

Evaluación de Regresión



La evaluación en modelos de regresión se centra en medir qué tan bien el modelo puede predecir valores numéricos continuos. Las métricas de evaluación comunes utilizadas en este contexto son:

1. Error Cuadrático Medio (Mean Squared Error, MSE)



El MSE es una de las métricas de evaluación más utilizadas en modelos de regresión. Se calcula como la media de las diferencias al cuadrado entre las predicciones del modelo y los valores reales. La fórmula es:

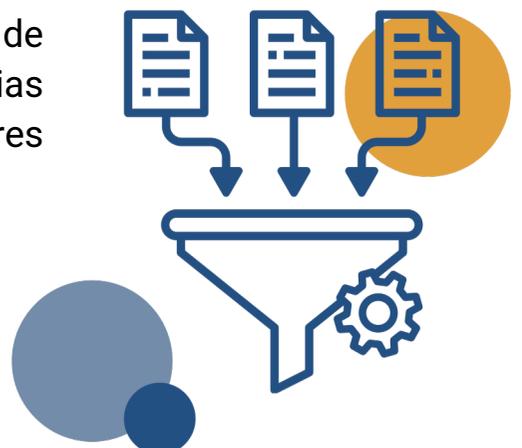
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

- Donde y_i son los valores reales, \hat{y}_i son las predicciones del modelo y n es el número total de muestras.

2. Error Absoluto Medio (Mean Absolute Error, MAE)

El MAE es otra métrica común utilizada en modelos de regresión. Se calcula como la media de las diferencias absolutas entre las predicciones del modelo y los valores reales. La fórmula es:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$



3. Raíz del Error Cuadrático Medio (Root Mean Squared Error, RMSE)



El RMSE es simplemente la raíz cuadrada del MSE. Proporciona una medida de la dispersión de los errores en las mismas unidades que la variable objetivo. La fórmula es:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

4. Coeficiente de Determinación (R² o R-cuadrado)

El coeficiente de determinación es una medida estadística que indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Es una medida de cuán bien el modelo se ajusta a los datos.

El valor de R² puede variar de 0 a 1, donde 1 indica un ajuste perfecto del modelo a los datos y 0 indica que el modelo no explica nada de la variabilidad de los datos.

Un valor negativo indica que el modelo es peor que simplemente usar la media como predicción.

La fórmula para el coeficiente de determinación es:



$$R^2 = \frac{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

Donde \bar{y} es la media de los valores reales y_i .

Estas métricas son útiles para evaluar diferentes aspectos del rendimiento de un modelo de regresión. El MSE, MAE y RMSE proporcionan información sobre la precisión y magnitud de los errores de predicción, mientras que el coeficiente de determinación R² proporciona información sobre la calidad del ajuste del modelo a los datos.

Es importante seleccionar las métricas de evaluación adecuadas según el contexto del problema y las preferencias del usuario.



Ejemplo

- El siguiente código utiliza un modelo de regresión lineal para predecir valores numéricos y luego calcula las métricas de error MSE, MAE, RMSE y R^2 utilizando las funciones proporcionadas por scikit-learn.
- Recuerda que puedes cambiar el modelo por cualquier otro modelo de regresión que desees evaluar, simplemente cambiando la inicialización del modelo y ajustándolo a tus datos.

```
from sklearn.datasets import make_regression
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error,
r2_score
import numpy as np

# Generar datos sintéticos para regresión
X, y = make_regression(n_samples=100, n_features=1, noise=10,
random_state=42)

# Dividir los datos en conjunto de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Inicializar y entrenar un modelo de regresión lineal
model = LinearRegression()
model.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred = model.predict(X_test)

# Calcular las métricas de error
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

# Imprimir las métricas de error
print("Mean Squared Error (MSE):", mse)
print("Mean Absolute Error (MAE):", mae)
print("Root Mean Squared Error (RMSE):", rmse)
print("R-squared (R²):", r2)
```