




Si h es un modelo constante $\boxed{h} \rightarrow \text{azul}$

accuracy = 98%



accuracy = tasa de predicciones correctas

$$\text{acc} = \frac{TP + TN}{TP + FP + FN + TN}$$



Si h es un modelo constante $\boxed{h} \rightarrow \text{azul}$

accuracy = 98%



accuracy = tasa de predicciones correctas

$$\text{acc} = \frac{TP + TN}{TP + FP + FN + TN}$$

MATRIZ DE CONFUSIÓN

		reales	
		yes	no
predichas	yes	TP	FP
	no	FN	TN

predichas

	reales	
	yes	no
yes	TP	FP
no	FN	TN

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

||

$$\text{recall}$$

$$\text{Specificity} = \frac{TN}{FP + TN}$$

La paradoja de los falsos positivos...

reales

predichas

	yes	no
yes	TP	FP
no	FN	TN

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

$$\text{Specificity} = \frac{TN}{FP + TN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

reales

predichos

	yes	no
yes	TP	FP
no	FN	TN

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

$$\text{Specificity} = \frac{TN}{FP + TN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

Combinándolas:

$$F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

ma's general

$$F_{\beta} = \frac{(1 + \beta^2) \cdot \text{precision} \cdot \text{recall}}{\beta^2 \text{ precision} + \text{recall}}$$

$\beta = 1$

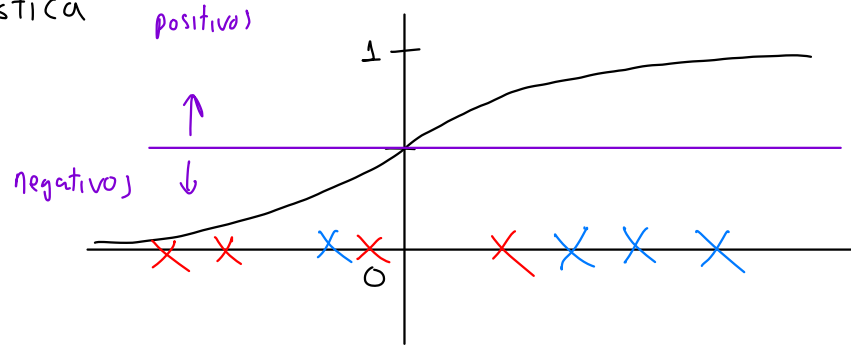
$$G\text{-mean} = \sqrt{\text{Sensitivity} \cdot \text{Specificity}}$$

Cuando el clasificador predice una **probabilidad** de pertenecer a una clase

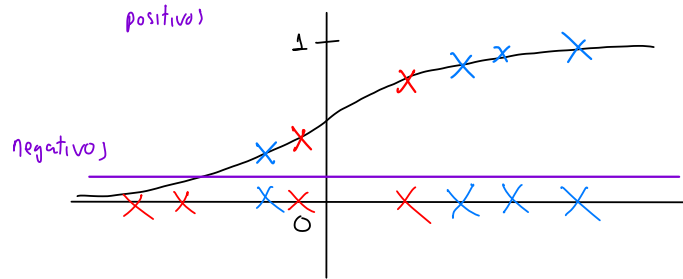
Ej: Regresión logística

Conf. Matrix

3	1
1	3



4	2
0	2

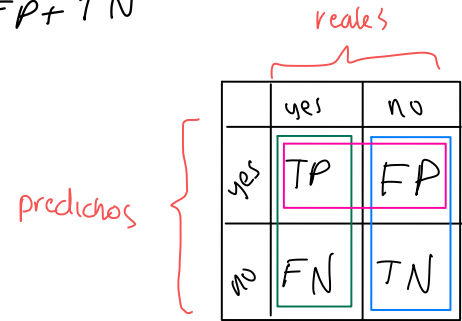


ROC (Receiver Operating Characteristic)

Gráfica que resume el comportamiento de un modelo al mostrar la tasa de falsos positivos y la tasa de verdaderos positivos con respecto a diferentes umbrales

$$\text{True positive rate (TPR)} = \text{recall} = \text{sensitivity} = \frac{TP}{TP + FN}$$

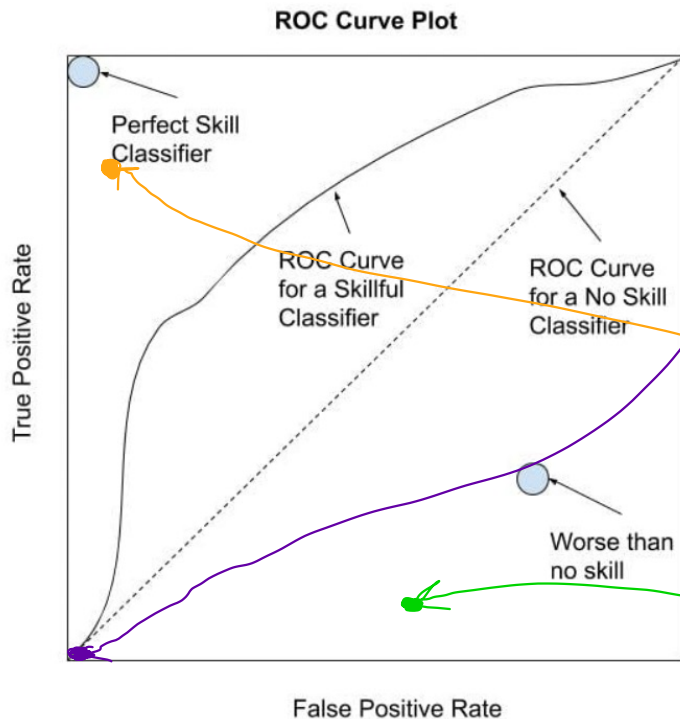
$$\text{False positive rate (FPR)} = 1 - \text{specificity} = \frac{FP}{FP + TN}$$



A confusion matrix diagram. The matrix is a 2x2 grid. The columns are labeled 'yes' and 'no' at the top, with a red bracket above them labeled 'reales'. The rows are labeled 'yes' and 'no' on the left, with a red bracket to the left labeled 'predichos'. The cells contain 'TP' (True Positive), 'FP' (False Positive), 'FN' (False Negative), and 'TN' (True Negative). The 'TP' cell is highlighted with a pink box, the 'FP' cell with a blue box, the 'FN' cell with a green box, and the 'TN' cell with a light blue box.

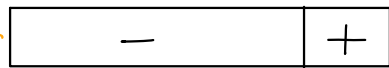
	yes	no
yes	TP	FP
no	FN	TN

ROC Curve



$FPR = 0 \quad TPR = 0$

Predicción
negativa
siempre

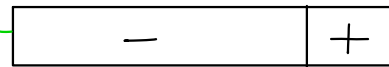


FP

FPR baja

TP

TPR alta



FP

FPR alta

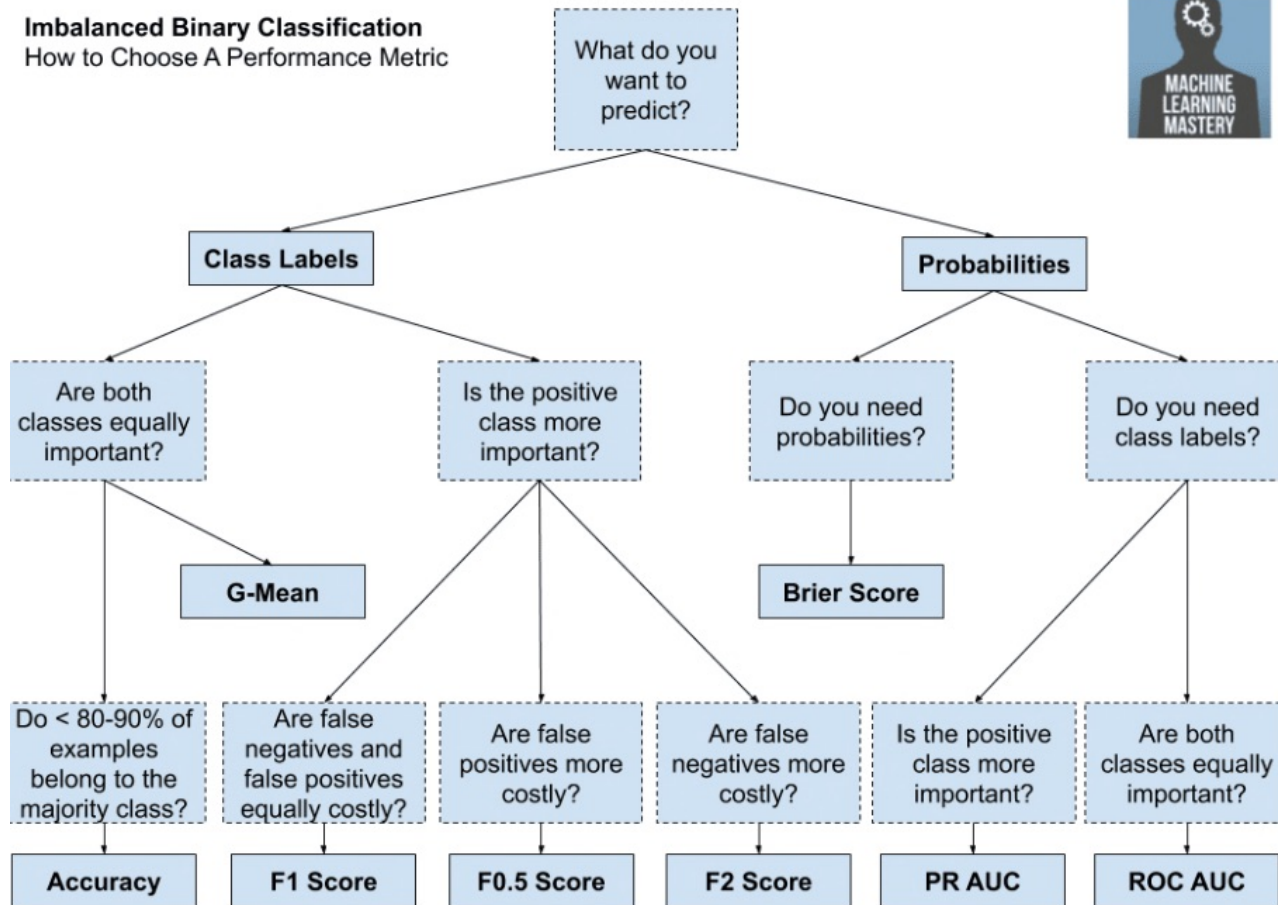
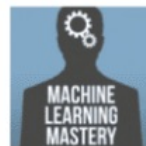
TP

TPR baja

AUC : Area under the (ROC) curve

Imbalanced Binary Classification

How to Choose A Performance Metric



© 2019 MachineLearningMastery.com All Rights Reserved.