Machine Learning

Ex: 04 Linear regression

Name: Athithraja. R

Reg.no: 2022503702

1) Linear regression(Hard Code)

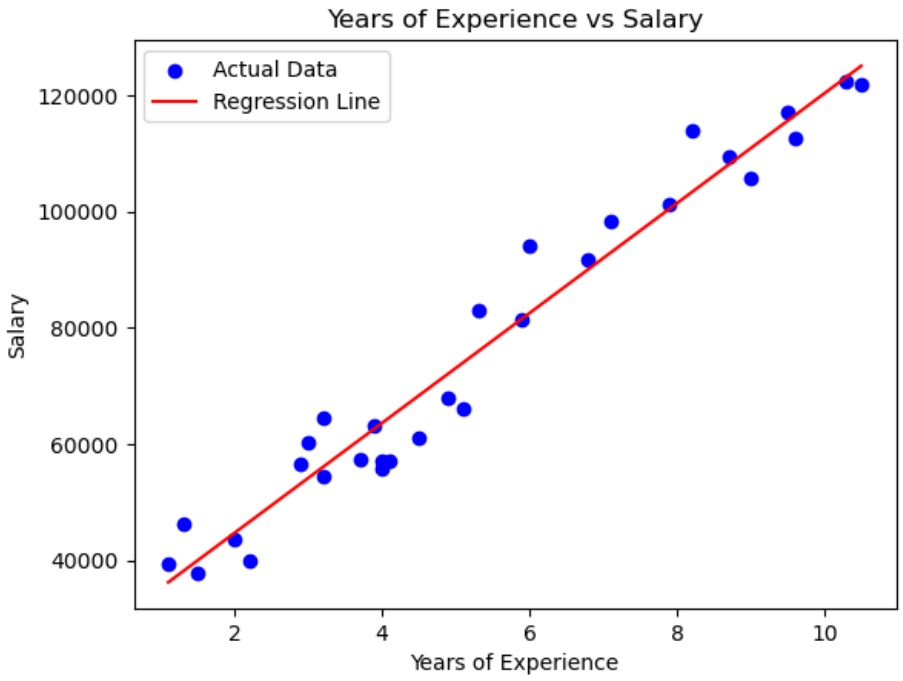
Code:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv('Salary_Data.csv')

X = df['YearsExperience']
y = df['Salary']

n = len(X)
sum_X = np.sum(X)
sum_Y = np.sum(y)
sum_XY = np.sum(X * y)
sum_X_squared = np.sum(X ** 2)

m = (n * sum_XY - sum_X * sum_Y) / (n * sum_X_squared - sum_X ** 2)
b = (sum_Y - m * sum_X) / n
```

```
print(f"Slope (m): {m}")

print(f"Intercept (b): {b}")


y_pred = m * X + b


plt.scatter(X, y, color='blue', label='Actual Data')

plt.plot(X, y_pred, color='red', label='Regression Line')

plt.title('Years of Experience vs Salary')

plt.xlabel('Years of Experience')

plt.ylabel('Salary')

plt.legend()

plt.show()
```

Output:

```
Slope (m): 9449.962321455077
Intercept (b): 25792.200198668685
```

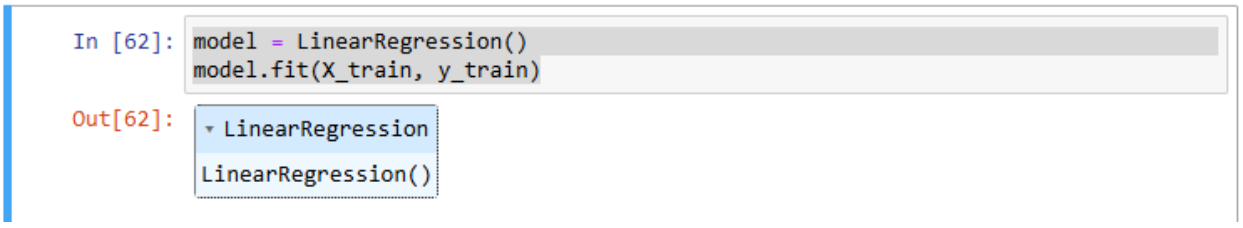


2) Linear Regression (sklearn):

Code:

```
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
df = pd.read_csv('Salary_Data.csv')
X = df[['YearsExperience']]
y = df['Salary']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
import matplotlib.pyplot as plt
plt.scatter(X, y, color='blue')
plt.plot(X, model.predict(X), color='red')
plt.title('Years of Experience vs Salary')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```

Output:




Years of Experience vs Salary
120000
100000
80000
60000
40000
Salary
2
4
6
8
10
Years of Experience

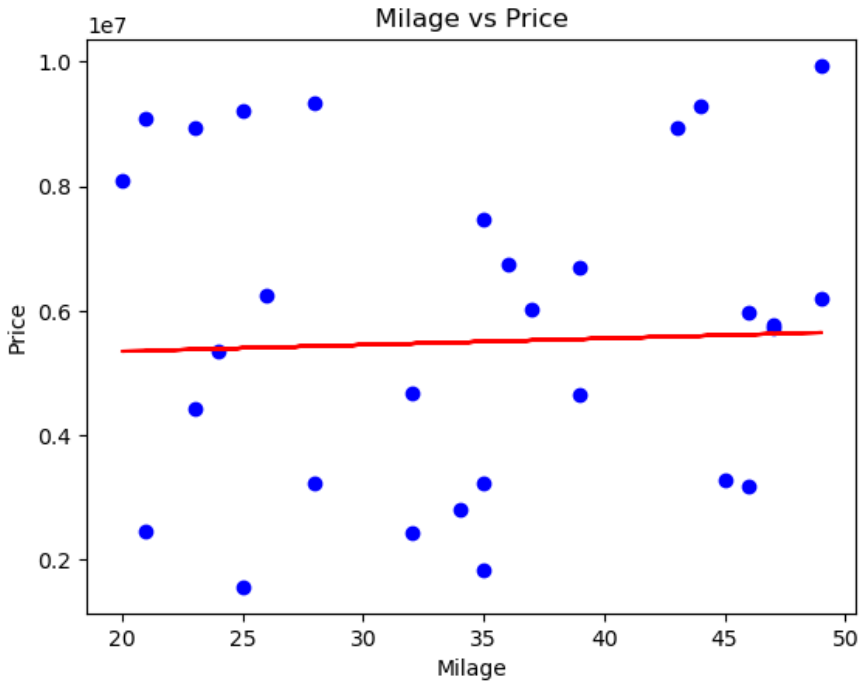
Actual Excersice:

1. Simple linear regression predict price of a car based on its milage:

Code:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
df = pd.read_excel('car_sales.xlsx')
X = df[['Milage']]
y = df['price']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
plt.scatter(X, y, color='blue')
plt.plot(X, model.predict(X), color='red')
plt.title('Milage vs Price')
plt.xlabel('Milage')
plt.ylabel('Price')
plt.show()
```

Output:



2. To Predict sales of car in us based on milage & price

Code:

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
df = pd.read_excel('car_sales 1.xlsx')
X = df[['Milage', 'price']]
y = df['sales']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
fig = plt.figure(figsize=(10, 7))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(df['Milage'], df['price'], df['sales'], color='blue', label="Actual Sales")
ax.scatter(X_test['Milage'], X_test['price'], y_pred, color='red', label="Predicted Sales",
marker='^')
ax.set_xlabel('Milage')
ax.set_ylabel('price')
ax.set_zlabel('sales')
ax.set_title('3D Scatter Plot of Mileage, Price & Sales')
ax.legend()
plt.show()
```

Output: