



UNIVERSITAS
Miguel Hernández

Tema 1. Regresión Lineal. Práctica.

José L. Sainz-Pardo Auñón

TÉCNICAS ESTADÍSTICAS PARA EL APRENDIZAJE II

Máster Universitario en Estadística Computacional
y Ciencia de Datos para la Toma de Decisiones.

Objetivo de la Práctica

- Utilizar diferentes técnicas de regresión para obtener el mejor modelo predictivo.
- Evaluar y comparar los modelos de regresión usando métricas estándar.
- Basar el análisis en bases de datos abiertas y reales para un enfoque práctico.

Base de Datos Utilizada

- Boston Housing Dataset (Disponible en 'scikit-learn'):
 - ▶ Predecir el precio de las viviendas en función de características socioeconómicas y geográficas.
- Bases de datos real que se usa frecuentemente para aplicar métodos de regresión.

1. Preprocesamiento de Datos

- Cargar los datos y manejar valores nulos, si es necesario.
- Realizar una exploración inicial:
 - ▶ Descripción estadística de las variables.
 - ▶ Gráficos de dispersión para evaluar relaciones entre las variables.

2. Selección de Variables

- Utilizar la matriz de correlación para identificar multicolinealidad.
- Aplicar el Factor de Inflación de la Varianza (VIF):
 - ▶ Eliminar variables altamente correlacionadas que aumenten la multicolinealidad.

3. Regresión Lineal Básica

- Ajustar un modelo de regresión lineal a los datos (utiliza el 80% de los datos como muestra de entrenamiento y el resto como conjunto de prueba).
- Evaluar el modelo usando:
 - ▶ Coeficientes de la regresión.
 - ▶ R^2 : Explicación de la varianza.
- Visualización:
 - ▶ Gráfico de los valores reales vs valores predichos.

5. Evaluación y Validación de Modelos

- Comparar los diferentes modelos basándose en:
 - ▶ R^2 ajustado.
- Analizar los residuos gráficamente y validar sus supuestos.

7. Mejora del Modelo

- Experimentar con diferentes transformaciones de variables como transformación logarítmica, cuadrática o raíz cuadrada de algunas variables.
- Probar nuevas interacciones entre variables que puedan mejorar el modelo.
- Añadir polinomios de mayor grado para capturar la no linealidad.

8. Más bases de datos

Puedes seguir realizando prácticas con las siguientes bases de datos:

- Auto MPG Dataset
(<https://archive.ics.uci.edu/dataset/9/auto+mpg>)
- Wine Quality Dataset
(<https://archive.ics.uci.edu/dataset/186/wine+quality>)
- Energy Efficiency Dataset (<https://archive.ics.uci.edu/dataset/242/energy+efficiency>)
- Concrete Compressive Strength (<https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength>)