

Evaluación de clasificadores

Christopher Expósito Izquierdo

Airam Expósito Márquez

Israel López Plata

Belén Melián Batista

J. Marcos Moreno Vega

{**cexposit, aexposim, ilopezpl, mbmelian, jmmoreno**}@ull.edu.es

Departamento de Ingeniería Informática y de Sistemas
Universidad de La Laguna



1. INTRODUCCIÓN

2. MATRIZ DE CONFUSIÓN

3. ESTADÍSTICOS

INTRODUCCIÓN

Evaluación (i)

- Antes de su uso, todo modelo obtenido desde los datos debe ser **evaluado para medir su calidad**.
- Es una tarea no trivial que puede depender de varios criterios, algunos de ellos bastante subjetivos y que pueden estar enfrentados entre sí.
- Debe recordarse que los patrones descubiertos deben tener tres cualidades: debe ser válidos (precisos), interesantes (útiles y novedosos) y comprensibles.

Evaluación (ii)

- **Etapas de entrenamiento y validación.** Para medir la calidad de los modelos predictivos deben definirse apropiadamente las etapas de entrenamiento y validación. De esta manera se asegura que las predicciones sean precisas y robustas.
 - **Datos de entrenamiento** (training dataset): conjunto de datos que se emplea para **construir el modelo**.
 - **Datos de validación** (test dataset): conjunto de datos que se usa para **validar el modelo** obtenido en la etapa anterior.

Evaluación (iii)

- **Técnicas de evaluación (i)**
 - **Validación simple:** se divide aleatoriamente el conjunto de datos en dos subconjuntos: uno para el entrenamiento y otro para la validación. Con el primero se construye el modelo que luego es validado con los datos del segundo. El conjunto de validación suele constar del 5% al 50% de los datos iniciales.

Evaluación (iv)

- **Técnicas de evaluación (ii)**
 - **Validación cruzada:** se divide aleatoriamente el conjunto de datos en dos subconjuntos.
 - Se construye el modelo con el primer subconjunto y se valida con el segundo. Se obtiene así un primer ratio de error (o precisión).
 - Se construye el modelo con el segundo subconjunto y se valida con el primero, obteniéndose un segundo ratio de error.
 - Finalmente, se construye el modelo con todos los datos y se valida sobre el conjunto total.
 - La precisión del clasificador se obtiene como la media de las precisiones anteriores.

Evaluación (v)

- **Técnicas de evaluación (iii)**
 - **Validación cruzada con k pliegues:** similar a la validación cruzada, pero dividiendo el conjunto de datos iniciales en k subconjuntos.
 - Se toma el primer subconjunto para validar y se construye el modelo con la unión del resto de subconjuntos.
 - Se repite el paso anterior con cada uno de los $k - 1$ subconjuntos restantes.
 - Finalmente, se construye el modelo con todos los datos y se valida sobre el conjunto total.
 - La precisión del clasificador se obtiene como la media de las precisiones anteriores.

Evaluación (vi)

- Las técnicas de evaluación anteriores han sido descritas usando la precisión del clasificador.
- Sin embargo, pueden usarse cuando cualquiera de las medidas de evaluación que se describirán a continuación.

Evaluación (vi)

- **Medidas de evaluación.** La medida de evaluación empleada depende, entre otros factores, del contexto de aplicación o de la tarea a evaluar.
 - **Clasificación.** Normalmente, se emplea como medida la **precisión**, que se obtiene como la proporción de instancias clasificadas correctamente.
 - No obstante, pueden considerarse otras medidas o estadísticos como los que se enumeran a continuación.
 - Para la definición de algunos de ellos es importante la **matriz de confusión**.

MATRIZ DE CONFUSIÓN

Descripción (i)

- Supongamos que se dispone de k clases, C_1, C_2, \dots, C_k , y sean
 - a_i el número de ejemplos que pertenecen a la clase C_i ,
 - p_j el número de ejemplos que el clasificador clasifica como perteneciente a la clase C_j ,
 - n_{ij} el número de ejemplos de la clase C_i que se clasifican como pertenecientes a la clase C_j , y
 - N el número total de ejemplos.
- La disposición de estos valores en forma de tabla recibe el nombre de **matriz de confusión**.

Matriz de confusión

Descripción (ii)

		Clase predicha						a_i
		C_1	C_2	\dots	C_j	\dots	C_k	
Clase real	C_1	n_{11}	n_{12}	\dots	n_{1j}	\dots	n_{1k}	a_1
	C_2	n_{21}	n_{22}	\dots	n_{2j}	\dots	n_{2k}	a_2
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	C_i	n_{i1}	n_{i2}	\dots	n_{ij}	\dots	n_{ik}	a_i
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	C_k	n_{k1}	n_{k2}	\dots	n_{kj}	\dots	n_{kk}	a_k
	p_j	p_1	p_2	\dots	p_j	\dots	p_k	N

Ejemplo

	C_1	C_2	C_3	C_4	a_i
C_1	130	74	2	6	212
C_2	96	99	6	16	217
C_3	3	4	207	4	218
C_4	6	12	4	177	199
P_j	235	189	219	203	845

ESTADÍSTICOS

True Positive (TP) rate

- También conocida como **Recall**, es la proporción de ejemplos que fueron clasificados como pertenecientes a la clase entre todos los ejemplos que verdaderamente pertenecen a esa clase. Es decir, es la **proporción de ejemplos que han sido bien clasificados**.

Matemáticamente se obtiene desde la siguiente expresión:

$$TP(C_i) = \frac{n_{ii}}{\sum_{j=1}^k n_{ij}} = \frac{n_{ii}}{a_i}$$

True Positive (TP) rate. Ejemplo

	C_1	C_2	C_3	C_4	a_i
C_1	130	74	2	6	212
C_2	96	99	6	16	217
C_3	3	4	207	4	218
C_4	6	12	4	177	199
p_j	235	189	219	203	845

$$TP(C_1) = \frac{130}{212} = 0.613$$

$$TP(C_2) = \frac{99}{217} = 0.456$$

$$TP(C_3) = \frac{207}{218} = 0.951$$

$$TP(C_4) = \frac{177}{199} = 0.889$$

False Positive (FP) rate

- Es la proporción de ejemplos que fueron clasificados como pertenecientes a la clase, pero que en realidad pertenecen a otra clase. Es decir, es la **proporción de ejemplos que han sido mal clasificados**.

Matemáticamente se obtiene desde la siguiente expresión:

$$FP(C_i) = \frac{\left(\sum_{i=1}^k n_{ij}\right) - n_{ii}}{\left(\sum_{j=1}^k a_j\right) - a_i}$$

False Positive (TP) rate. Ejemplo

	C_1	C_2	C_3	C_4	a_i
C_1	130	74	2	6	212
C_2	96	99	6	16	217
C_3	3	4	207	4	218
C_4	6	12	4	177	199
p_j	235	189	219	203	845

$$FP(C_1) = \frac{96 + 3 + 6}{217 + 218 + 199} = 0.165$$

$$FP(C_2) = \frac{74 + 4 + 12}{212 + 218 + 199} = 0.173$$

$$FP(C_3) = \frac{2 + 6 + 4}{212 + 217 + 199} = 0.019$$

$$FP(C_4) = \frac{6 + 16 + 4}{212 + 217 + 218} = 0.040$$

Precision

- Es la proporción de ejemplos que realmente pertenecen a la clase de entre los que fueron clasificados como pertenecientes a la clase.

Matemáticamente se obtiene desde la siguiente expresión:

$$Precision(C_i) = \frac{n_{ii}}{\sum_{j=1}^k n_{ij}}$$

Precision. Ejemplo

	C_1	C_2	C_3	C_4	a_i
C_1	130	74	2	6	212
C_2	96	99	6	16	217
C_3	3	4	207	4	218
C_4	6	12	4	177	199
P_j	235	189	219	203	845

$$Precision(C_1) = \frac{130}{235} = 0.553$$

$$Precision(C_2) = \frac{99}{189} = 0.524$$

$$Precision(C_3) = \frac{207}{219} = 0.945$$

$$Precision(C_4) = \frac{177}{203} = 0.872$$

F-measure

- Media armónica entre **Precision** y **Recall**.

Matemáticamente se obtiene desde la siguiente expresión:

$$F - measure(C_i) = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Precision. Ejemplo

	C_1	C_2	C_3	C_4	a_i
C_1	130	74	2	6	212
C_2	96	99	6	16	217
C_3	3	4	207	4	218
C_4	6	12	4	177	199
p_j	235	189	219	203	845

$$F(C_1) = 2 \cdot \frac{0.553 \cdot 0.613}{0.553 + 0.613} = 0.582$$

$$F(C_2) = 2 \cdot \frac{0.524 \cdot 0.456}{0.524 + 0.456} = 0.488$$

$$F(C_3) = 2 \cdot \frac{0.945 \cdot 0.950}{0.945 + 0.950} = 0.947$$

$$F(C_4) = 2 \cdot \frac{0.872 \cdot 0.889}{0.872 + 0.889} = 0.881$$

Índice kappa (i)

- Se trata de un índice de concordancia propuesto por Cohen que, en clasificación, mide la concordancia existente entre las frecuencias de ocurrencia de cada clase y las frecuencias obtenidas al usar el clasificador.

Matemáticamente se obtiene desde la siguiente expresión:

$$K = \frac{P_o - P_e}{1 - P_e}$$

donde

$$P_o = \frac{n_{11} + n_{22} + \dots + n_{kk}}{N}$$

y

$$P_e = \frac{a_1 \cdot p_1 + a_2 \cdot p_2 + \dots + a_k \cdot p_k}{N^2}$$

Índice kappa (ii)

- Landis y Koch propusieron, y desde entonces se usa ampliamente, la siguiente escala de valoración del índice kappa:

Valor de kappa	Grado de concordancia
$K < 0.00$	No hay
$0.00 < K < 0.20$	Insignificante
$0.20 \leq K < 0.40$	Discreta
$0.40 \leq K < 0.60$	Moderada
$0.60 \leq K < 0.80$	Sustancial
$0.80 \leq K \leq 1.00$	Casi perfecta

Índice kappa. Ejemplo

	C_1	C_2	C_3	C_4	a_i
C_1	130	74	2	6	212
C_2	96	99	6	16	217
C_3	3	4	207	4	218
C_4	6	12	4	177	199
P_j	235	189	219	203	845

$$P_o = \frac{130 + 99 + 207 + 177}{845}$$

$$= 0.724$$

$$P_e = \frac{212 \cdot 235 + \dots + 199 \cdot 203}{845 \cdot 845}$$

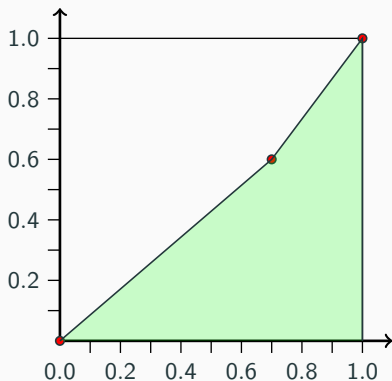
$$= 0.250$$

$$K = \frac{0.724 - 0.250}{1 - 0.250} = 0.632$$

Área bajo la curva ROC (i)

- En unos ejes de coordenadas (x, y) , la curva ROC viene determinada por los puntos $(0, 0)$, $(FP(C_i))$, $TP(C_i)$ y $(1, 1)$.
- Es útil porque muestra de manera sencilla y gráfica el comportamiento del clasificador para la correspondiente clase.

Área bajo la curva ROC (ii)

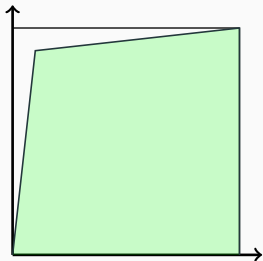


$$TP(C_i) = 0.614$$

$$FP(C_i) = 0.692$$

$$\text{Área ROC}(C_i) = 0.459$$

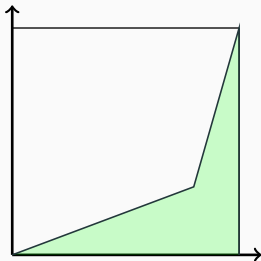
Área bajo la curva ROC (ii)



Buen clasificador

Alto TP

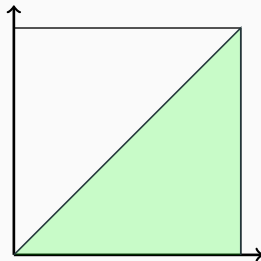
Bajo FP



Mal clasificador

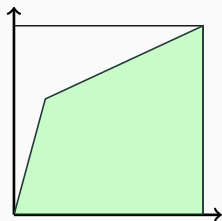
Bajo TP

Alto FP



Mal clasificador

Área bajo la curva ROC. Ejemplo

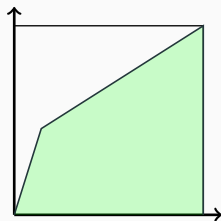


Clase C_1

$$TP(C_1) = 0.613$$

$$FP(C_1) = 0.166$$

$$ROC(C_1) = 0.784$$

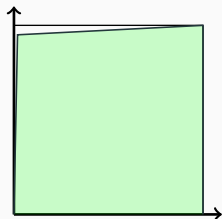


Clase C_2

$$TP(C_2) = 0.456$$

$$FP(C_2) = 0.143$$

$$ROC(C_2) = 0.758$$

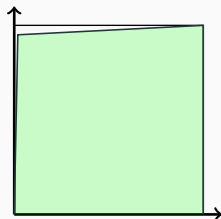


Clase C_3

$$TP(C_3) = 0.950$$

$$FP(C_3) = 0.019$$

$$ROC(C_3) = 0.977$$



Clase C_4

$$TP(C_4) = 0.889$$

$$FP(C_4) = 0.040$$

$$ROC(C_4) = 0.881$$

Mean absolute error (MAE)

- Mide el error absoluto medio entre el verdadero valor de la clase y el valor predicho.

Matemáticamente se obtiene desde la siguiente expresión:

$$MAE = \frac{|c_1 - ec_1| + |c_2 - ec_2| + \dots + |c_N - ec_N|}{N}$$

donde

- c_i , es el valor de clase del i -ésimo ejemplo, y
- ec_i , es el valor de clase del i -ésimo ejemplo predicho por el clasificador

Root mean squared error (RMSE)

- Mide el error cuadrático medio entre el verdadero valor de la clase y el valor predicho.

Matemáticamente se obtiene desde la siguiente expresión:

$$RMSE = \sqrt{\frac{(c_1 - ec_1)^2 + (c_2 - ec_2)^2 + \dots + (c_N - ec_N)^2}{N}}$$

donde

- c_i , es el valor de clase del i -ésimo ejemplo, y
- ec_i , es el valor de clase del i -ésimo ejemplo predicho por el clasificador

Relative absolute error (RAE)

- Mide el error absoluto relativo entre el verdadero valor de la clase y el valor predicho.

Matemáticamente se obtiene desde la siguiente expresión:

$$RAE = \frac{|c_1 - ec_1| + |c_2 - ec_2| + \dots + |c_N - p_N|}{|\bar{c} - ec_1| + |\bar{c} - ec_2| + \dots + |\bar{c} - ec_N|}$$

donde

- c_i , es el valor de clase del i -ésimo ejemplo,
- ec_i , es el valor de clase del i -ésimo ejemplo predicho por el clasificador, y
- \bar{c} , es el valor medio de clase

Root relative squared error (RRSE)

- Mide el error cuadrático relativo entre el verdadero valor de la clase y el valor predicho.

Matemáticamente se obtiene desde la siguiente expresión:

$$RRSE = \frac{(c_1 - ec_1)^2 + (c_2 - ec_2)^2 + \dots + (c_N - ec_N)^2}{(\bar{c} - c_1)^2 + (\bar{c} - c_2)^2 + \dots + (\bar{c} - c_N)^2}$$

donde

- c_i , es el valor de clase del i -ésimo ejemplo,
- ec_i , es el valor de clase del i -ésimo ejemplo predicho por el clasificador, y
- \bar{c} , es el valor medio de clase

Esta obra está bajo una licencia de Creative Commons.
Reconocimiento - No comercial - Compartir igual

