

Módulo 2: Conceptos generales de Machine Learning

2.5. Métricas de desempeño de modelos

Rafael Zambrano

rafazamb@gmail.com

Evaluación de modelos de regresión

- Error absoluto medio MAE (*Mean Absolute Error*)

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

Real	Predicción
32	35
23	28
85	95

$$MAE = \frac{1}{3} (|32 - 35| + |23 - 28| + |85 - 97|) = \frac{1}{3} (3 + 5 + 10) = 6$$

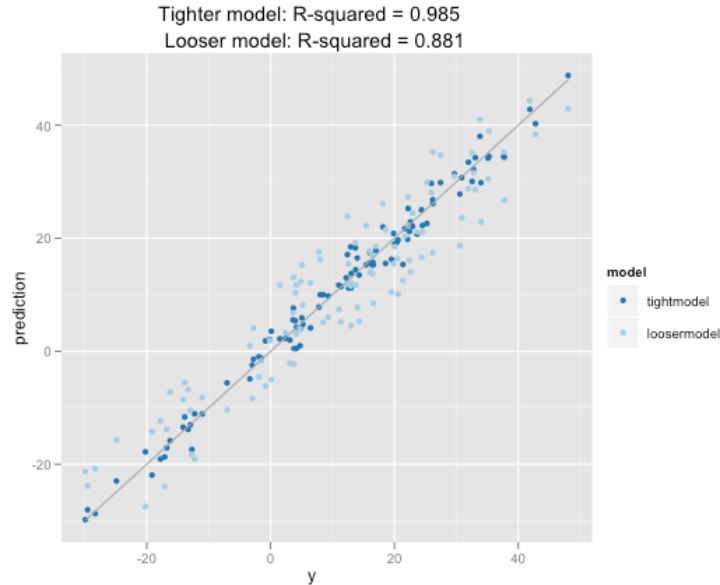
- Error cuadrático medio RMSE (*Root Mean Square Error*): Tiene la ventaja de penalizar errores grandes

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

$$RMSE = \sqrt{\frac{1}{3} (3^2 + 5^2 + 10^2)} = 6.68$$

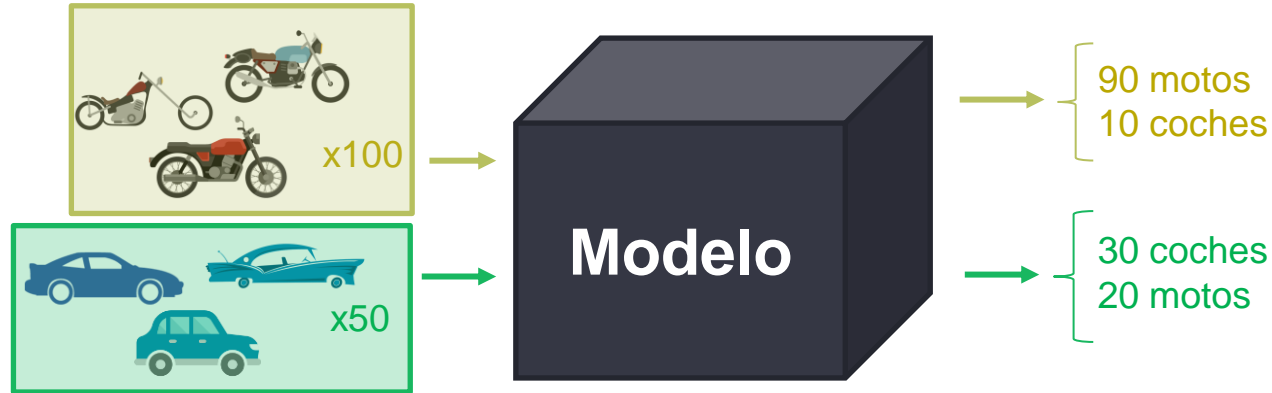
Evaluación de modelos de regresión

- R^2 : cuadrado de la correlación entre los valores reales y las predicciones (valor entre 0 y 1)



Evaluación de modelos de clasificación

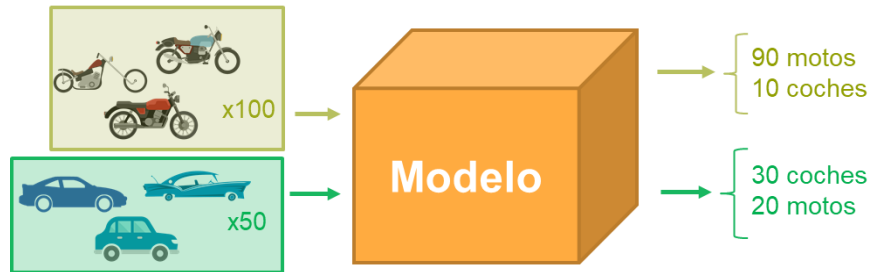
- **Ejemplo:** Clasificador de motos



Evaluación de modelos de clasificación:

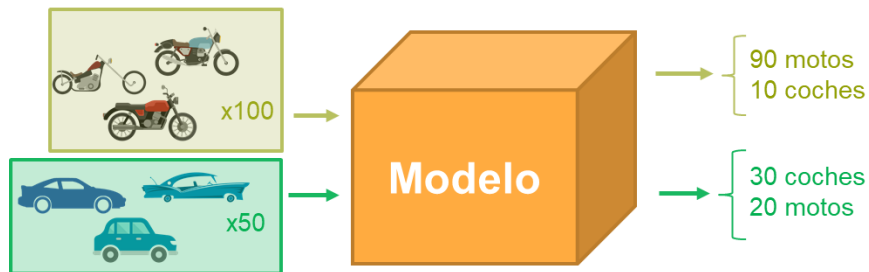
Matriz de confusión

- True Positives (TP)
- True Negatives (TN)
- False Positives (FP)
- False Negatives (FN)



		REALIDAD	
		MOTOS	COCHES
PREDICCIÓN	MOTOS	TP 90	FP 20
	COCHES	FN 10	TN 30

Evaluación de modelos de clasificación



		REALIDAD	
		MOTOS	COCHES
PREDICCIÓN	MOTOS	TP 90	FP 20
	COCHES	FN 10	TN 30

- **Accuracy:** En general, ¿cómo de bueno es el clasificador? $(TP+TN)/Total = (90+30)/150 = 80\%$

- **Precision:** Cuando predice “Sí”, ¿cuántas veces acierta? $TP/(TP+FP) = 90/(90+20) = 82\%$

- **Recall:** En los casos reales de “Sí”, ¿cuánto predice correctamente? $TP/(TP+FN) = 90/(90+10) = 90\%$

- **Specificity:** En los casos reales de “NO”, ¿cuánto predice correctamente? $TN/(TN+FP) = 30/(30+20) = 60\%$

- **F1 Score:** Combina precision y recall en una sola métrica $\in [0,1]$: $(2*Precision*Recall)/(Precision+Recall) = 0.81$

- **Ratio de falsos positivos:** $1 - Specificity = FP/(TN+FP) = 1-0.6 = 40\%$

- **Mejora:** ¿Cómo mejora el modelo a una decisión aleatoria?

$$\text{Prior} = P(\text{moto}) = 100/150 = 66\%$$

$$\Rightarrow \text{Mejora} = \text{Precision}/\text{Prior} = 82/66 = 1.24$$

Evaluación de modelos de clasificación

Ejemplo: Detección de fraude en tarjetas. De 1000 clientes, solo 10 cometieron fraude. Se entrena un modelo y se prueba con el conjunto de test, obteniendo la siguiente matriz de confusión

PREDICCIÓN		FRAUDE	NO FRAUDE
	FRAUDE	0	0
	NO FRAUDE	10	990

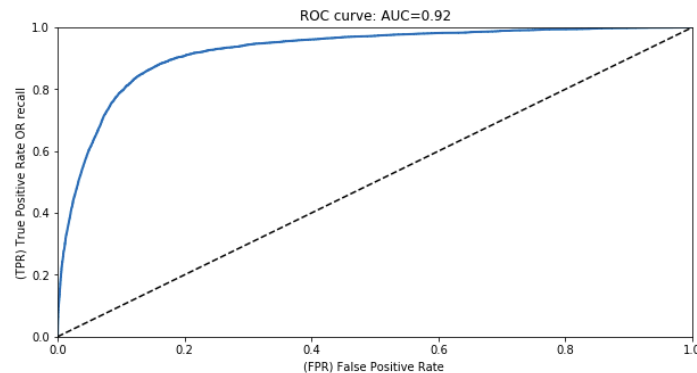
- **Accuracy:** En general, ¿cómo de bueno es el clasificador? $(TP+TN)/Total = (0+990)/1000 = 99\%$
- **Precision:** Cuando predice “Sí”, ¿cuántas veces acierta? $TP/(TP+FP) = 0/(0+0) = 0\%$
- **Recall:** En los casos reales de Sí, ¿cuánto predice correctamente? $TP/(TP+FN) = 0/(0+10) = 0\%$
- **Specificity:** En los casos reales de NO, ¿cuánto predice correctamente? $TN/(TN+FP) = 990/(990+0) = 100\%$
- **F1 Score:** Combina precision y recall en una sola métrica $\in [0,1]$: $(2*Precision*Recall)/(Precision+Recall) = 0$
- **Ratio de falsos positivos:** $1 - Specificity = FP/(TN+FP) = 1-0 = 100\%$
- **Mejora:** ¿Cómo mejora el modelo a una decisión aleatoria?

$$\text{Prior} = P(\text{fraude}) = 10/1000 = 1\%$$

$$\Rightarrow \text{Mejora} = \text{Precision}/\text{Prior} = 0/1 = 0$$

Evaluación de modelos de clasificación: Curva ROC

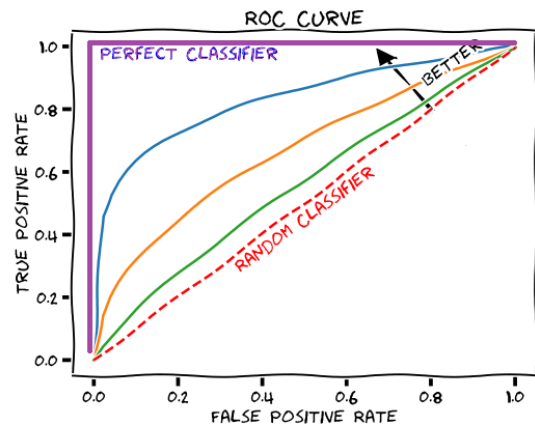
- En los modelos de clasificación binarios, existe un compromiso entre el error de falsos positivos y el de falsos negativos, pudiendo aumentar uno para disminuir el otro, y viceversa.
- Ejemplos:
 - Quiero que mi modelo detecte todos los fraudes de tarjetas: habrá muchos falsos positivos (baja precisión) y pocos falsos negativos (mayor recall)
 - Quiero que mi modelo detecte solo los casos reales de fraude de tarjetas: habrá muchos falsos negativos (bajo recall) y pocos falsos positivos (alta precisión)
- La **curva ROC** relaciona el recall con el ratio de falsos positivos



Evaluación de modelos de clasificación: Curva ROC

- En las curvas ROC, nos interesa que la curva se acerque lo máximo posible a la esquina superior izquierda de la gráfica, de manera que el hecho de aumentar el recall no haga que nuestro modelo introduzca más falsos positivos.
- En este caso también podemos calcular el ROC AUC (área bajo la curva), que también nos sirve como métrica para resumir la curva y poder comparar modelos.

- $AUC = 1$: Clasificador perfecto
- $AUC = 0.5$: Clasificador aleatorio



¡Gracias!

Contacto: Rafael Zambrano

rafazamb@gmail.com