

Ciencia de Datos con Python

Una breve introducción

by Walter Casas



Accuracy

Mide el performance del modelo:

$$accuracy = \frac{PrediccionesCorrectas}{TotalPredicciones(Correctas+Incorrectas)}$$

Caso de uso: **Predicción de SPAM**

		Predicciones	
		HAM	SPAM
Real	HAM	970	7
	SPAM	11	12

Correctas	Total	Accuracy
982	1000	0,98



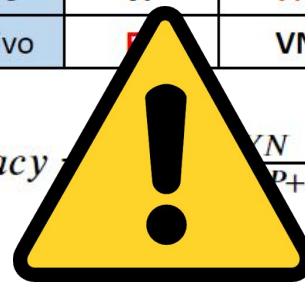
Eventos raros, porcentaje de aparición bajo.

[illegible]

Confusion Matrix

	Predicciones	
Real	Positivo	Negativo
Positivo	VP	FN
Negativo	FP	VN

$$accuracy = \frac{TN}{P+FN}$$



Necesitamos nuevas métricas.





Métricas para Clasificación

Métricas para clasificación

Precision

Real	Predicciones	
	Positivo	Negativo
Positivo	VP	FN
Negativo	FP	VN

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

Recall

Real	Predicciones	
	Positivo	Negativo
Positivo	VP	FN
Negativo	FP	VN

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



Detectando SPAM

Probaremos las nuevas métricas en nuestro ejemplo.

		Predicciones	
		HAM	SPAM
Real	HAM	970	7
	SPAM	11	12

$$precision = \frac{VP}{VP+FP} = \frac{12}{12+7} = 0.63$$

$$recall = \frac{VP}{VP+FN} = \frac{12}{12+11} = 0.52$$



Encontrando el equilibrio

F1-Score

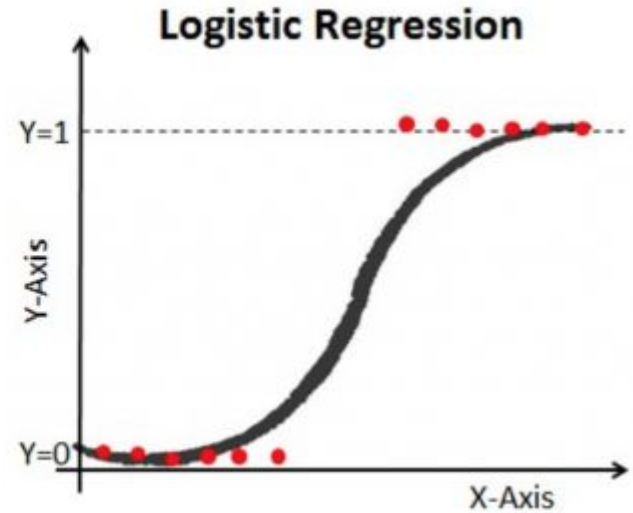
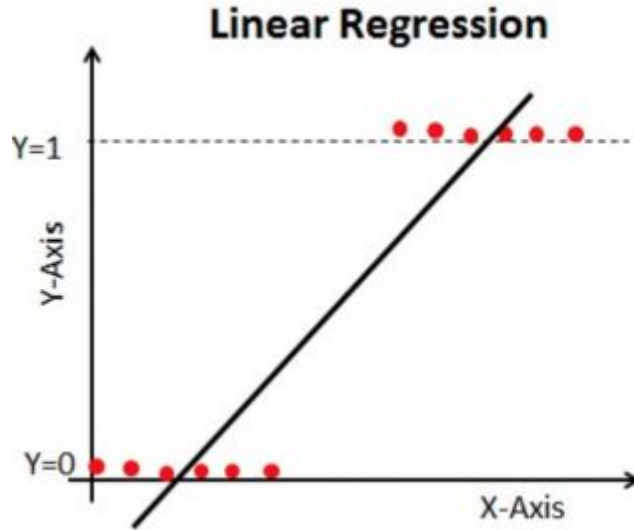
$$F1 = 2 * \frac{precision * recall}{precision + recall}$$





Regresión Logística

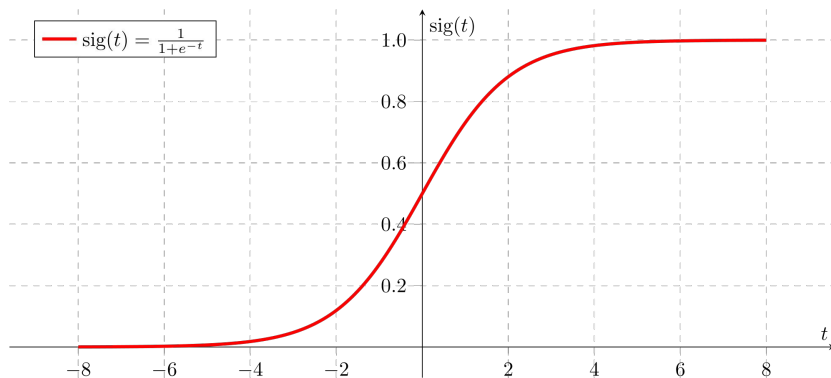
RL para Clasificación Binaria



RL para Clasificación Binaria

La regresión logística nos devuelve probabilidades.

Probabilidad $> 0.5 \rightarrow$ target = 1



$$\text{sig}(t) = \frac{1}{1+e^{-t}}$$



Sensibilidad y Especificidad

True Positive Ratio

TPR

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

False Positive Ratio

FPR

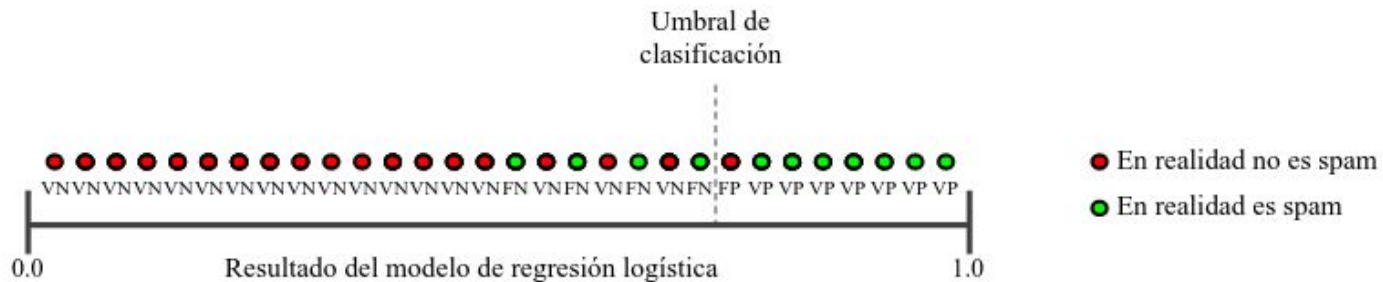
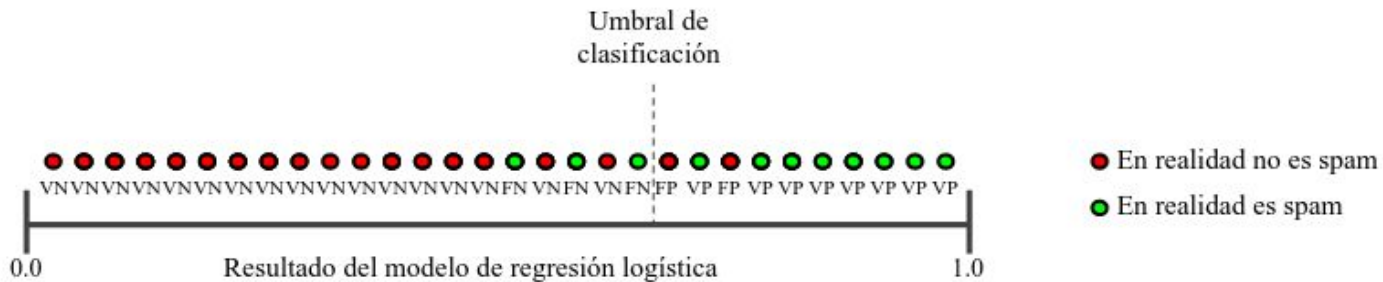
$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{FPR} = 1 - \text{Specificity}$$

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



Podemos variar el punto de corte



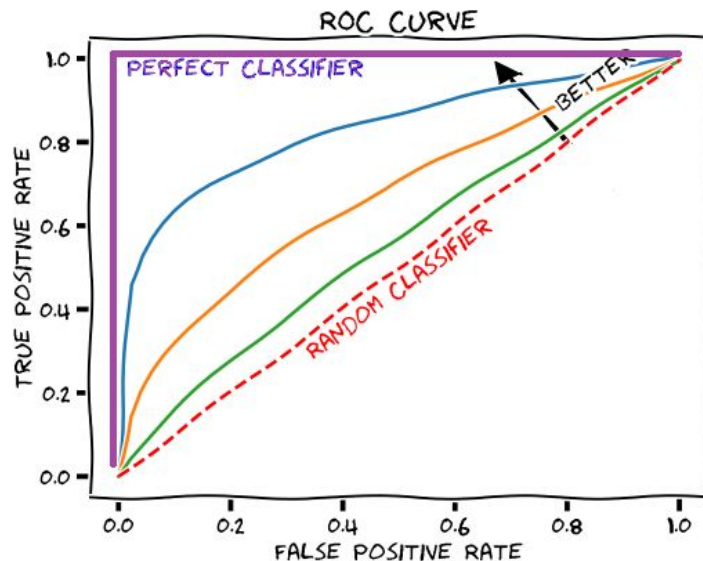
La curva ROC

Si el punto de corte = 0:

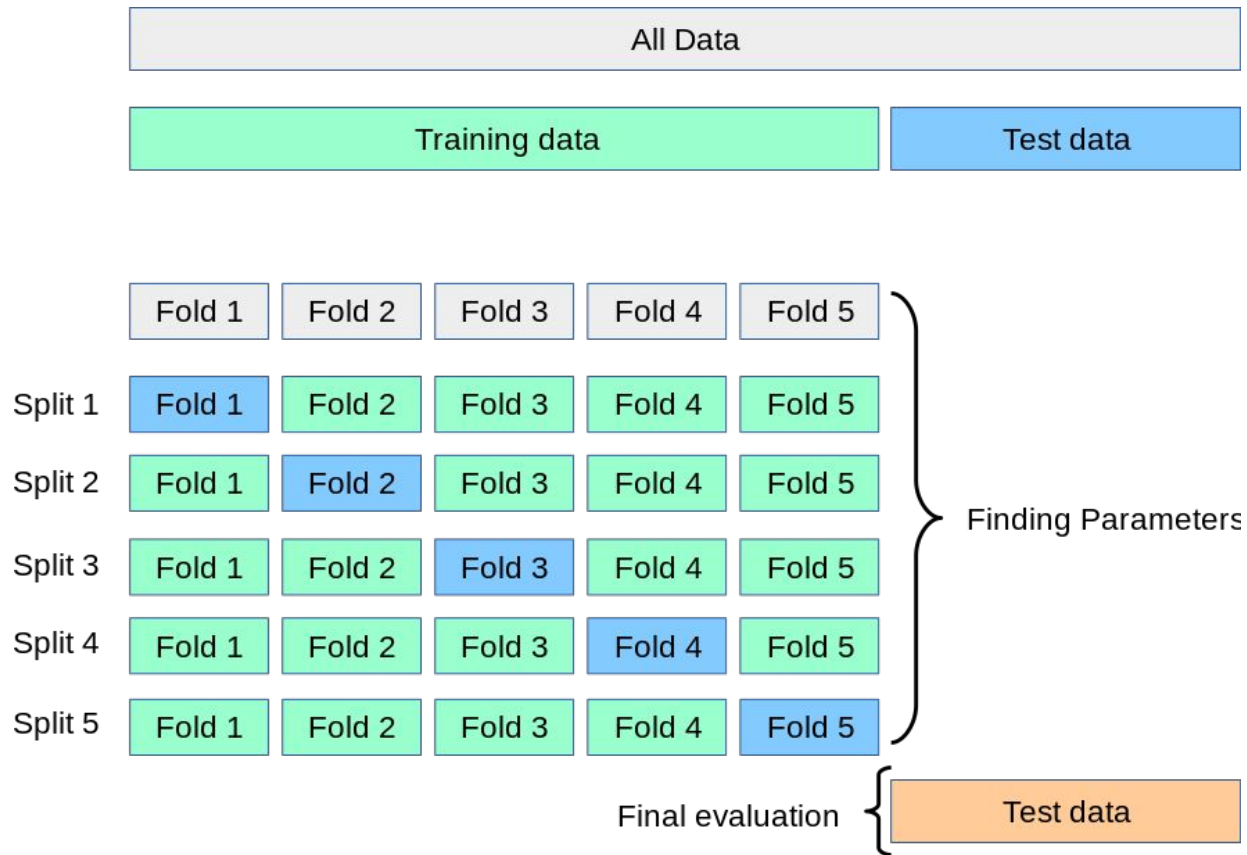
- Todos son 1, además $TPR = FPR = 1$

Si el punto de corte = 1:

- Todos son 0, además $TPR = FPR = 0$



Validación Cruzada





Gracias!

Contáctame:

 [walterpcasas](https://www.linkedin.com/in/walterpcasas)

 wperezc@uni.pe

 [Surf Code](https://www.youtube.com/c/SurfCode)

