01 PAO25-25 - PYTHON, Data Types

Luis Pilaguano

Abre este Jupyter en Google Colab

Regresión Lineal: Coste de un incidente de seguridad

En este ejercicio se explican los fundamentos básicos de la regresión lineal aplicada a un caso de uso sencillo relacionado con la Ciberseguridad.

Enunciado del ejercicio

El ejercicio consiste en predecir el coste de un incidente de seguridad en base al número de equipos que se han visto afectados. El conjunto de datos es generado de manera aleatoria.

0. Imports

In [1]:
```
# Instalacion de librerías externas
!pip install pandas
!pip install numpy
!pip install matplotlib
!pip install scikit-learn
```

```
Requirement already satisfied: pandas in c:\users\user\anaconda3\envs\machinelear
ningpracticas\lib\site-packages (2.2.3)
Requirement already satisfied: numpy>=1.26.0 in c:\users\user\anaconda3\envs\mach
inelearningpracticas\lib\site-packages (from pandas) (2.3.1)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\user\anaconda3
\envs\machinelearningpracticas\lib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\user\anaconda3\envs\machi
nelearningpracticas\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\user\anaconda3\envs\mac
hinelearningpracticas\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\user\anaconda3\envs\machinele
arningpracticas\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
Requirement already satisfied: numpy in c:\users\user\anaconda3\envs\machinelearn
ingpracticas\lib\site-packages (2.3.1)
Requirement already satisfied: matplotlib in c:\users\user\anaconda3\envs\machine
learningpracticas\lib\site-packages (3.10.0)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\user\anaconda3\envs\m
achinelearningpracticas\lib\site-packages (from matplotlib) (1.3.1)
Requirement already satisfied: cycler>=0.10 in c:\users\user\anaconda3\envs\machi
nelearningpracticas\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\user\anaconda3\envs
\machinelearningpracticas\lib\site-packages (from matplotlib) (4.55.3)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\user\anaconda3\envs
\machinelearningpracticas\lib\site-packages (from matplotlib) (1.4.8)
Requirement already satisfied: numpy>=1.23 in c:\users\user\anaconda3\envs\machin
elearningpracticas\lib\site-packages (from matplotlib) (2.3.1)
Requirement already satisfied: packaging>=20.0 in c:\users\user\anaconda3\envs\ma
chinelearningpracticas\lib\site-packages (from matplotlib) (24.2)
Requirement already satisfied: pillow>=8 in c:\users\user\anaconda3\envs\machinel
earningpracticas\lib\site-packages (from matplotlib) (11.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\user\anaconda3\envs\m
achinelearningpracticas\lib\site-packages (from matplotlib) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\user\anaconda3\en
vs\machinelearningpracticas\lib\site-packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in c:\users\user\anaconda3\envs\machinele
arningpracticas\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.17.
0)
Requirement already satisfied: scikit-learn in c:\users\user\anaconda3\envs\machi
nelearningpracticas\lib\site-packages (1.7.0)
Requirement already satisfied: numpy>=1.22.0 in c:\users\user\anaconda3\envs\mach
inelearningpracticas\lib\site-packages (from scikit-learn) (2.3.1)
Requirement already satisfied: scipy>=1.8.0 in c:\users\user\anaconda3\envs\machi
nelearningpracticas\lib\site-packages (from scikit-learn) (1.16.0)
Requirement already satisfied: joblib>=1.2.0 in c:\users\user\anaconda3\envs\mach
inelearningpracticas\lib\site-packages (from scikit-learn) (1.5.1)
Requirement already satisfied: threadpoolctl>=3.1.0 in c:\users\user\anaconda3\en
vs\machinelearningpracticas\lib\site-packages (from scikit-learn) (3.6.0)
```

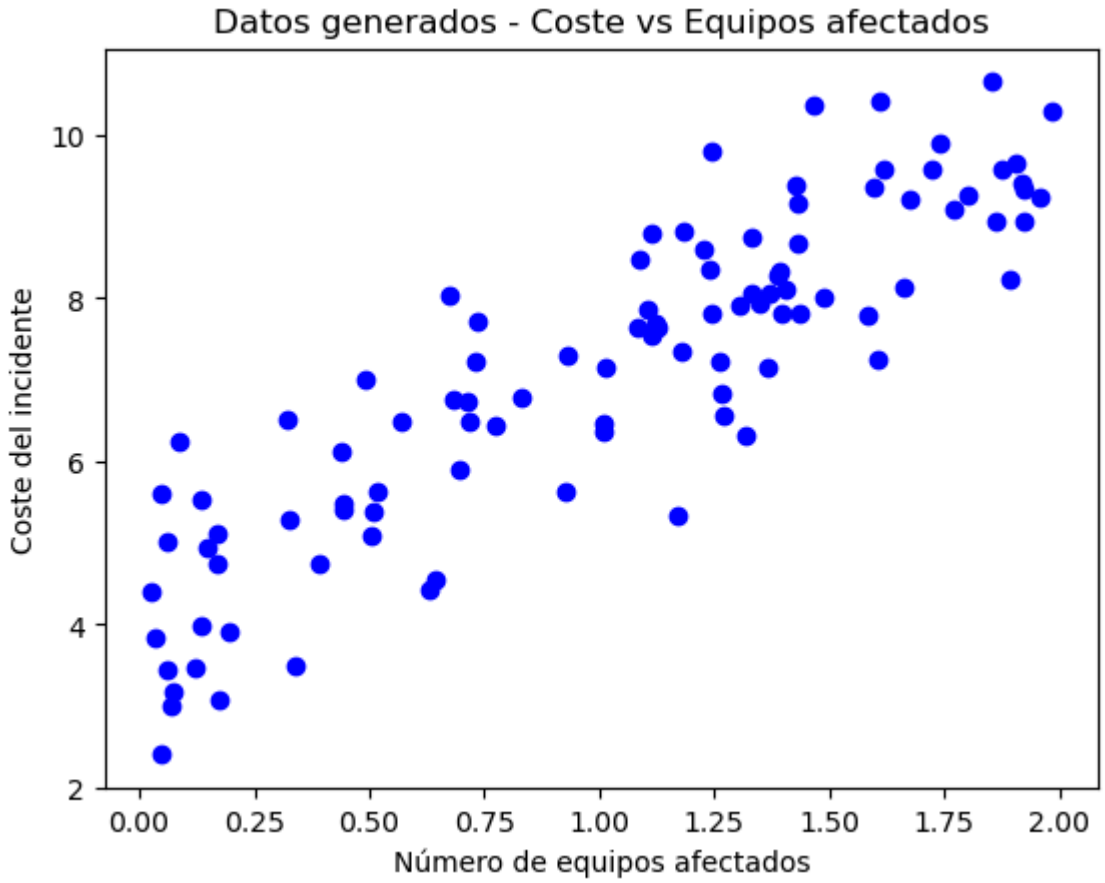
1. Generación del conjunto de datos

In [2]:
```python
import numpy as np

X = 2 * np.random.rand(100, 1)
y = 4 + 3 * X + np.random.randn(100, 1)

print("La longitud del conjunto de datos es:", len(X))
```

```
La longitud del conjunto de datos es: 100
```

2. Visualización del conjunto de datos

In [3]:
```
# tu respuesta aqui
import matplotlib.pyplot as plt

plt.scatter(X, y, color='blue')
plt.xlabel("Número de equipos afectados")
plt.ylabel("Coste del incidente")
plt.title("Datos generados - Coste vs Equipos afectados")
plt.show()
```



3. Modificación del conjunto de datos

In [4]:
```
y = y.ravel()  # convierte y a vector 1D
```

4. Construcción del modelo

In [5]:
```
#Tu respuesta aqui
from sklearn.linear_model import LinearRegression
# Crear una instancia del modelo de regresión lineal
modelo = LinearRegression()
# Entrenar (ajustar) el modelo usando los datos X (número de equipos afectados)
modelo.fit(X, y)
# Mostrar el intercepto (b0) de la recta ajustada
print(f"Intercepto (b0): {modelo.intercept_}")
# Mostrar el coeficiente (pendiente b1) que indica cuánto cambia el coste por ca
print(f"Coeficiente (b1): {modelo.coef_[0]}")
```

```
Intercepto (b0): 4.1267400224438475
Coeficiente (b1): 2.9316270924630676
```

5. Predicción de nuevos ejemplos

In [6]:
```python
#Tu respuesta aqui
# Nuevo dato: 1.5 equipos afectados
X_nuevo = np.array([[1.5]])
prediccion = modelo.predict(X_nuevo)
print(f"Predicción del coste para {X_nuevo[0][0]} equipos afectados: {prediccion
```

```
Predicción del coste para 1.5 equipos afectados: 8.52
```

Git Hub

In [ ]:
```python

```