

# Tema 6: Comparación, selección y evaluación de modelos

Minería de Datos

# ¿Para que sirve la evaluación?

- Hemos construido un modelo o varios basados en los datos, pero necesitamos saber como de bueno es.
- Necesitamos comparar los datos que hemos obtenido.
- ¿Nuestro modelo generaliza?
- ¿Este modelo es útil en fase de explotación?

## > Selección de métricas

- El experto en aprendizaje automático de Fayrix habla de las **métricas de rendimiento** que se utilizan comúnmente en la ciencia de datos para evaluar y realizar los modelos de aprendizaje automático.
- 1º Entender la tarea:
  - Según los requisitos previos, debemos comprender qué tipo de problemas estamos tratando de resolver:
    - Clasificación
    - Regresión
    - Categorización

## > Clasificación

- Matriz de confusión

Esta matriz se utiliza para evaluar la precisión de un clasificador y se presenta en la tabla a continuación.

		Resultado de la predicción		
		Positivo	Negativo	
Valor actual	Positivo	TP	FN	TP + FN
	Negativo	FP	TN	FP + TN

## > Clasificación

### Matriz de confusión

Esta matriz se utiliza para evaluar la precisión de un clasificador y se presenta en la tabla a continuación.

		Resultado de la predicción		
		Positivo	Negativo	
Valor actual	Positivo	TP	FN	TP + FN
	Negativo	FP	TN	FP + TN

**Error tipo I**  
(falso positivo)



**Error tipo II**  
(falso negativo)



## > Clasificación

### Exactitud

Indica el número de elementos clasificados correctamente en comparación con el número total.

Tenga en cuenta que la métrica de exactitud tiene limitaciones: no funciona bien con las clases desequilibradas que pueden tener muchos elementos de la misma clase e incluir algunas otras clases.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

TP = total positivos

TN = total negativos

FP = falsos positivos

FN = falsos negativos

## > Clasificación

### Exhaustividad / Sensibilidad

La métrica de exhaustividad muestra la cantidad de verdaderos positivos que el modelo ha clasificado en función del número total de valores positivos.

$$recall = \frac{TP}{TP + FN}$$

TP = total positivos

TN = total negativos

FP = falsos positivos

FN = falsos negativos

### > Clasificación

### Exhaustividad / Sensibilidad

La métrica de exhaustividad muestra la cantidad de verdaderos positivos que el modelo ha clasificado en función del número total de valores positivos.

$$recall = \frac{TP}{TP + FN}$$

TP = total positivos

TN = total negativos

FP = falsos positivos

FN = falsos negativos



## > Clasificación

### Precisión

Esta métrica representa el número de verdaderos positivos que son realmente positivos en comparación con el número total de valores positivos predichos.

$$precision = \frac{TP}{TP + FP}$$

TP = total positivos

TN = total negativos

FP = falsos positivos

FN = falsos negativos

## > Clasificación

### Puntuación F1

Esta métrica es la combinación de las métricas de precisión y exhaustividad y sirve de compromiso entre ellas. La mejor puntuación F1 es igual a 1 y la peor a 0.

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

## > Regresión

### Error Medio Absoluto (EMA)

Esta métrica de regresión es el valor medio de la diferencia absoluta entre el valor real y el valor predicho.

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |\text{original}_t - \text{predict}_t|$$

## > Regresión

### Error Cuadrático Medio (ECM)

El error cuadrático medio (ECM) calcula el valor medio de la diferencia al cuadrado entre el valor real y el predicho para todos los puntos de datos.

En esta métrica, el impacto de los errores es mayor. Cuanto menor sea el ECM, más precisas serán nuestras predicciones. ECM = 1 es el punto óptimo.

$$MSE = \frac{1}{n} \sum_{t=1}^n (original_t - predict_t)^2$$

***El MSE tiene algunas ventajas frente al MAE:***

1. El **MSE** destaca grandes errores entre los pequeños.
2. El **MSE** es diferenciable, lo que ayuda a encontrar los valores mínimos y máximos utilizando los métodos matemáticos de manera más efectiva.

## > Regresión

### Raíz del Error Cuadrático Medio (RECM)

El RECM es la raíz cuadrada del ECM. Es fácil de interpretar en comparación con el ECM y utiliza valores absolutos más pequeños, lo que es útil para los cálculos informáticos.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (original_t - predict_t)^2}$$

## > Clasificación por categorías

### Coefficiente Tau de Kendall

El coeficiente tau de Kendall muestra la correlación entre las dos listas de elementos clasificados según el número de pares concordantes y discordantes: en cada caso tenemos dos rangos (máquina y predicción humana). En primer lugar, los elementos clasificados se convierten en una matriz de comparación por pares con la correlación entre el rango actual y otros. Un par concordante significa que el rango de algoritmo se correlaciona con el rango humano. En el caso opuesto será un par discordante. Por lo tanto, este coeficiente se define de la siguiente manera

$$\tau = \frac{(\text{número de pares coincidentes}) - (\text{número de pares no coincidentes})}{n * (n - 1) / 2}$$

Los valores de  $\tau$  varían de 0 a 1. Cuanto más  $|\tau|$  se aproxime a 1, tanto mejor será el ranking. Por ejemplo, cuando el valor de  $\tau$  se aproxima a -1, la clasificación es igual de precisa, sin embargo, el orden de sus ítems debería ser inverso. Esto es bastante consistente con los indicadores de estimación que asignan el rango más alto a los mejores valores, mientras que durante el ranking humano los mejores reciben los rangos más bajos.  $\tau = 0$  indica la falta de correlación entre los rangos.

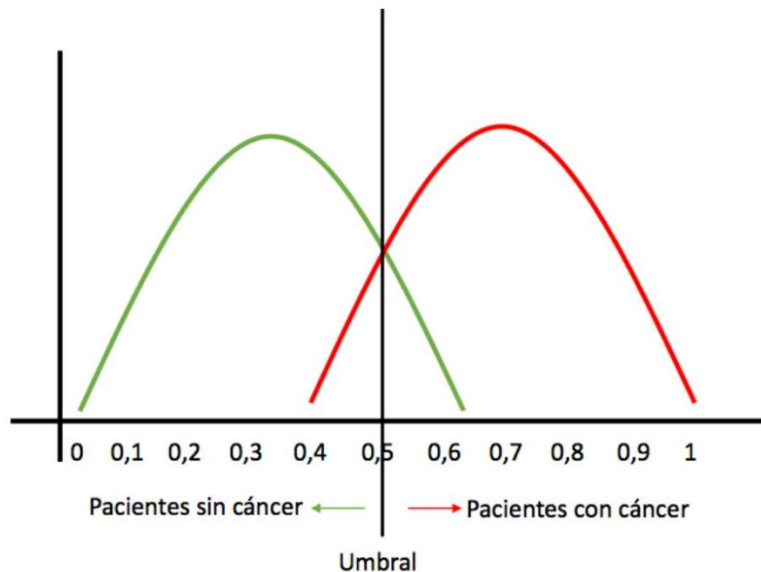
### > Clasificación

- **Curva ROC y Área bajo la curva (AUC)**
  - Esta es una de las métricas de evaluación más importante para verificar el rendimiento de cualquier modelo de clasificación.
  - ROC viene de las características de funcionamiento del receptor y AUC del área bajo la curva.
  - La curva ROC nos dice qué tan bueno puede distinguir el modelo entre dos cosas. Mejores modelos pueden distinguir con precisión entre los dos, mientras que un modelo pobre tendrá dificultades para distinguir entre los dos.

## > Clasificación

### Curva ROC y Área bajo la curva (AUC)

Supongamos que tenemos un modelo que predice si un paciente tiene cáncer o no, el resultado es el siguiente:

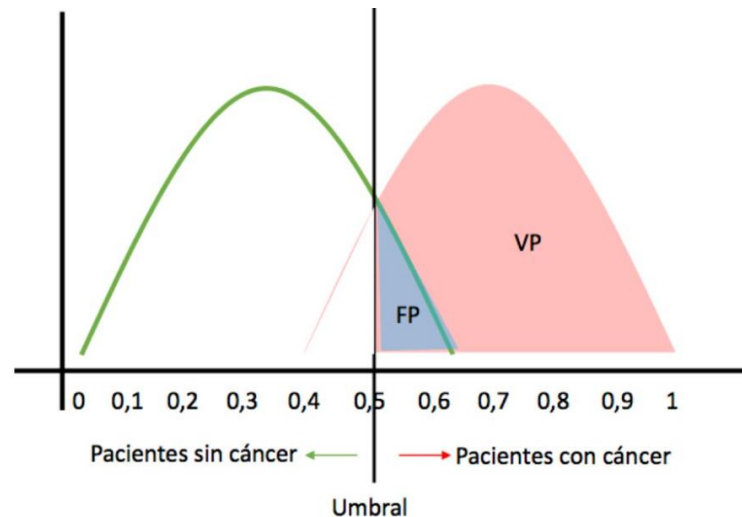




### > Clasificación

### Curva ROC y Área bajo la curva (AUC)

Ahora debemos elegir un valor en donde establecemos el corte o un valor umbral, por encima del cual predeciremos a todos como positivos, tienen cáncer, y por debajo del cual predeciremos como negativos, NO cáncer. Este umbral lo establecemos en 0.5.

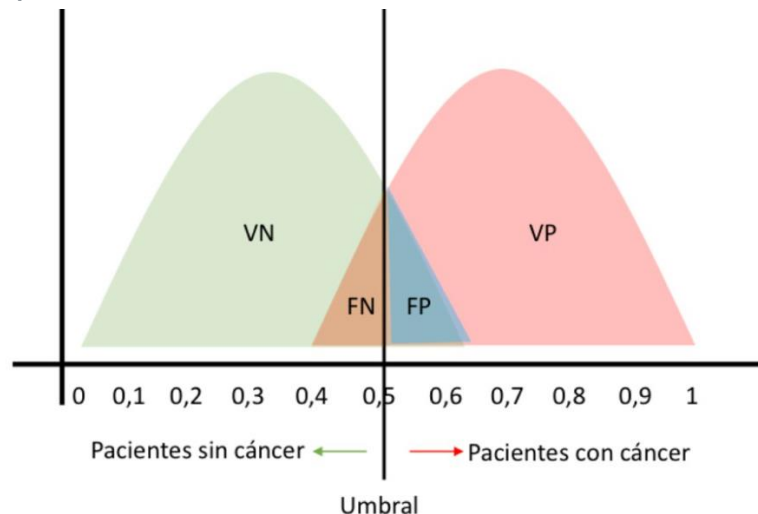


### > Clasificación

### Curva ROC y Área bajo la curva (AUC)

Tomando los conceptos aprendidos en la matriz de confusión, todos los valores positivos por encima del umbral serán “verdaderos positivos” y los valores negativos por encima del umbral serán “falsos positivos”, ya que se predicen incorrectamente como positivos.

Todos los valores negativos por debajo del umbral serán “verdaderos negativos” y los valores positivos por debajo del umbral serán “falsos negativos”, ya que se pronostican incorrectamente como negativos.



### > Clasificación

### Curva ROC y Área bajo la curva (AUC)

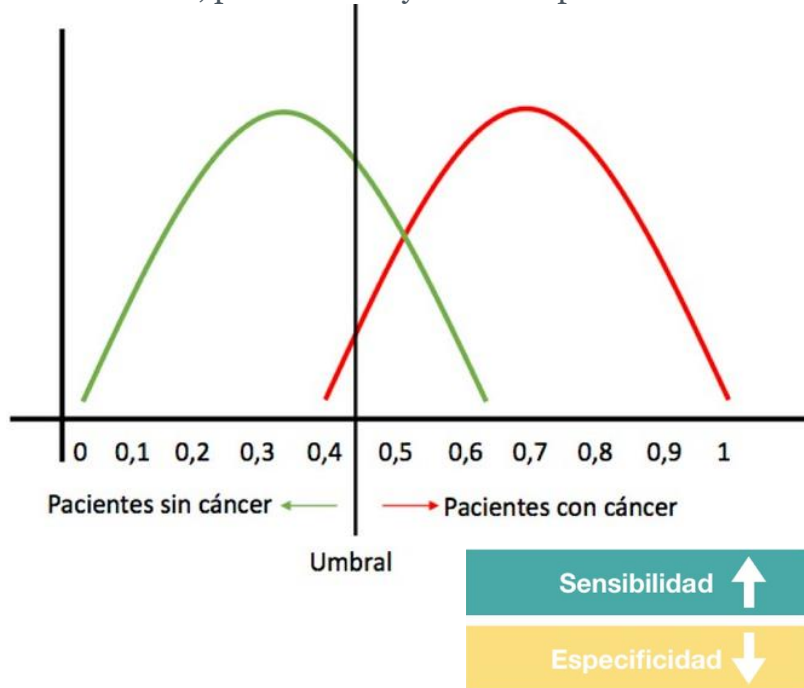
Aquí, tenemos una idea básica de que el modelo predice valores correctos e incorrectos con respecto al conjunto de umbrales.

Recordamos dos conceptos previos:

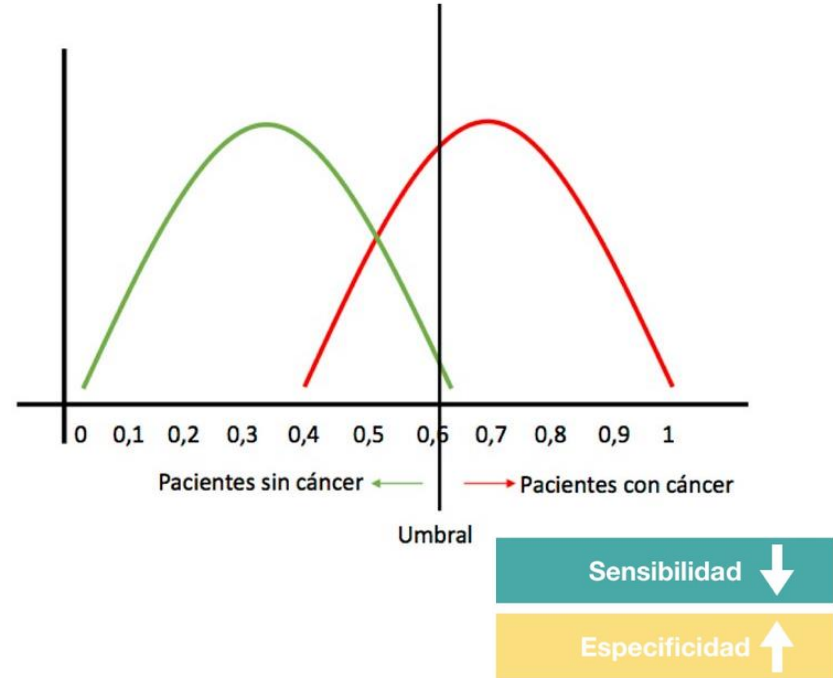
La sensibilidad o recall, es la proporción de pacientes que se identificaron correctamente por tener cáncer, es decir verdadero positivo, sobre el número total de pacientes que realmente tienen la enfermedad.

Por su parte, especificidad es la proporción de pacientes que se identificaron correctamente por no tener cáncer, verdadero negativo, sobre el número total de pacientes que no tienen la enfermedad.

Si volvemos a nuestra gráfica anterior, si disminuimos el valor del umbral, obtenemos más valores negativos, aumentando la sensibilidad, pero disminuyendo la especificidad.



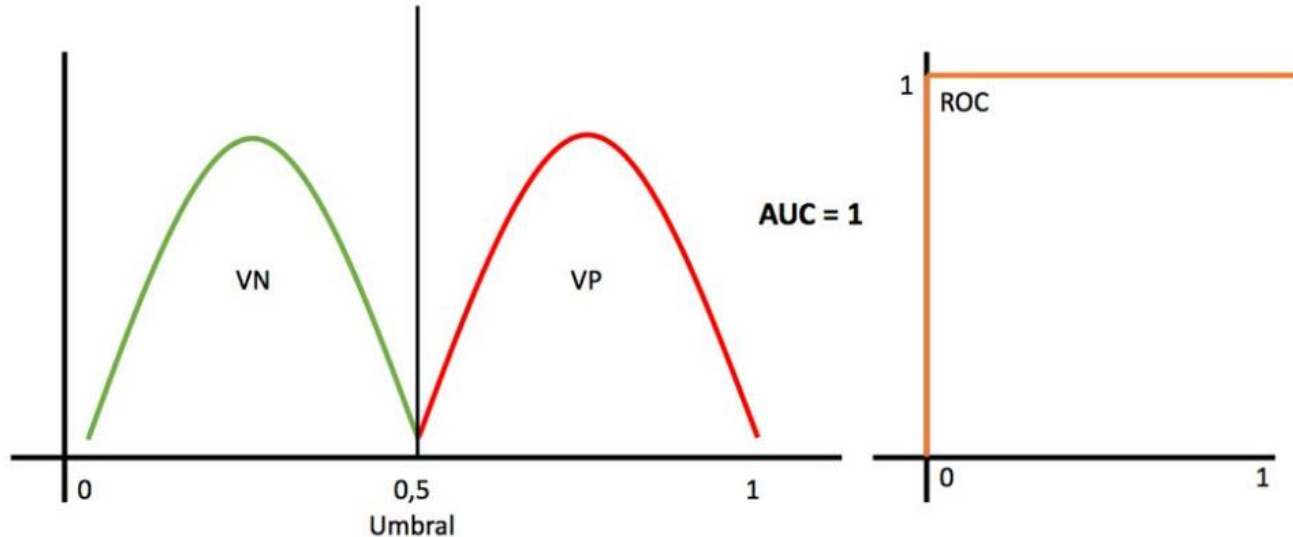
En cambio, si aumentamos el umbral, obtenemos más valores negativos, lo que aumenta la especificidad y disminuye la sensibilidad.



## > Clasificación

### Área bajo la curva (AUC)

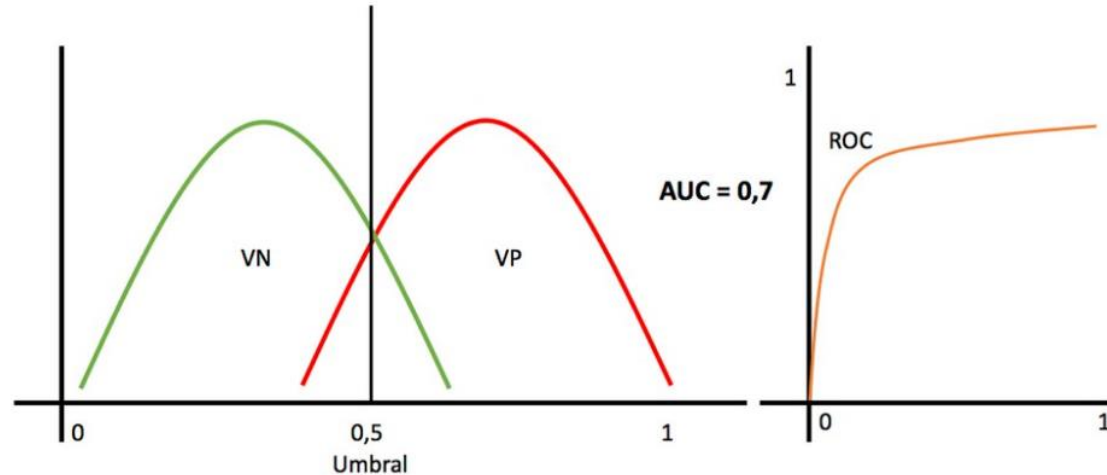
El AUC es el área bajo la curva ROC. Este puntaje nos da una buena idea de qué tan bien funciona el modelo.



## > Clasificación

### Área bajo la curva (AUC)

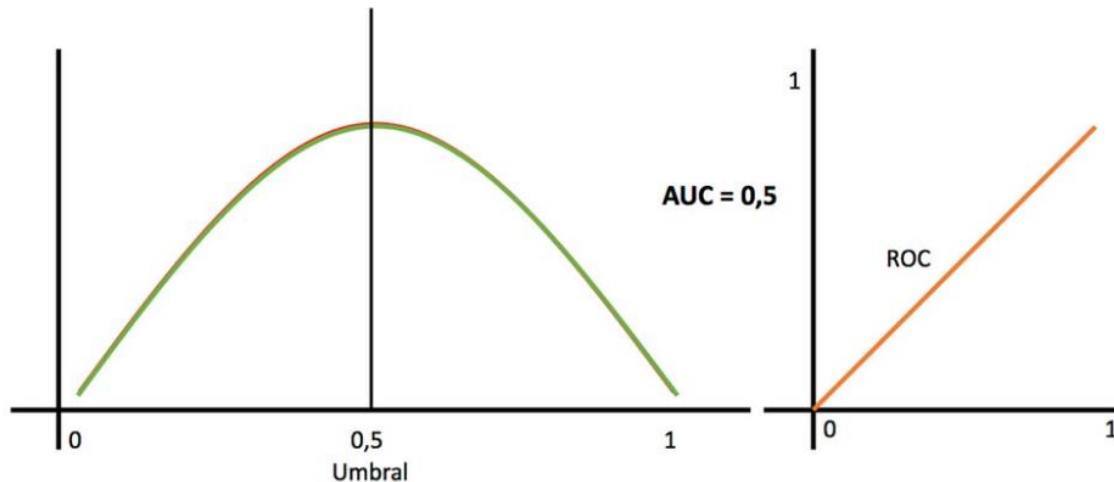
Cuando dos distribuciones se superponen, introducimos errores. Dependiendo del umbral, podemos minimizarlos o maximizarlos. Cuando AUC es 0.7, significa que hay 70% de probabilidad de que el modelo pueda distinguir entre clase positiva y clase negativa.



### > Clasificación

### Área bajo la curva (AUC)

Esta es la peor situación. Cuando el AUC es aproximadamente 0.5, el modelo no tiene capacidad de discriminación para distinguir entre clase positiva y clase negativa.



# Gracias