

Métricas

DRA. CONSUELO VARINIA GARCIA MENDOZA



Métricas

Existen varias métricas para evaluar el desempeño de los clasificadores

- Matriz de confusión
- TP,TN,FP & FN
- Exactitud (Accuracy)
- Precisión (Precision)
- Exhaustividad(Recall)
- Valor F (F1 Score)
- Averages of F1 Score

Ejemplo 1. Clasificador Binario

- 30 pacientes que se realizaron la prueba de Covid
- 13 resultados positivos
 - 7 tenían Covid
 - 6 no tenían Covid
- 17 resultados negativos
 - 2 tenían Covid
 - 15 no tenían Covid

Matriz de Confusión

Clases reales	Clases predichas	
	Positivo	Negativo
	Positivo	Negativo
	7 (TP)	2 (FN)
	6 (FP)	15 (TN)

Accuracy

Porcentaje de predicción correcta, es decir el porcentaje de casos que el modelo ha acertado lo que se traduce en capacidad de generalización. Es la métrica más utilizada

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{22}{30} = 0.73 = 73\%$$

Clases reales	Clases predichas	
	Positivo	Negativo
	Negativo	
	Positivo	7 (TP) 2 (FN)
	Negativo	6 (FP) 15 (TN)

Recall

Porcentaje de instancias por clase que se recuperaron correctamente

$$Recall_P = \frac{TP}{TP+FN} = \frac{7}{7+2} = \frac{7}{9} = 0.77 = 77\%$$

$$Recall_N = \frac{TN}{TN+FP} = \frac{15}{15+6} = \frac{15}{21} = 0.71 = 71\%$$

Clases reales	Clases predichas	
	Positivo	Negativo
	Positivo	Negativo
	7 (TP)	2 (FN)
	6 (FP)	15 (TN)

Precision

Porcentaje de predicción correcta por clase

$$Precision_P = \frac{TP}{TP+FP} = \frac{7}{7+6} = \frac{7}{13} = 0.53 = 53\%$$

$$Precision_N = \frac{TN}{TN+FN} = \frac{15}{2+15} = \frac{15}{17} = 0.88 = 88\%$$

Clases reales	Clases predichas	
	Positivo	Negativo
	Positivo	7 (TP) 2 (FN)
Negativo	6 (FP)	15 (TN)

F1

Combina precision y recall mediante la media armónica para hacer un resumen del rendimiento del modelo

$$F1_1 = 2 * \frac{Precision_P * Recall_P}{Precision_P + Recall_P} = 2 * \frac{0.53 * 0.77}{0.53 + 0.77} = 0.63$$

$$F1_2 = 2 * \frac{Precision_N * Recall_N}{Precision_N + Recall_N} = 2 * \frac{0.88 * 0.71}{0.88 + 0.71} = 0.78$$

$$Precision_P = 0.53 \quad Precision_N = 0.88$$

$$Recall_P = 0.77 \quad Recall_N = 0.71$$

Resumen

Clase	Acuracy	Recall	Precision	F1 score
Positivo	0.73	0.77	0.53	0.63
Negativo		0.71	0.88	0.78

Ejemplo 2. Clasificador Multi-clase

Supongamos que hemos entrenado algún clasificador multiclase con tres clases y en la etapa de pruebas se obtuvo el siguiente resultado:

No. Instancia	1	2	3	4	5	6	7	8	9	10	11	12
Clases reales	1	1	2	1	2	3	2	3	1	3	3	3
Clases predichas	1	1	2	2	3	3	2	3	1	2	3	3

Clases reales	Clases predichas			
		1	2	3
	1	3	1	0
	2	0	2	1
	3	0	1	4

Accuracy

Porcentaje de predicción correcta

$$Accuracy = \frac{d}{i} = \frac{9}{12} = 0.75 = 75\%$$

donde

d : instancias predichas correctamente (instancias en la diagonal de la matriz de confusión)

i : total de instancias

Clases reales	Clases predichas			
		1	2	3
	1	3	1	0
	2	0	2	1
	3	0	1	4

Recall

Porcentaje de instancias por clase que se recuperaron correctamente

$$Recall_j = \frac{c_j}{rc_j}$$

Donde

c_j : instancias de la clase j predichas correctamente

rc_j : instancias que pertenecen a la clase j

Para el ejemplo

$$Recall_1 = \frac{c_1}{rc_1} = \frac{3}{4} = .75 = 75\%$$

$$Recall_2 = \frac{c_2}{rc_2} = \frac{2}{3} = 0.67 = 67\%$$

$$Recall_3 = \frac{c_3}{rc_3} = \frac{4}{5} = 0.8 = 80\%$$

Clases reales	Clases predichas			
		1	2	3
	1	3	1	0
	2	0	2	1
	3	0	1	4

Precision

Porcentaje de predicción correcta por clase

$$Precision_j = \frac{c_j}{pc_j}$$

donde

c_j : instancias de la clase j predichas correctamente

pc_j : instancias predichas como clase j

Para el ejemplo

$$Precision_1 = \frac{c_1}{pc_1} = \frac{3}{3} = 1 = 100\%$$

$$Precision_2 = \frac{c_2}{pc_2} = \frac{2}{4} = 0.5 = 50\%$$

$$Precision_3 = \frac{c_3}{pc_3} = \frac{4}{5} = 0.8 = 80\%$$

Clases reales	Clases predichas			
		1	2	3
	1	3	1	0
	2	0	2	1
	3	0	1	4

F1

Combina precision y recall
mediante la media armónica

$$F1_j = 2 * \frac{Precision_j * Recall_j}{Precision_j + Recall_j}$$

Para el ejemplo

$$Precision_1 = 1 \quad Precision_2 = 0.5 \quad Precision_3 = 0.8$$

$$Recall_1 = 0.75 \quad Recall_2 = 0.67 \quad Recall_3 = 0.8$$

$$F1_1 = 2 * \frac{Precision_1 * Recall_1}{Precision_1 + Recall_1} = 2 * \frac{1 * 0.75}{1 + 0.75} = 0.86$$

$$F1_2 = 2 * \frac{Precision_2 * Recall_2}{Precision_2 + Recall_2} = 2 * \frac{0.5 * 0.67}{0.5 + 0.67} = 0.57$$

$$F1_3 = 2 * \frac{Precision_3 * Recall_3}{Precision_3 + Recall_3} = 2 * \frac{0.8 * 0.8}{0.8 + 0.8} = 0.8$$

Averages of F1 Score

Como observamos en la tabla tenemos un valor de F1 score por clase, podemos promediarlos para obtener un único valor que describa el rendimiento global del clasificador.

Existen tres métodos para obtener este valor global

- Macro Average
- Weighted Average
- Micro Average

Clase	Acuracy	Recall	Precision	F1 score
1	0.75	0.75	1	0.86
2		0.67	0.5	0.57
3		0.8	0.8	0.8

Macro Average

La puntuación macro average F1 score o macro F1 score se calcula utilizando la media aritmética (es decir, la media no ponderada) de todas las puntuaciones F1 por clase.

Clase	Acuracy	Recall	Precision	F1 score	Macro-average F1 score
1	0.75	0.75	1	0.86	$\frac{0.86 + 0.57 + 0.8}{3} = 0.74$
2		0.67	0.5	0.57	
3		0.8	0.8	0.8	

Weighted Average

Weighted Average F1 se calcula tomando la media de todas las puntuaciones F1 por clase teniendo en cuenta el soporte de cada clase.

El soporte se refiere al número de ocurrencias reales de la clase en el conjunto de datos. El peso se refiere esencialmente a la proporción del soporte de cada clase en relación con la suma de todos los valores de soporte.

Clases reales	Clases predichas			
		1	2	3
	1	3	1	0
	2	0	2	1
	3	0	1	4

Clase	Acuracy	Recall	Precision	F1 score	Support	Support Proportion	Weighted average F1 score
1	0.75	0.75	1	0.86	4	$\frac{4}{12}=0.33$	$(0.86*0.33)+(0.57*0.25)+(0.8*0.41)=0.7543$
2		0.67	0.5	0.57	3	$\frac{3}{12}=0.25$	
3		0.8	0.8	0.8	5	$\frac{5}{12}=0.41$	
					12	1	

¿Cómo calculamos TP, FP y FN en clasificación single label multiclase?

- Multiclase

- OvR
- OvO

Clases reales	Clases predichas			
		1	2	3
	1	3	1	0
	2	0	2	1
	3	0	1	4

Clase	True Positive (TP)	False Positive (FP) 	False Negative (FN) —
1	3	0	1
2	2	2	1
3	4	1	1
Total			

Micro Average

Estos resultados significan que en los casos de clasificación multiclase en los que cada observación tiene una única etiqueta (single label)

micro-F1 = micro-precisión = micro-recall = acuracy

Clase	True Positive (TP)	False Positive (FP)	False Negative (FN)	Micro average values
1	3	0	1	$\text{micro-recall} = \frac{TP}{TP + FN} = \frac{9}{9 + 3} = 0.75$
2	2	2	1	
3	4	1	1	
Total	9	3	3	$\text{micro-precision} = \frac{TP}{TP + FP} = \frac{9}{9 + 3} = 0.75$
				$\text{micro-F1} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} = \frac{9}{9 + \frac{1}{2}(3 + 3)} = 0.75$