

# Regresion Lineal

## Nombre: David Egas

Desarrollo de un ejemplo simple de regresión lineal en python.

El dataset que se va a usar es el Salary\_Data, contiene información referente a los años de experiencia y el salario que se gana, se procede a aplicar un modelo de regresión lineal para calcular el salario que se debe ganar según los años de experiencia, comparar los resultados obtenidos con los reales y ver el error.

## Importaciones necesarias

```
In [29]: # Importar Las Librerías
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

```
In [30]: # Importar el data set
dataset = pd.read_csv('Salary_Data.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values

print(dataset[:5])
```

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

## Dividir los datos en entrenamiento y testeo

```
In [31]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 1/3, random_stat
```

## Crear el modelo de regresión, usando sklearn

```
In [32]: from sklearn.linear_model import LinearRegression
regression = LinearRegression()
regression.fit(X_train, y_train)
```

```
Out[32]: LinearRegression()
```

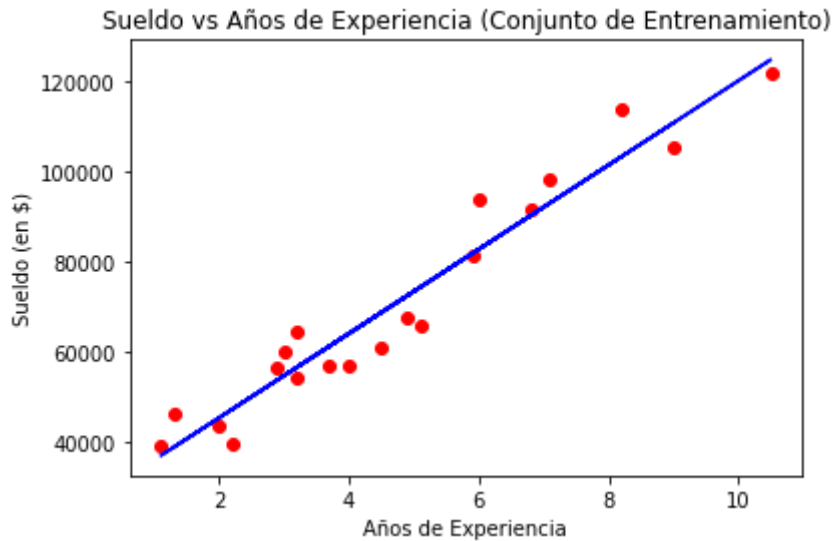
## Predecir el conjunto de testeo

```
In [33]: y_pred = regression.predict(X_test)
```

## Resultados para entrenamiento

```
In [34]: plt.scatter(X_train, y_train, color = "red")
plt.plot(X_train, regression.predict(X_train), color = "blue")
plt.title("Sueldo vs Años de Experiencia (Conjunto de Entrenamiento)")
plt.xlabel("Años de Experiencia")
```

```
plt.ylabel("Sueldo (en $)")
plt.show()
```



## Resultados para testeo

```
In [35]: plt.scatter(X_test, y_test, color = "red")
plt.plot(X_train, regression.predict(X_train), color = "blue") #es la misma recta con
plt.title("Sueldo vs Años de Experiencia (Conjunto de Testing)")
plt.xlabel("Años de Experiencia")
plt.ylabel("Sueldo (en $)")
plt.show()
```



El modelo no muestra una diferencia considerable con los datos reales, por lo que se comprueba que es un modelo aceptable y se puede usar para predecir con valores desconocidos.

In [ ]: