Evaluación de modelos de respuesta cualitativa binaria

En este anexo se presenta un resumen de las principales métricas que se utilizan habitualmente para la evaluación de la precisión de los modelos de respuesta cualitativa binaria. La terminología, las definiciones y las relaciones entre dichas métricas están extraídas de la página web <http://es.wikipedia.org/wiki/Curva_ROC>.

**Métricas discretas**

Al ser binario el *output*, podemos definir una matriz de tamaño 2x2 que contenga el valor real en las filas y la previsión del modelo en las columnas, mediante el establecimiento de un punto de corte por encima del cual consideramos que el modelo prevé que un determinado caso será positivo, y viceversa:

Valor real

P’

Falsos positivos

Verdaderos negativos

Verdaderos positivos

Falsos negativos

Previsión del modelo

N’

P N

La matriz anterior nos proporciona cuatro métricas primarias de evaluación:

* Verdaderos Positivos (VP): casos que el modelo predice como positivos y realmente son positivos.
* Verdaderos Negativos (VP): casos previstos por el modelo como negativos y que realmente son negativos.
* Falsos Positivos (FP): casos negativos que el modelo predice como positivos, también denominado error de tipo I.
* Falsos Negativos (FP): casos positivos que el modelo predice como negativos, también denominado error de tipo II.

A partir de las métricas primarias, se obtienen cuatro métricas derivadas dividiendo cada una de ellas por el total de casos reales positivos o negativos, según corresponda:

* Razón de Verdaderos Positivos o Sensibilidad: .
* Razón de Falsos Positivos: .
* Razón de Verdaderos Negativos o Especifidad: .
* Razón de Falsos Negativos: .

De forma similar, dividiendo por el total de casos previstos positivos o negativos, según corresponda, obtenemos otras cuatro métricas derivadas:

* Valor Predictivo Positivo: .
* Valor Predictivo Negativo: .
* Razón de Falsas Alarmas: .
* Razón de Falsos Descartes: .

Por último, podemos obtener la precisión del modelo dividiendo el número total de aciertos, tanto positivos como negativos, por el número total de casos:

* Precisión: .

**Métricas continuas**

Alternativamente, podemos definir métricas y herramientas gráficas que no dependen de un punto de corte, tales como:

* Curva ROC (Característica Operativa del Receptor, por sus siglas en inglés): cada punto representa un par de valores {RVP, RFP} obtenido a partir de un determinado punto de corte. A partir de la curva ROC, se pueden calcular diversas métricas, de las cuales la más utilizada es el área bajo la curva ROC (AUC). Ésta se puede interpretar como la probabilidad con la que el modelo puntuará más alto un caso positivo que uno negativo, es decir que ordene correctamente dos casos extraídos aleatoriamente. Obviamente, es un valor entre 0.5 y 1.
* Coeficiente de Gini (CG): representa el poder de discriminación del modelo, medido entre 0 (nulo poder) y 1 (discriminación perfecta). Se obtiene a partir de la curva de Lorenz, si bien está relacionado con el área bajo la curva ROC: CG.
* Coeficiente de Kolmogorov-Smirnov (KS). Es la máxima diferencia entre la razón de verdaderos positivos y la razón de falsos positivos: . En consecuencia, es también la máxima distancia entre la curva ROC y la línea de no discriminación (recta de 45º y pendiente unitaria que se obtiene con los pares RVP y RFP resultantes de aplicar un modelo uniforme aleatorio.
* Desviación típica de los residuos: es la raíz cuadrada de la varianza, que representa la media aritmética de las desviaciones con respecto a la [media](http://es.wikipedia.org/wiki/Media_aritm%C3%A9tica), elevadas al cuadrado.