

# Métricas

---



# Métricas para la regresión



- MSE, RMSE, R-Squared
- MAE
- (R)MSPE, MAPE
- (R)MSLE

# MSE



Error cuadrático medio:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Se utiliza cuando no hay una preferencia hacia el método de solución o no se conoce otra métrica.

# RMSE



Raíz del error cuadrático medio:

$$RSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

Similar al MSE, la ventaja es que permite analizar el error en la misma escala de las etiquetas. Es más fácil de comprender el error.

A pesar de que son similares, no son directamente intercambiables si se fuera a aplicar gradiente descendiente.

$$\frac{\delta RSE}{\delta \hat{y}_i} = \frac{1}{2\sqrt{MSE}} \frac{\delta MSE}{\delta \hat{y}_i}$$

# R-squared



Métrica que califica la calidad de la regresión entre 0 y 1.

$$R^2 = 1 - \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2} = 1 - \frac{MSE}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2} \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

0 significa que la predicción es tan mala como predecir para todo valor una constante igual a  $\bar{y}$ , 1 significa que el término entre paréntesis es siempre cero por lo que la predicción es perfecta.

# MAE



Promedio de las diferencias absolutas entre las predicciones y las etiquetas:

Fácil de justificar.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Es usado en finanzas. Permite decir cuántas veces es mejor un modelo que el otro. Un error de 10 dólares es dos veces peor que un error de 5 dólares.

Más robusto que el MSE, no es influenciado por los outliers, muestras atípicas.

# Supongamos que tenemos un modelo para dos tiendas:



Para la tienda 1:

Se predice para una muestra el valor de 9, pero se vendió a 10, el MSE es 1.

Para la tienda 2:

Se predice para una muestra el valor de 999, pero se vendió a 1000, el MSE es 1.

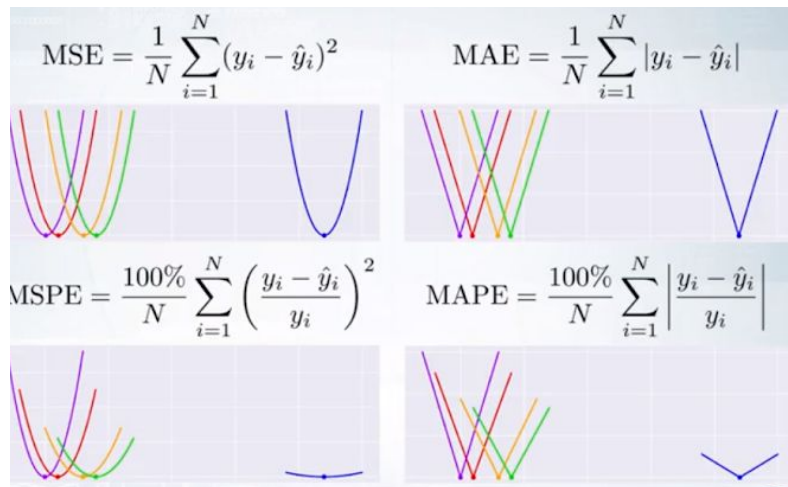
Cuál error es más crítico? El primero!

# Los errores relativos pueden ser más informativos:

$$MSPE = \frac{100\%}{N} \sum_{i=1}^N \left( \frac{y_i - \hat{y}_i}{y_i} \right)^2$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

El costo que se paga por arreglar un error depende del tamaño de la etiqueta. Mientras más pequeña la etiqueta más costoso. MAPE pierde su capacidad para ignorar las muestras atípicas.

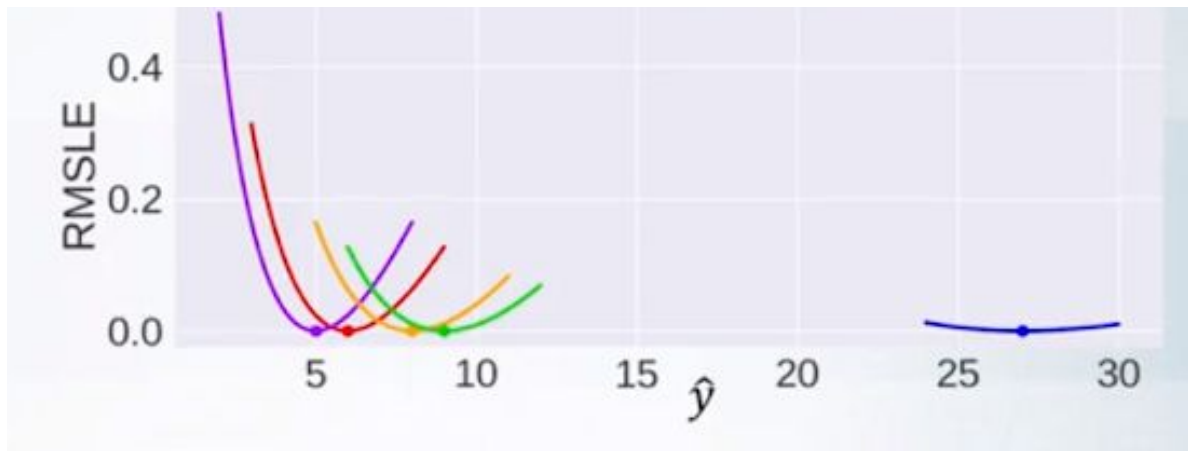




# RMSLE

$$RMSLE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\log(y_i + 1) - \log(\hat{y}_i + 1))^2}$$

Se preocupa más por los valores relativos que por los absolutos. Según esta métrica es mejor siempre predecir más que el valor de la etiqueta, que un valor inferior.



# Métricas para la clasificación



- Accuracy score.
- LogLoss

# Accuracy score



$$Accuracy = \frac{1}{N} \sum_{i=1}^N [\hat{y}_i - y_i]$$

Esta métrica habla de cuan frecuente es la predicción correcta. Un valor de 0.9 dice que de cada 10 predicciones, 9 son buenas.

Tiene un problema. Cuando se tienen clases desbalanceadas:

Datos:

10 Gatos y 90 Perros

Si siempre se predice Perro, el clasificador tiene un accuracy de 0.9!! Error!

# LogLoss



Para un problema binario:

$$LogLoss = -\frac{1}{N} \sum_{i=1}^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)$$

Se preocupa por maximizar las predicciones soft.

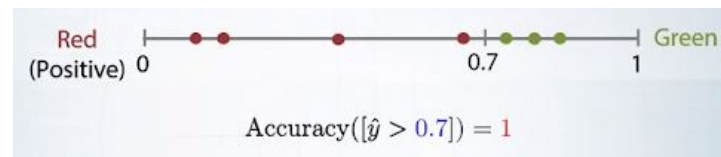
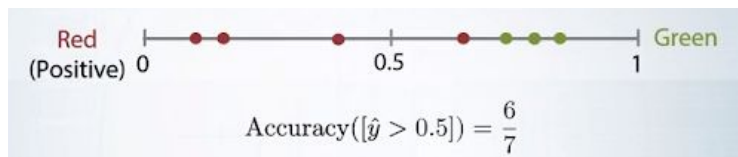
Para problemas multiclase:

$$LogLoss = -\frac{1}{N} \sum_{i=1}^N \sum_{l=1}^L y_{il} \log(\hat{y}_{il})$$

Probabilidad de pertenecer a la clase l:  $\hat{y}_{il}$

# AUC Area Under Curve ROC

Para problemas de clasificación binaria podemos cambiar el valor del umbral de la salida soft para obtener una mejor o peor clasificación.



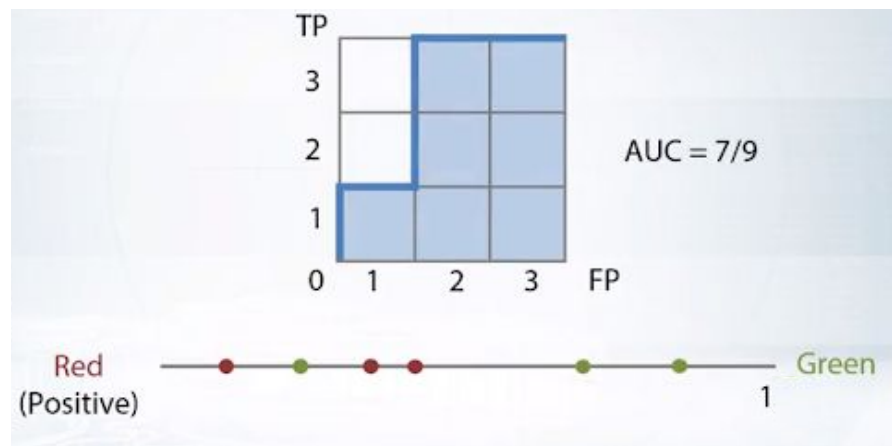
Para construir la curva ROC, se deben primero definir dos conceptos:

**Verdaderos positivos (TP):** Elementos que eran de la clase i y fueron clasificados como clase i.

**Falsos positivos (FP):** Elementos que no eran de la clase i pero fueron clasificados como de la clase i.

# AUC Area Under Curve ROC

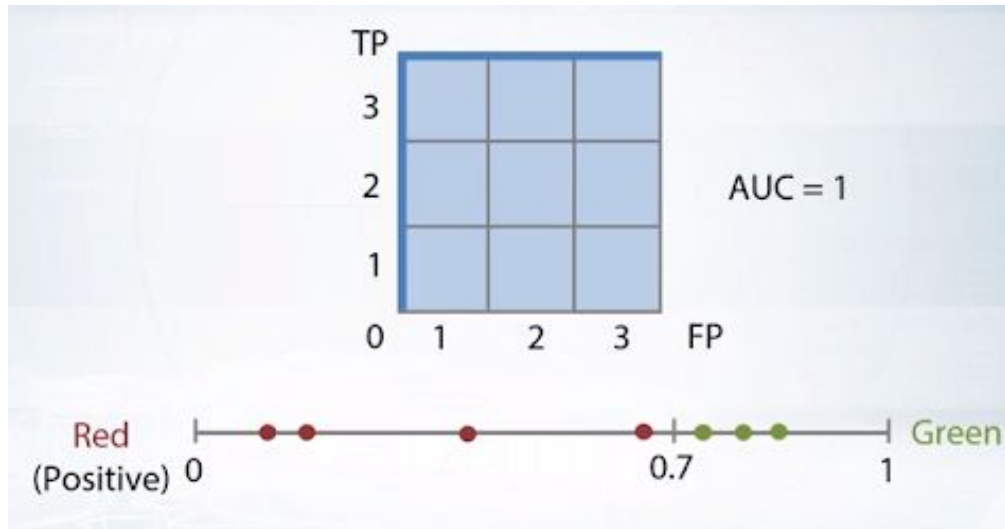
Para un umbral de 0.5, se tiene:



Se organizan las predicciones para la clase positiva y luego las de la clase negativa. Se va de izquierda a derecha contando el número de verdaderos positivos, y de falsos positivos, el primer rojo es un verdadero positivo, TP=1, el verde es un falso positivo (Se predijo como verde), FP= 1, el siguiente rojo es un verdadero positivo, TP=2, y así sucesivamente.

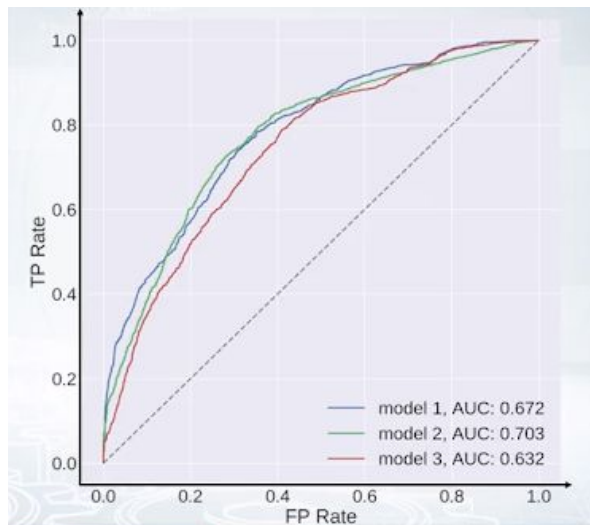
# AUC Area Under Curve ROC

Para un umbral de 0.7, se tiene:



# Para un clasificador real:

La línea punteada representa un clasificador random. AUC=0.5





# Matriz de confusión

		Actual class		
		Cat	Dog	Rabbit
Predicted class	Cat	5	2	0
	Dog	3	3	2
	Rabbit	0	1	11

**True positive (TP):** Se predice la clase positiva y en realidad era de la clase positiva.

**True negative (TN):** Se predice la clase negativa y en realidad era de la clase negativa.

**False positive (FP):** Se predice la clase positiva y en realidad era de la clase negativa.

**False negative (FN):** Se predice la clase negativa y en realidad era de la clase positiva.

La clase positiva en realidad es figurativa, es mi clase de interés. Por ejemplo en un problema multiclase de 3 clases, si quiero analizar que tan bien reconoce mi clasificador la clase perro, la clase perro sería la clase positiva, y las demás las clase negativa.

# Para cada clase se calcula el F-Score



Una medida análoga al error es el F-Score. Este se calcula para cada clase. Es un ponderado de la precisión y la sensibilidad. Donde la precisión es cuantas muestras de las que predije como clase positiva son en realidad de la clase positiva. Y la sensibilidad es de las muestras que eran positivas, cuantas se predijeron como positivas.

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

Precisión

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

Sensibilidad

$$F_1 = 2 \frac{P \times R}{P + R}$$