LAB # 06

SUPERVISED LEARNING (LINEAR REGRESSION)

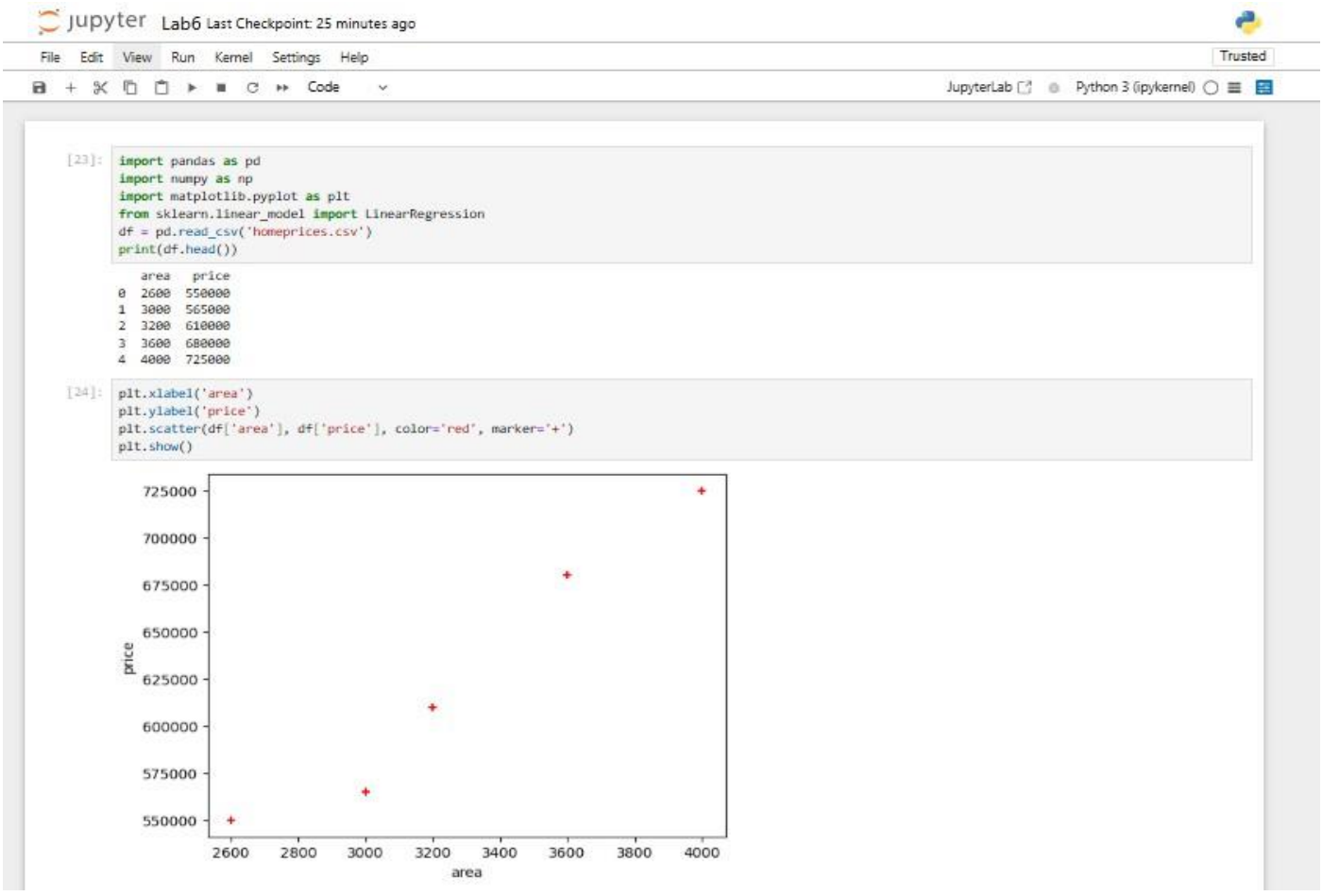
OBJECTIVES:

Implementing supervised learning, linear regression algorithm for training, testing and classification.

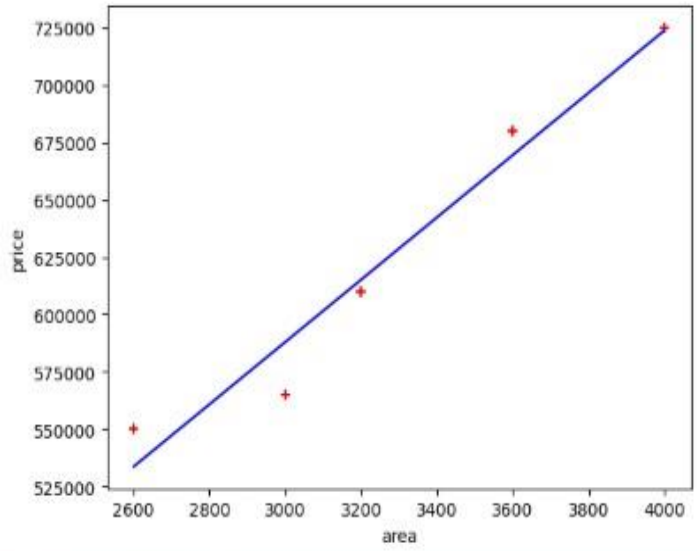
Lab Task

area	price
2600	550000
3000	565000
3200	610000
3600	680000
4000	725000

1. Implement linear regression algorithm on above dataset predict price of home with areas in the dataset by using (homeprices.csv).



```
[25]: X = df[['area']]
      y = df['price']

[26]: reg = LinearRegression()
      reg.fit(X, y)

[26]: LinearRegression()

[27]: predictions = reg.predict(X)
      print("Predicted Prices:", predictions)

Predicted Prices: [533664.38356164 587979.45205479 615136.98630137 669452.05479452
 723767.12328767]

[28]: plt.scatter(df['area'], df['price'], color='red', marker='+')
      plt.plot(df['area'], predictions, color='blue')
      plt.xlabel('area')
      plt.ylabel('price')
      plt.show()
```



2. Implement linear regression using table 1 in such a way that the:

Predict price of a home with area = 5000 Sqr. Ft.
Predict price of a home with area = 8000 Sqr. Ft.
Predict price of a home with area = 9000 Sqr. Ft.

```
[7]: areas = [[5000], [8000], [9000]]
     predicted_prices = reg.predict(areas)
     for area, price in zip(areas, predicted_prices):
         print(f"Predicted price for area {area[0]} sq ft: {price}")

Predicted price for area 5000 sq ft: 859554.7945205481
Predicted price for area 8000 sq ft: 1266917.8082191783
Predicted price for area 9000 sq ft: 1402705.479452055
C:\Users\wajiz.pk\anaconda3\Lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but LinearRegression was fitted wi
th feature names
  warnings.warn(

[8]: m = reg.coef_[0]
     b = reg.intercept_
     for area in [5000, 8000, 9000]:
         calculated_price = m * area + b
         print(f"Calculated price for area {area} sq ft: {calculated_price}")

Calculated price for area 5000 sq ft: 859554.7945205481
Calculated price for area 8000 sq ft: 1266917.8082191783
Calculated price for area 9000 sq ft: 1402705.479452055
```

Home Tasks:

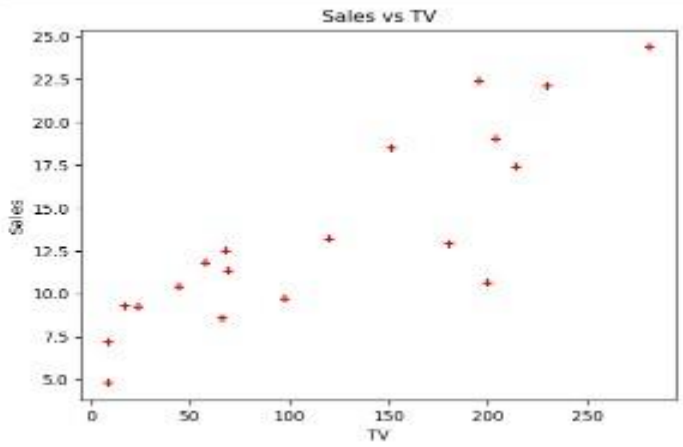
1. Implement Linear Regression on the given Advertising dataset to predict the sales of a product based on the amount spent on TV, Radio, and Newspaper advertisements?

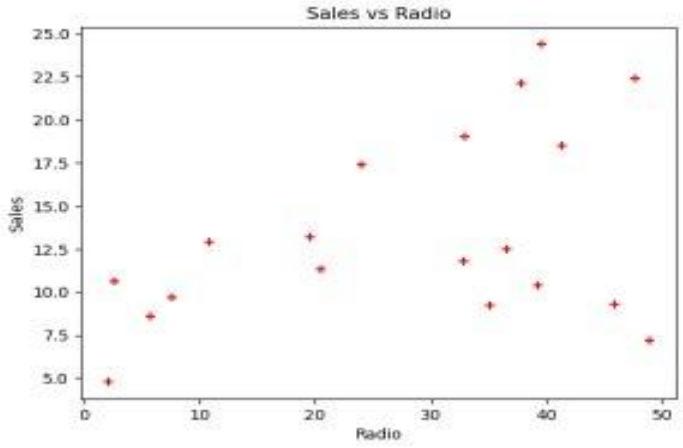
TV	Radio	Newspaper	Sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	9.3
151.5	41.3	58.5	18.5
180.8	10.8	58.4	12.9
8.7	48.9	75	7.2
57.5	32.8	23.5	11.8
120.2	19.6	11.6	13.2
8.6	2.1	1	4.8
199.8	2.6	21.2	10.6
66.1	5.8	24.2	8.6
214.7	24	4	17.4
23.8	35.1	65.9	9.2
97.5	7.6	7.2	9.7
204.1	32.9	46	19
195.4	47.7	52.9	22.4
67.8	36.6	114	12.5
281.4	39.6	55.8	24.4
69.2	20.5	18.3	11.3

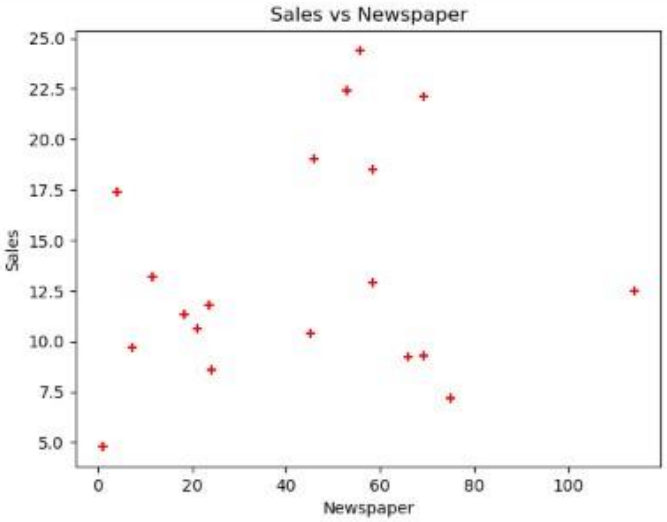
[30]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
df = pd.read_csv("Advertising.csv")
print(df)
```

```
      TV  Radio  Newspaper  Sales
0  230.1   37.8       69.2   22.1
1   44.5   39.3       45.1   10.4
2   17.2   45.9       69.3    9.3
3  151.5   41.3       58.5   18.5
4  180.8   10.8       58.4   12.9
5    8.7   48.9       75.0    7.2
6   57.5   32.8       23.5   11.8
7  120.2   19.6       11.6   13.2
8    8.6    2.1        1.0    4.8
9  199.8    2.6       21.2   10.6
10  66.1    5.8       24.2    8.6
11 214.7   24.0        4.0   17.4
12  23.8   35.1       65.9    9.2
13  97.5    7.6        7.2    9.7
14 204.1   32.9       46.0   19.0
15 195.4   47.7       52.9   22.4
16  67.8   36.6      114.0   12.5
17 281.4   39.6       55.8   24.4
18  69.2   20.5       18.3   11.3
```

[31]:
```
features = ['TV', 'Radio', 'Newspaper']
for feature in features:
    plt.scatter(df[feature], df['Sales'], label=feature, color='red', marker='+')
    plt.xlabel(feature)
    plt.ylabel('Sales')
    plt.title(f'Sales vs {feature}')
    plt.show()
```







[32]:
```
X = df[['TV', 'Radio', 'Newspaper']]
y = df['Sales']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

[33]:
```
model = LinearRegression()
model.fit(X_train, y_train)
```

[33]:
```
LinearRegression()
```

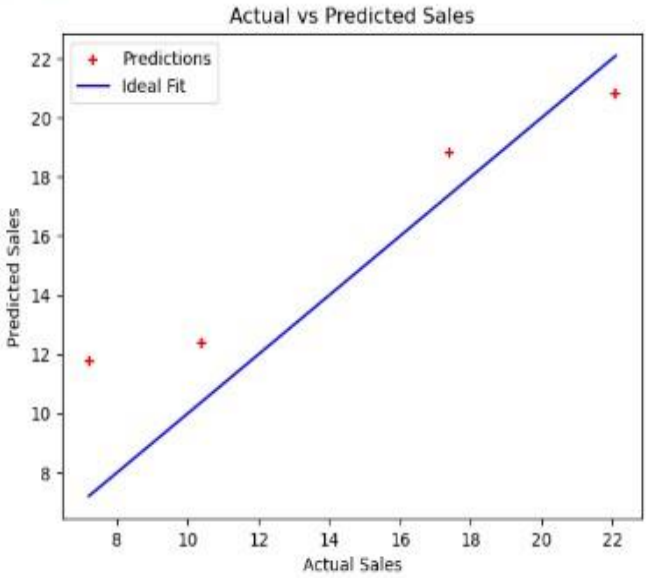
[34]:
```
y_pred = model.predict(X_test)
print(f"Coefficients: {model.coef_}")
print(f"Intercept: {model.intercept_}")
```

```
Coefficients: [ 0.05021002  0.20298284 -0.02443115]
Intercept: 3.254071939471327
```

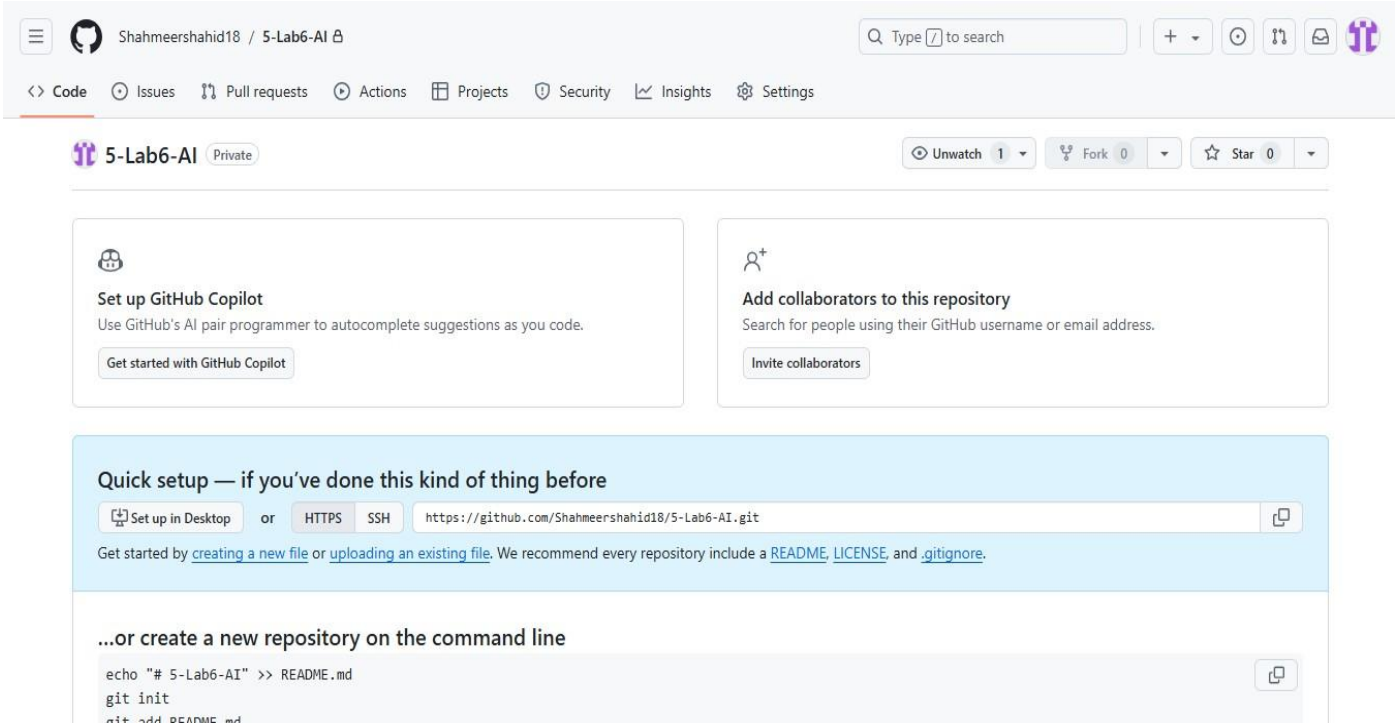
[35]:
```
sample_data = [[300, 50, 20]]
predicted_sales = model.predict(sample_data)
print(f"Predicted Sales: {predicted_sales[0]}")
```

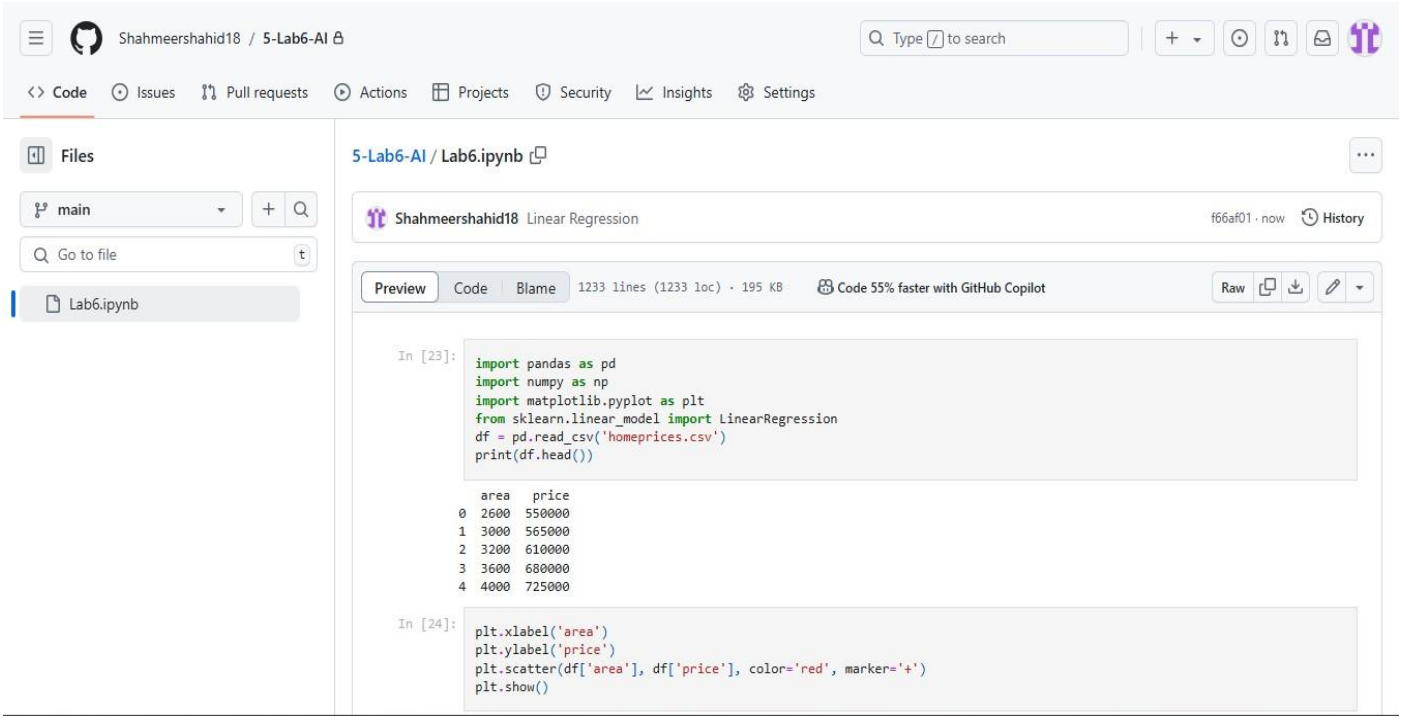
```
Predicted Sales: 27.97759822277094
C:\Users\wajiz.pk\anaconda3\Lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but LinearRegression was fitted wi
th feature names
  warnings.warn(
```

[38]:
```
plt.scatter(y_test, y_pred, color='red', marker='+', label='Predictions')
plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)], color='blue', label='Ideal Fit')
plt.xlabel("Actual Sales")
plt.ylabel("Predicted Sales")
plt.title("Actual vs Predicted Sales")
plt.legend()
plt.show()
```



Uploaded File On GitHub:




Shahmeershahid18 / 5-Lab6-AI
Type / to search
Code
Issues
Pull requests
Actions
Projects
Security
Insights
Settings
Files
main
Go to file
Lab6.ipynb
5-Lab6-AI / Lab6.ipynb
Shahmeershahid18 Linear Regression
f66af01 · now
History
Preview
Code
Blame
1233 lines (1233 loc) · 195 KB
Code 55% faster with GitHub Copilot
Raw
In [23]:
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
df = pd.read_csv('homeprices.csv')
print(df.head())
area price
0 2600 550000
1 3000 565000
2 3200 610000
3 3600 680000
4 4000 725000
In [24]:
plt.xlabel('area')
plt.ylabel('price')
plt.scatter(df['area'], df['price'], color='red', marker='+')
plt.show()