ML (CSE(AIML) - 5th Semester)

SUPRATIM NAG/AIML/22/057

Linear Regression --- In Personal Dataset

In [ ]:
```
import pandas as pd
import seaborn as sns
import numpy as np
from sklearn import preprocessing, svm
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [ ]:
```
file_path='../Personal_Datasets/KOLKATA_Housing.csv'
df=pd.read_csv(file_path)
```

In [ ]:
```
df_binary = df[['avg_area_house_age', 'Price']].copy()
df_binary_columns = ['avg_area_house_age', 'Price']
df_binary.head()
```

Out[ ]:

	avg_area_house_age	Price
0	6.234568	1500000.988
1	5.567890	1200000.099
2	7.890123	1300000.568
3	6.123457	1400000.123
4	5.987654	1250000.235

In [ ]:
```
x = df['avg_area_house_age']
y = df['Price']
```

In [ ]:
```
X = np.array(df_binary['avg_area_house_age']).reshape(-1,1)
Y = np.array(df_binary['Price']).reshape(-1,1)
#df_binary.dropna(inplace=True)
x_train,x_test,y_train,y_test=train_test_split(X,Y,test_size=0.33)
```

In [ ]:
```
lr=LinearRegression()
lr.fit(x_train,y_train)
```

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

In [ ]:
```
print(lr.score(x_test,y_test))
lr.coef_
lr.intercept_
```

```
m=lr.intercept_
b=lr.coef_
print('y=',m,'*X+',b)
```

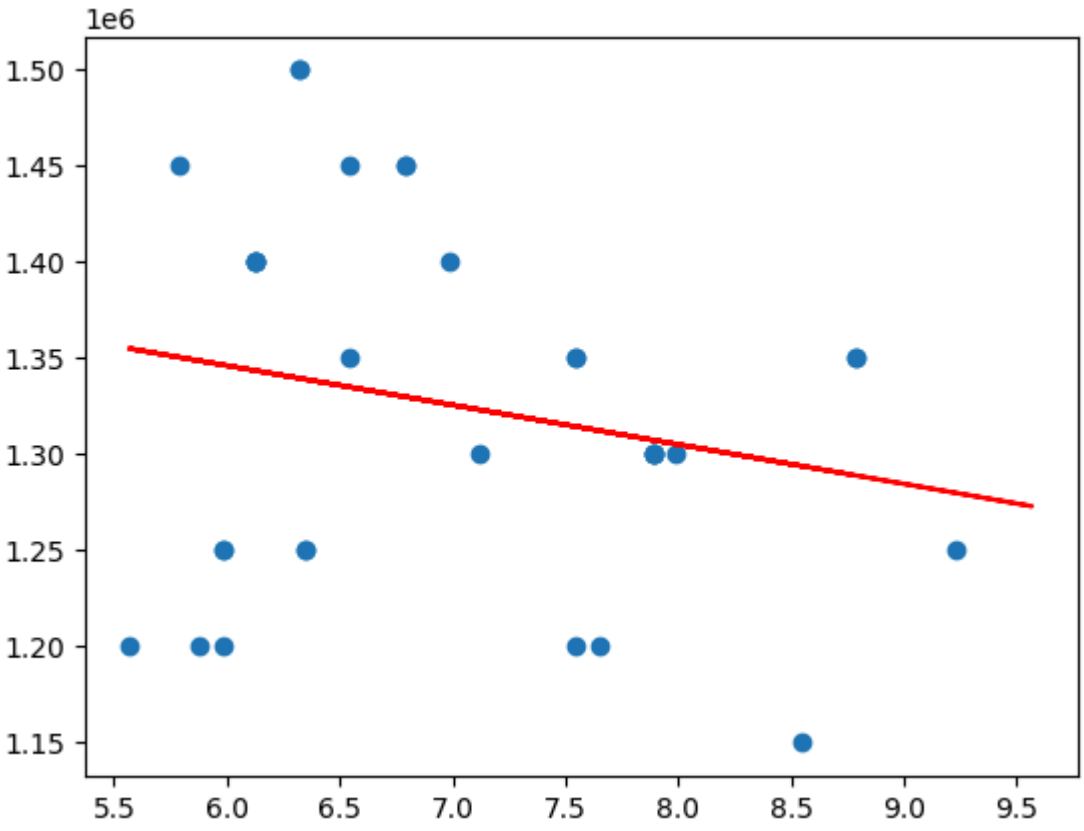
```
-0.004984411305229752
y= [1469295.29990309] *X+ [[-20506.47372358]]
```

In [ ]:
```
pred = lr.predict(x_test)
print(pred)
```

```
[[1323218.32043843]
 [1355117.50737952]
 [1330076.59662081]
 [1293598.0862293 ]
 [1323218.32043843]
 [1343724.79416201]
 [1323218.32043843]
 [1289547.19666359]
 [1341446.29728895]
 [1355117.50740003]
 [1294104.1682627 ]
 [1314610.64198628]
 [1355117.50737952]
 [1273091.61250573]
 [1350583.07034438]
 [1346509.62390365]
 [1346509.62390365]]
```

In [ ]:
```
plt.scatter(x_train,y_train)
plt.plot(x_test,pred,color='red')
```

Out[ ]: `[<matplotlib.lines.Line2D at 0x2747fe8d990>]`



In [ ]:
```
plt.scatter(x_test,y_test)
plt.plot(x_test,pred,color='red')
```

Out[ ]: [<matplotlib.lines.Line2D at 0x2747ff94650>]