

Modelización mediante Regresión Lineal con Anaconda

Enunciado

Una empresa del sector tecnológico quiere estimar el salario mensual de sus empleados en función de distintas variables relacionadas con su perfil profesional. Para ello, se dispone de un conjunto de datos históricos que será utilizado para construir un modelo de regresión lineal.

Herramienta

Anaconda (Jupyter Notebook).

Conjunto de datos

Se proporciona un fichero llamado `empleados.csv` que contiene las siguientes variables:

edad → Edad del empleado (años)

experiencia → Años de experiencia laboral

formacion → Nivel formativo (1 = FP, 2 = Grado, 3 = Máster)

horas_extra → Número medio de horas extra mensuales

salario → Salario mensual en euros (variable dependiente)

Se pide

- Estudiar el proceso completo de creación de un modelo de regresión lineal.
- Aplicar técnicas básicas de análisis exploratorio de datos.
- Construir y evaluar un modelo predictivo.
- Interpretar los coeficientes obtenidos.

Tareas a realizar

1. Carga y exploración de datos

- Importa las librerías necesarias (pandas, numpy, matplotlib, sklearn).
- Carga el fichero empleados.csv.
- Muestra las primeras filas del conjunto de datos.
- Comprueba si existen valores nulos o inconsistencias.

2. Análisis exploratorio

- Calcula estadísticas descriptivas básicas.
- Representa gráficamente la relación entre las variables independientes y el salario.
- Comenta brevemente qué variables parecen tener mayor relación con el salario.

3. Preparación de los datos

- Separa las variables independientes (X) y la variable dependiente (y).
- Divide el conjunto de datos en conjunto de entrenamiento y de prueba (por ejemplo, 80 % / 20 %).
- Justifica brevemente por qué es necesaria esta división.

4. Construcción del modelo

- Crea un modelo de regresión lineal.
- Entrena el modelo con los datos de entrenamiento.
- Muestra la ecuación del modelo obtenido.

5. Evaluación del modelo

- Realiza predicciones sobre el conjunto de prueba.
- Calcula al menos una métrica de evaluación (R^2 , MAE o MSE).
- Interpreta el resultado obtenido.

6. Interpretación

- Explica qué significado tienen los coeficientes del modelo.
- Indica qué variable tiene mayor impacto en el salario según el modelo.
- Comenta una posible limitación del modelo construido.