Métricas para Regresión Lineal en Big Data

Experto en Big Data

29 de noviembre de 2025

1. Introducción

En modelos de regresión lineal, es fundamental evaluar el rendimiento mediante métricas cuantitativas. A continuación, se detallan las principales métricas, sus fórmulas, ventajas y desventajas.

2. Métricas Principales

3. Resumen de Selección

Métrica	**Usar Cuando...**	**Evitar Cuando...**
MSE	Penalizar errores grandes.	Hay muchos *outliers*.
RMSE	Error en unidades originales.	*Outliers* problemáticos.
MAE	Robustez frente a *outliers*.	Necesitas diferenciabilidad.
R^2	Medir varianza explicada.	Hay sobreajuste.
R^2 Ajustado	Comparar modelos con más variables.	Modelos simples.
MAPE	Errores en porcentaje.	Valores cercanos a cero.

Cuadro 1: Guía rápida para selección de métricas.

$R^2 ajustado$ Supongamos que tenemos una regresión con los siguientes datos iniciales:

Cuadro 2: Placeholder Caption

Métrica	Valor
Número de Observaciones (n)	30
Variables Predictoras (k)	2
R2 del Modelo	0.70

1. Modelo Base (2 Variables).Primero, calculemos el $\mathbf{R^2}$ ajustado para el modelo base

- Fórmula: $\bar{R}^2 = 1 - \left[(1 - R^2) \cdot \frac{n-1}{n-k-1}\right]$
- Sustitución: $\bar{R}^2 = 1 - \left[(1 - 0{,}70) \cdot \frac{30-1}{30-2-1}\right]$
- Cálculo: $\bar{R}^2 = 1 - \left[(0{,}30) \cdot \frac{29}{27}\right]$
- Resultado: $\bar{R}^2 \approx 1 - [0{,}30 \cdot 1{,}074] \approx 1 - 0{,}322 = \mathbf{0{,}678}$

$$\text{Modelo Base: } \bar{R}^2 \approx 0{,}678$$

2. Adición de una Variable Irrelevante
   Ahora, agregamos una tercera variable predictora ($k = 3$). Esta nueva variable no tiene ningún poder explicativo real sobre la variable dependiente.Como resultado de añadirla, el R^2 estándar siempre aumenta, aunque sea mínimamente:

Cuadro 3: Placeholder Caption

Métrica	Valor
Número de Observaciones (n)	30
Variables Predictoras (k)	3
R2 del Modelo	0.705

- Fórmula: $\bar{R}^2 = 1 - \left[(1 - 0{,}705) \cdot \frac{30-1}{30-3-1}\right]$
- Sustitución: $\bar{R}^2 = 1 - \left[(0{,}295) \cdot \frac{29}{26}\right]$
- Cálculo: $\bar{R}^2 = 1 - \left[(0{,}30) \cdot \frac{29}{27}\right]$
- Resultado: $\bar{R}^2 \approx 1 - \left[(0{,}30) \cdot \frac{29}{27}\right] \approx 1 - 0{,}329 = \mathbf{0{,}671}$

$$\text{Modelo Base: } \bar{R}^2 \approx 0{,}671$$

Métrica	Modelo Base (k=2)	Modelo Expandido (k=3)
R2 Estándar	0.700	0.705 (Aumenta)
R2 Ajustado ($\bar{R}^2$)	0.678	0.671 (Disminuye)

Cuadro 4: Placeholder Caption

Conclusión:

- El $\mathbf{R^2}$ nos engañaría, sugiriendo que el modelo de 3 variables es "mejor"(0.705 ¿0.700)
- .El $\mathbf{R^2}$ ajustado ($\bar{\mathbf{R}}^2$) detecta que el beneficio del R^2 no justifica la penalización por la variable adicional (k aumentó). Por lo tanto, el $\bar{R}^2$ disminuye (0.671 ¡0.678), indicando que el modelo de 2 variables es más parsimonioso y preferible

Esto demuestra por qué el R^2 ajustado es una métrica crucial para la selección de modelos, ya que prioriza la simplicidad (parsimonia) junto con la bondad de ajuste.