

Error en Modelos de Regresión

Error en Modelos de Regresion

1. Error Total

El error total mide la diferencia entre los valores reales (y) y su promedio ($y\text{-barra}$). Es una medida de la variabilidad total en los datos que el modelo intenta explicar.

Formula:

$$\text{Error Total} = \text{suma}(y_i - y\text{-barra})^2$$

2. Error de Residuos (Error No Explicado)

Los residuos son las diferencias entre los valores reales (y_i) y los valores predichos (y_i sombrero).

Formula:

$$\text{Residuos} = y_i - y_i \text{ sombrero}$$

Este error refleja la parte de la variabilidad que el modelo no puede explicar.

3. Error Explicado

Es la parte de la variabilidad total que el modelo de regresion logra capturar. Representa como las predicciones se ajustan al promedio de los valores reales.

Formula:

$$\text{Error Explicado} = \text{suma}(y_i \text{ sombrero} - y\text{-barra})^2$$

Error en Modelos de Regresión

4. Medidas de Error

Para cuantificar los errores, se utilizan metricas:

- Error Cuadratico Medio (MSE): $MSE = (1/n) \sum (y_i - \hat{y}_i)^2$
- Raiz del Error Cuadratico Medio (RMSE): $RMSE = \sqrt{MSE}$
- Error Absoluto Medio (MAE): $MAE = (1/n) \sum |y_i - \hat{y}_i|$
- Coeficiente de Determinacion (R2): $R^2 = 1 - (\text{Suma de errores cuadraticos} / \text{Suma de errores totales})$

5. Tipos de Error por Causa

- Error de Sesgo: Ocurre cuando el modelo es demasiado simple para capturar la complejidad de los datos.
- Error de Varianza: Sucede cuando el modelo es demasiado complejo y se ajusta excesivamente a los datos de entrenamiento.
- Error Irreducible: Es el error inherente a los datos debido a ruido u otros factores impredecibles.

6. Estrategias para Manejar Errores

- Seleccionar el modelo adecuado.
- Validacion cruzada.
- Regularizacion (Lasso, Ridge).
- Aumentar los datos.
- Eliminar ruido en el preprocesamiento.