

Cómo calcular las métricas a partir de la matriz de confusión cuando hay más de dos clases

Tenemos las clases a, b y c, correspondientes a un pronóstico: bueno, regular y malo. Tenemos la siguiente matriz de confusión.

=== Confusion Matrix ===

a b c <-- classified as

49 1 0 | a = bueno

0 47 3 | b = regular

0 2 48 | c = malo

Vamos a calcular las métricas siguientes para las tres clases:

- Tasa de verdaderos positivos
- Tasa de falsos positivos
- Precisión

Clase a=bueno

Primero nos centramos en la clase a=bueno. Para calcular sus métricas consideramos que la clase a es la clase positiva y las instancias del resto de clases son miembros de la clase negativa. La clase positiva viene marcada en la matriz con color rojo y la negativa en azul:

a b c <-- classified as

49 1 0 | a = bueno

0 47 3 | b = regular

0 2 48 | c = malo

A efectos de los cálculos para la clase A, esta matriz es equivalente:

a b+c <-- classified as

49 1 | a = bueno

0 100 | b+c = regular+malo

100 es el resultado de sumar lo que no está ni en la fila de la clase a ni en la columna de la clase a.

Como la fila indica la clase real de las instancias, tenemos 49+1 instancias de la clase positiva (a=bueno) y 47+3+2+48 instancias de la clase negativa (la unión de las clases b y c).

Calculamos las métricas para la clase a.

Tasa de verdaderos positivos (TPR o True Positive Rate):

$$TPR \text{ Clase A} = \frac{TP}{TP + FN} = \frac{49}{49 + 1} = 0.98$$

En este caso nos centramos en las cifras que aparecen en la fila correspondiente a la clase a (instancias que realmente son de la clase a). Hay 1 instancia de la clase A catalogada como de la clase B, por lo tanto sólo hay un falso negativo. El resto de instancias, 49, se han catalogado correctamente como de la clase A.

Tasa de falsos positivos (FPR o False Positive Rate):

$$FPR \text{ Clase A} = \frac{FP}{FP + TN} = \frac{0}{0 + 47 + 3 + 2 + 48} = 0$$

Los falsos positivos son aquellos que son de la clase b o c y se han catalogado en la clase a. Serían aquellos que están en la columna "classified as a", exceptuando los 49 de la diagonal que son los verdaderos positivos. Son los rodeados en la matriz de confusión a continuación:

a b c <-- classified as

49 1 0 | a = bueno

0 47 3 | b = regular

0 2 48 | c = malo

Los TN o verdaderos negativos, cuando estamos centrados en la clase a, son todas aquellas instancias que no son de la clase a y no se han catalogado como a, así que son el total de instancias de la clase b y c que no se han catalogado como de la clase a. Lo rodeado en la matriz de confusión a continuación:

a b c <-- classified as

49 1 0 | a = bueno

0 47 3 | b = regular

0 2 48 | c = malo

El denominador del FPR no es más que la suma de todas las instancias negativas (las de la clase b y c).

Respecto a la precisión, nos centramos en la columna que corresponde a las instancias catalogadas como a:

a b c <-- classified as

49 1 0 | a = bueno

0 47 3 | b = regular

0 2 48 | c = malo

$$\text{Precisión Clase A} = \frac{TP}{TP + FP} = \frac{49}{49 + 0 + 0} = 1$$

Clase B=regular

Calculamos ahora las métricas para la clase B. Entonces ahora la clase positiva es la b y las negativas son las clases a y c, esto es, todo lo que no es clase b. Marco en la matriz de confusión a la clase positiva en rojo y a las negativas en azul:

a b c <-- classified as

49 1 0 | a = bueno

0 47 3 | b = regular

0 2 48 | c = malo

Para calcular la TPR nos tenemos que centrar en las cifras que aparecen en la fila de la clase b.

Hay 47 instancias correctamente clasificadas en la clase b luego serán los TP y hay 3 instancias catalogadas en otra clase, luego son falsos negativos (FN)

$$TPR \text{ Clase B} = \frac{TP}{TP + FN} = \frac{47}{47 + 3} = 0.94$$

Para calcular la FPR consideramos negativos a todas las instancias que no son de la clase B, luego serían las marcadas en azul en la matriz anterior. Los falsos positivos son 3, que son aquellas instancias que no siendo de la clase b (ya sean de la clase a o c) han sido catalogadas como b, una instancia de la clase a ha sido catalogada como b y dos instancias de la clase c se han catalogado como b, luego su suma son los falsos positivos de la clase b.

a b c <-- classified as

49 1 0 | a = bueno

0 47 3 | b = regular

0 2 48 | c = malo

$$FPR \text{ Clase B} = \frac{FP}{FP + TN} = \frac{1 + 2}{1 + 2 + 49 + 0 + 0 + 48} = 0.03$$

Para calcular la precisión nos centramos en la columna "classified as b". Tenemos 1+47+2 instancias clasificadas como b, siendo 1 de ellas de la clase a (luego es un FP) y 2 de ellas de la clase c(luego es un FP):

$$\text{Precisión Clase B} = \frac{TP}{TP + FP} = \frac{47}{47 + 1 + 2} = 0.94$$

Clase c=malo

A continuación calculamos las métricas para la clase C, y por tanto se considera positiva a la tercera fila y negativas a las otras dos filas.

a b c <-- classified as

49 1 0 | a = bueno

0 47 3 | b = regular

0 2 48 | c = malo

$$\text{TPR Clase C} = \frac{TP}{TP + FN} = \frac{48}{48 + 2} = 0.96$$

$$\text{FPR Clase C} = \frac{FP}{FP + TN} = \frac{3}{3 + 49 + 1 + 47} = 0.03$$

$$\text{Precisión Clase C} = \frac{TP}{TP + FP} = \frac{48}{48 + 3} = 0.9411$$

Estos son los cálculos que hace Weka para mostrar los siguientes datos en su salida:

	TP Rate	FP Rate	Precision	Recall	Class
	0,980	0,000	1,000	0,980	bueno
	0,940	0,030	0,940	0,940	regular
	0,960	0,030	0,941	0,960	malo
Weighted Avg.	0,960	0,020	0,960	0,960	