

Práctica: cálculo del Coeficiente de Determinación R^2

El objetivo de esta práctica es que implementes el cálculo del coeficiente de determinación R^2 desde cero. Este coeficiente se utiliza para evaluar la precisión de un modelo de regresión lineal. Al finalizar, deberás probar tu implementación con un conjunto de datos propuestos y entender cómo se calcula y qué representa el valor de R^2 .

Introducción

El **coeficiente de determinación** R^2 es una medida que indica qué tan bien se ajusta un modelo de regresión lineal a los datos. En términos simples, indica la proporción de la variabilidad total de la variable dependiente (valores reales) que es explicada por el modelo (valores predichos).

El valor de R^2 oscila entre 0 y 1:

- Un valor de $R^2 = 1$ indica que el modelo explica perfectamente la variación en los datos.
- Un valor de $R^2 = 0$ indica que el modelo no explica ninguna de las variaciones.
- Un valor intermedio refleja qué porcentaje de la variación es explicada por el modelo, cuanto más cercano a 1, mejor es el ajuste del modelo.

El R^2 se calcula con la siguiente fórmula:

$$R^2 = 1 - \frac{\sum (y_{\text{real}} - y_{\text{predicho}})^2}{\sum (y_{\text{real}} - \overline{y_{\text{real}}})^2}$$

En esta práctica, implementarás esta fórmula sin utilizar funciones predefinidas de librerías como scikit-learn.

Instrucciones para la implementación

1. Crea una función con el nombre que quieras, que tome como entrada 2 arrays de NumPy: “y_real” (los valores reales) y “y_predicho” (los valores predichos por un modelo).
2. Dentro de la función transforma la fórmula a una implementación de Python que devuelva el R^2 como un valor decimal.
3. Pruebas. Utiliza este conjunto de datos para probar tu implementación:

```
y_real = np.array([3, -0.5, 2, 7, 4.2])
y_predicho = np.array([2.5, 0.0, 2, 8, 4.1])
```

El coeficiente de determinación (R^2) es: 0.95

4. Compara el valor obtenido en tu implementación con la función `r2_score` de la librería `sklearn` y verifica si coinciden.