

# Métricas de clasificación

Cristian López Del Alamo

[clopezd@utec.edu.pe](mailto:clopezd@utec.edu.pe)

Willian Bermudez Navarro

Alejandra Bustinza Cornejo

Anibal Lajo Atayupanqui

IPRODAM3D - Research group

2022



# Métricas de clasificación

Cristian López Del Alamo

[clopezd@utec.edu.pe](mailto:clopezd@utec.edu.pe)

Willian Bermudez Navarro

Alejandra Bustinza Cornejo

Anibal Lajo Atayupanqui

IPRODAM3D - Research group

2022



# 1

## Introducción

UTEC

## ¿Qué son las Métricas de Clasificación?

- ★ Una métrica de clasificación es un número que mide el rendimiento de su modelo
- ★ ¿Cómo sabemos si nuestro modelo funciona correctamente?
- ★ Permiten evaluar el rendimiento de los modelos de aprendizaje automático
- ★ Normalmente, el rendimiento se presenta en un rango de 0 a 1 (aunque no siempre)

# Matriz de Confusión

Es una tabla cruzada entre los valores actuales y los valores Predichos.

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

		PREDICTED		
		Positive (1)	Negative (0)	Total
ACTUAL	Positive (1)	TP = 20	FN = 5	25
	Negative (0)	FP = 10	TN = 15	25
	Total	30	20	50

## Precision - Precisión

¿Qué proporción de Positivos pronosticados es verdaderamente Positivo?.

$$\textit{Precisión} = \frac{TP}{TP+FP}$$

## Recall - Exhaustividad

¿Qué proporción de Positivos reales se clasifica correctamente?

$$\textit{Recall} = \frac{TP}{TP+FN}$$

# Accuracy-Exactitud

- ★ La exactitud es la proporción de resultados verdaderos entre el número total de casos examinados.
- ★ Considera la suma de los elementos correctamente clasificados en el numerador y todas las entradas que hay en la matriz de confusión en el denominador.

¿Estoy interesado en predecir el mayor número de individuos en la clase correcta, sin preocuparme por la distribución de clases y otros indicadores?

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

# Balanced Accuracy

- Esta métrica es particularmente útil cuando las dos clases están desequilibradas, es decir, una clase aparece mucho más que la otra.

¿Cuán probable es que un individuo de esa clase sea clasificado correctamente?

		PREDICTED		
		Positive (1)	Negative (0)	Total
ACTUAL	Classes			
	Positive (1)	TP = 20	FN = 5	25
	Negative (0)	FP = 10	TN = 15	25
Total		30	20	50

$$\text{Balanced Accuracy} = \frac{\frac{TP}{Total_{row1}} + \frac{TN}{Total_{row2}}}{2}$$



# F1 Score

- Evalúa el desempeño del modelo de clasificación a partir de la matriz de confusión
- Es un número entre 0 y 1 y es la media armónica de precisión y Recall.

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

# F1-Score Binary

$$Precision = \frac{20}{30} = 0.66 \quad Recall = \frac{20}{25} = 0.80$$

$$F1-Score = 2 \cdot \left( \frac{0.66 \cdot 0.80}{0.66 + 0.80} \right) = 0.72$$

		PREDICTED	
		Positive (1)	Negative (0)
ACTUAL	Classes		
	Positive (1)	TP = 20	FN = 5
	Negative (0)	FP = 10	TN = 15
Total		30	20

# F1-Score Multiclass

- ★ F1-Score debe involucrar a todas las clases
- ★ Surgen 2 nuevas métricas:
  - ❑ Micro F1-Score
  - ❑ Macro F1-Score

# Macro-Score

Se calculan con una Precisión promedio para cada clase predicha y el Recall promedio para cada clase real

$$Precision_k = \frac{TP_k}{TP_k + FP_k}$$

$$Recall_k = \frac{TP_k}{TP_k + FN_k}$$

$$MacroAveragePrecision = \frac{\sum_{k=1}^K Precision_k}{K}$$

$$MacroAverageRecall = \frac{\sum_{k=1}^K Recall_k}{K}$$

$$Macro\ F1-Score = 2 * \left( \frac{MacroAveragePrecision * MacroAverageRecall}{MacroAveragePrecision^{-1} + MacroAverageRecall^{-1}} \right)$$

# Micro-Score

La idea de Micro F1-score es considerar todas a las unidades juntas, sin tomar en cuenta diferencias entre las clases

$$\text{Micro Average Precision} = \frac{\sum_{k=1}^K TP_k}{\sum_{k=1}^K \text{Total Column}_k} = \frac{\sum_{k=1}^K TP_k}{\text{Grand Total}}$$

$$\text{Micro Average F1} = \frac{\sum_{k=1}^K TP_k}{\text{Grand Total}}$$

$$\text{Micro Average Recall} = \frac{\sum_{k=1}^K TP_k}{\sum_{k=1}^K \text{Total Row}_k} = \frac{\sum_{k=1}^K TP_k}{\text{Grand Total}}$$

# Micro y Macro-Score Resumen

- ★ Macro F1-Score calcula la Precision y Recall a nivel de clase, por lo que cada clase tiene el mismo peso, ya sea grande o pequeña.
- ★ Micro F1-Score trabaja en base al tamaño: La Precision se calcula a nivel de conjunto de datos y cada unidad tiene la misma importancia.

# Cross Entropy

Cross-Entropy se utiliza para evaluar la similitud entre dos funciones de distribución.

$$H(p, q) = - \int_{D_x} p(x) \log q(x)$$

$$H(p, q) = - \sum_{D_x} p(x) \log q(x)$$

En la imagen anterior, están las fórmulas para calcular el Cross-Entropy, considerando las distribuciones genéricas  $p(x)$  y  $q(x)$  para variables  $X$  continuas o discretas, respectivamente, y  $D_x$  siendo el dominio de  $X$ .

En clasificación Multi-clase, tanto la respuesta como la predicción son consideradas como dos variables discretas aleatorias: asumen valores en  $\{1, \dots, K\}$  y cada número representa una clase distinta.

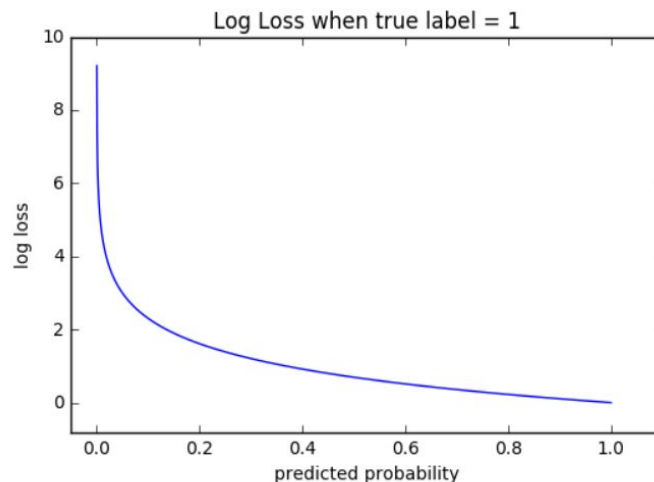
# Log Loss/Binary Cross Entropy

Donde  $p$  es la probabilidad de predecir 1.

Log loss disminuye ya que estamos bastante seguros de nuestra predicción de 1.

Se recomienda que deba introducir pesos de clase para penalizar más los errores minoritarios.

$$-(y \log(p) + (1 - y) \log(1 - p))$$





# Categorical Cross entropy

Si hay N muestras que pertenecen a M clases, entonces la Entropía cruzada categórica es la suma de los valores de -ylogp:

$$\text{LogarithmicLoss} = \frac{-1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} * \log(p_{ij})$$

- $y_{ij}$  es 1 si la muestra  $i$  pertenece a la clase  $j$
- $p_{ij}$  es la probabilidad que predice nuestro clasificador de que la muestra  $i$  pertenezca a la clase  $j$ .

En general, usamos la entropía cruzada categórica en el caso de las redes neuronales.  
Minimizar la entropía cruzada categórica da una mayor precisión para la clasificación.

# AUC

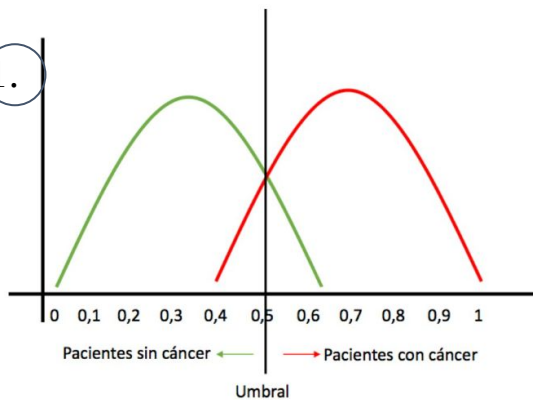
AUC: Área bajo la curva.  
ROC: Receiver Operator  
characteristic

$$\text{Sensitivity} = \text{TPR} = \text{Recall} = \frac{TP}{TP + FN}$$

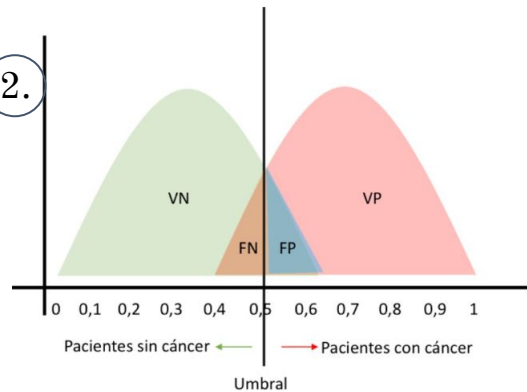
$$1 - \text{Specificity} = \text{FPR} = \frac{FP}{FP + TN}$$

# AUC

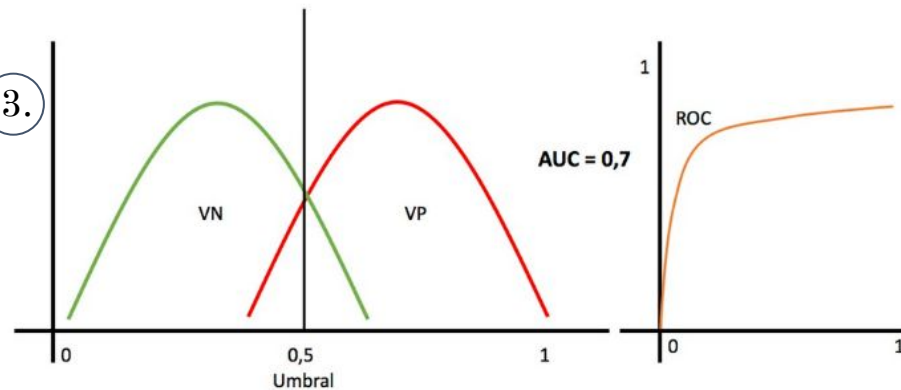
1.



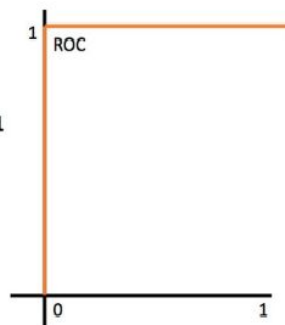
2.



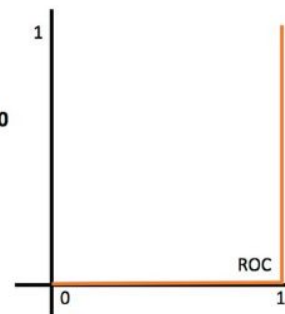
3.



AUC = 1



AUC = 0



# Matthews Correlation Coefficient (MCC)

- ★ MCC tiene un rango de  $[-1,1]$
- ★ Valores cercanos a 1 indican muy buena predicción, de hecho existe una fuerte correlación positiva entre la predicción y las etiquetas verdaderas.
- ★ Una fuerte correlación implica que las dos variables concuerdan fuertemente, por lo tanto, los valores pronosticados serán muy similares a la clasificación real.
- ★ MCC es igual a 0, no hay correlación entre nuestras variables: el clasificador está asignando aleatoriamente las unidades a las clases sin ningún vínculo con su verdadero valor de clase.
- ★ MCC también puede ser negativo, en este caso la relación entre las clases verdaderas y predichas es de tipo inverso. Incluso si esta es una situación altamente indeseable, esto sucede a menudo debido a errores de configuración en el modelado.
- ★ Una fuerte correlación inversa significa que el modelo aprendió cómo clasificar los datos pero cambia sistemáticamente todas las etiquetas.

# MCC para clasificación binaria

		PREDICTED		Total
		Positive (1)	Negative (0)	
ACTUAL	Positive (1)	TP = 20	FN = 5	25
	Negative (0)	FP = 10	TN = 15	25
	Total	30	20	50

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}}$$

# MCC para clasificación multiclase

Matrix C		PREDICTED classification				
	K Classes	k=1	k=2	k=3	k=4	Total
ACTUAL classification	k=1	$C_{11}=50$	37	24	39	150
	k=2	10	$C_{22}=480$	5	3	$t_{k=2} = 498$
	k=3	14	10	$C_{33}=765$	1	790
	k=4	0	2	9	$C_{44}=101$	112
	Total	74	$p_{k=2} = 529$	803	144	$s = 1550$

$$MCC = \frac{c \times s - \sum_k^K p_k \times t_k}{\sqrt{(s^2 - \sum_k^K p_k^2)(s^2 - \sum_k^K t_k^2)}}$$

- $c = \sum_k^K C_{kk}$  the total number of elements correctly predicted
- $s = \sum_i^K \sum_j^K C_{ij}$  the total number of elements
- $p_k = \sum_i^K C_{ki}$  the number of times that class  $k$  was predicted (column total)
- $t_k = \sum_i^K C_{ik}$  the number of times that class  $k$  truly occurred (row total)

- ★ MCC parece depender de elementos correctamente clasificados.
- ★ Esta suma incluye también los elementos mal clasificados por el modelo y cubre entidades multiplicativas que son más débiles que el producto  $c \times s$ .
- ★ Ya que no cambian cuando aplicamos diferentes modelos en el mismo conjunto de datos.

# Ventajas y desventajas de MCC

- ★ MCC incluye todas las entradas de la matriz de confusión tanto en el numerador como en el denominador.
- ★ No hay un valor mínimo fijo y cambia cada vez entre -1 y 0 según el número y la distribución real de las clases en el conjunto de datos inicial
- ★ Si hay resultados desequilibrados en la predicción del modelo, el valor final de MCC muestra fluctuaciones muy amplias dentro de su rango de  $[-1; +1]$  durante el período de entrenamiento del modelo