

Aprendizaje Supervisado

Métricas de clasificación

October 23, 2021

1 Slides

- Definiciones y ejemplos

2 Ejercicios

Evaluar un clasificador es en general más difícil que evaluar un regresor debido a que disponemos de muchas métricas de evaluación de modelos. Diferentes tipos de problemas que requieren métricas específicas.

- Matriz de confusión
- Accuracy
- Precision - Recall - F1
- ROC - AUC

Definición

Table	Predicted Negative	Predicted Positive	Subtotals
Actual Negative	TN	FP	TN+FP
Actual Positive	FN	TP	FN+TP
Subtotals	TN+FN	FP+TP	TOTAL

Table: Matriz de confusión

- Error tipo 1: FP
- Error tipo 2: FN
- Accuracy = $\frac{TN+TP}{TOTAL}$

Ejemplo

Table	Predicted Negative	Predicted Positive	Subtotals
Actual Negative	723	38	761
Actual Positive	89	64	153
Subtotals	812	102	914

Table: Ejemplo 1

- Error tipo 1: $FP = 38$
- Error tipo 2: $FN = 89$
- Accuracy = $\frac{723+64}{914} = 86.11\%$

Ejemplo

Si siempre predecimos la clase negativa:

Table	Predicted Negative	Predicted Positive	Subtotals
Actual Negative	761	0	761
Actual Positive	153	0	153
Subtotals	914	0	914

Table: Ejemplo Dummy

- Error tipo 1: $FP = 0$
- Error tipo 2: $FN = 153$
- Accuracy = $\frac{761+0}{914} = 83.20\%$

Precision

- **Precision:** $\frac{TP}{FP+TP}$
- Predicted Positive = $FP + TP$

Recall

- **Recall:** $\frac{TP}{FN+TP}$
- Actual Positive = $FN + TP$

F1 score

- **F1:** media armónica de precision y recall
- $F1 = \frac{2Precision*Recall}{Precision+Recall}$

Ejemplo

Table	Predicted Negative	Predicted Positive	Subtotals
Actual Negative	723	38	761
Actual Postive	89	64	153
Subtotals	812	102	914

Table: Ejemplo 1

- Precision = $64 / 102 = 62.75\%$
- Recall = $64/153 = 41.83\%$
- F1 = 50.19%

Caso 1

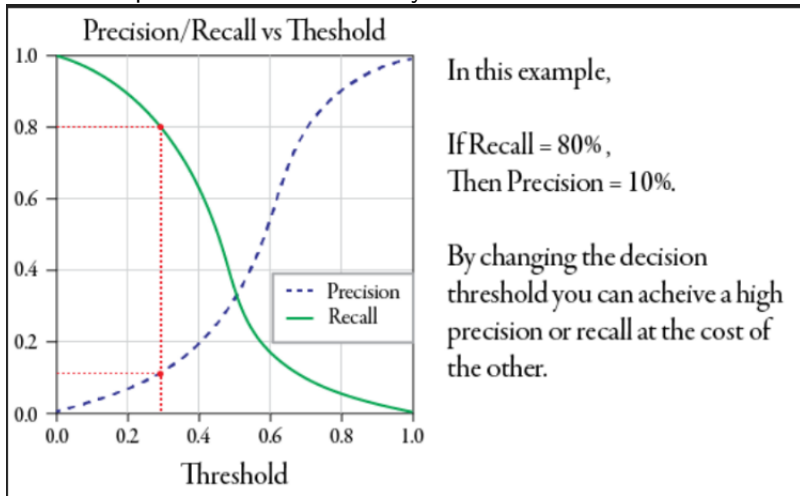
En detección de fraude de movimientos de tarjetas de crédito, no nos interesa bajo precision, sino que alto recall (poder detectar los TP entre los realmente fraudulentos que son los Actual Positive para este caso). De otra forma: si predecimos fraude cuando no lo hay, no sería lo más grave.

Caso 2

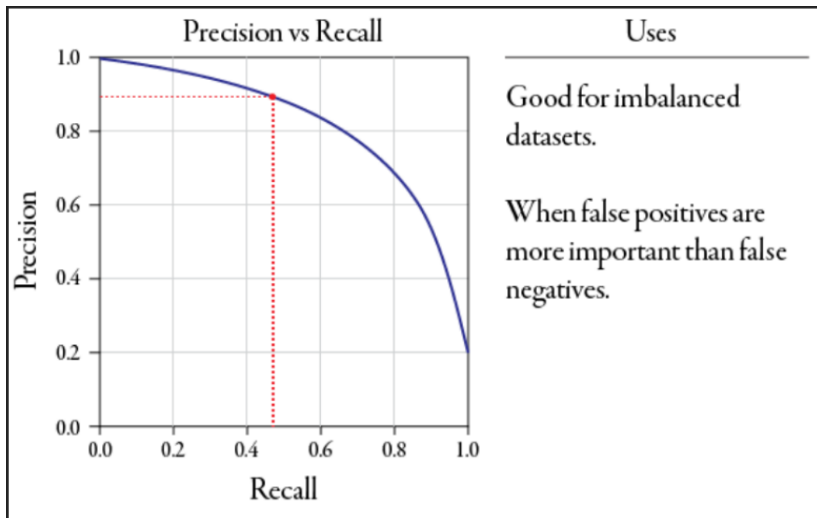
En detección veneno de hongos comestibles, nos interesa alto precision, ya que queremos que efectivamente sean comestibles los que estamos prediciendo que lo son. El peor escenario sería predecir comestible hongos venenosos.

Trade off

Aumentar precision reduce recall y viceversa!



Trade off

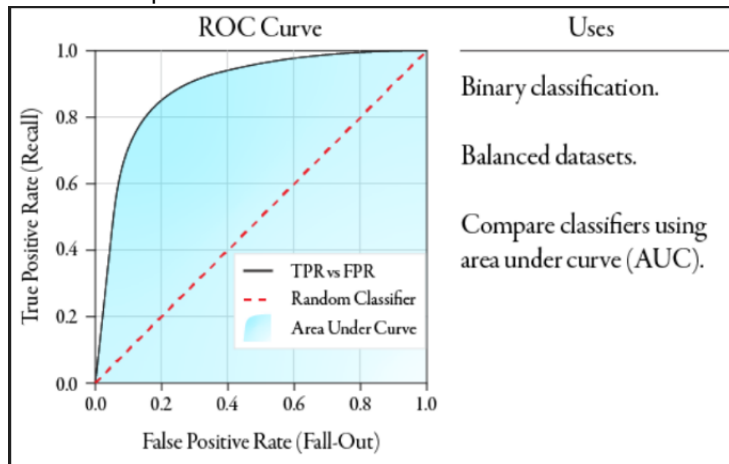


Curva ROC

$$\text{FallOut} = \frac{FP}{FN+TP}$$

FN+TP= Actual Positive

ROC curve plots Recall contra FallOut:



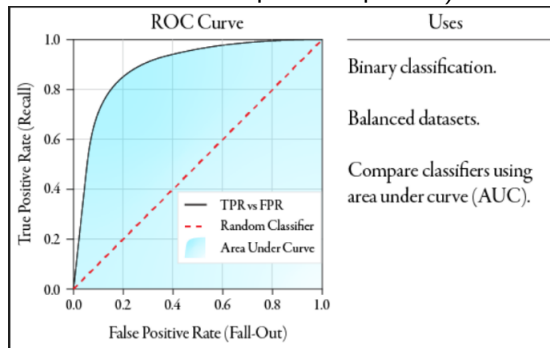
Curva ROC

AUC:= area under ROC curve

Perfect classifier: $AUC = 1$

Random classifier: $AUC = 0.5$

Buscamos clasificador con curva ROC mas alejada de la diagonal (más cercano al vértice superior izquierdo).



Ejercicios

- 1 Libro ILRS2 4.7.1 y 4.7.2 (página 171)
- 2 Escribir funciones para calcular métricas y aplicarlas
- 3 Dada la siguiente tabla, usando las funciones anteriores calcular precision, recall, F1-score y graficar la curva ROC. ¿Qué nivel de threshold da el mejor modelo?

Threshold	TP	FP	TN	FN
0.0	50	50	0	0
0.1	48	47	3	2
0.2	47	40	9	4
0.3	45	31	16	8
0.4	44	23	22	11
0.5	42	16	29	13
0.6	36	12	34	18
0.7	30	11	38	21
0.8	20	4	43	33
0.9	12	3	45	40
1.0	0	0	50	50