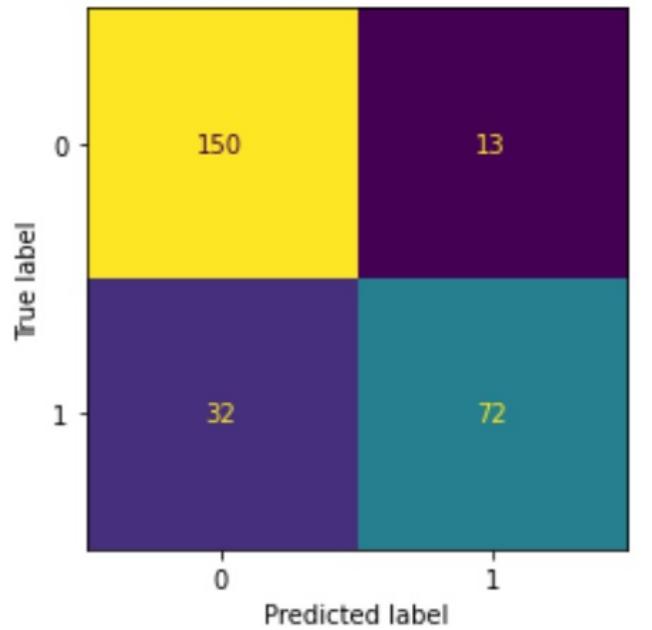
Accuracy (Exactitud)

$$acc = \frac{1 + 0 + 2}{5} = \frac{3}{5} = 0.60$$

Predicciones

	0	1	
Verdad	0 [150, 13]	163	
	1	32, 72	104
	182	85	<u>267</u>

	precision	recall	f1-score	support
0	0.82	0.92	0.87	163
1	0.85	0.69	0.76	104
accuracy			0.83	267
macro avg	0.84	0.81	0.82	267
weighted avg	0.83	0.83	0.83	267



0 : No sobrevive

1 : Sobre vive

$$\text{precisión (no sobrevive)} = 150/182$$

$$\text{precisión (sobrevive)} = 72/85$$

$$\text{recall (no sobrevive)} = 150/163$$

$$\text{recall (sobrevive)} = 72/104$$

$$f1(\text{no sobrevive}) = 2(0.82 \times 0.92) / (0.82 + 0.92)$$

$$f1(\text{sobrevive}) = 2(0.85 \times 0.69) / (0.85 + 0.69)$$

$$\text{Accuracy} = \frac{150 + 72}{267}$$