Practical - 12

AIM

Implement Linear Regression problem.

PROBLEM

Based on a dataset consisting of an existing set of prices and area/size of the houses, predict the estimated price of a given house.

CODE & OUTPUT

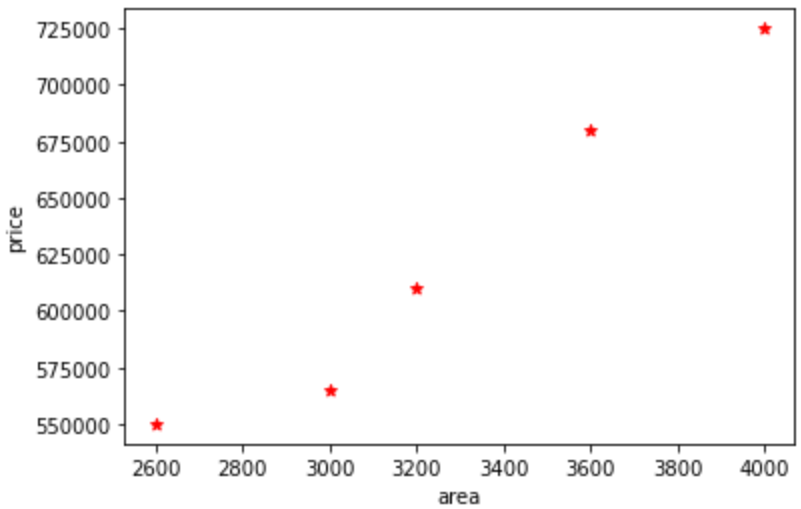
In [2]:
```
import matplotlib
import pandas as pd
import numpy as np
from sklearn import linear_model
import matplotlib.pyplot as plt
```

In [3]:
```
df = pd.read_csv("homeprices.csv")
print(df)

plt.xlabel('area')
plt.ylabel('price')
plt.scatter(df.area, df.price, color = "red", marker="*")
```

```
   area   price
0  2600  550000
1  3000  565000
2  3200  610000
3  3600  680000
4  4000  725000
```

Out[3]:
```
<matplotlib.collections.PathCollection at 0x141760006d0>
```



In [4]:
```
new_df = df.drop("price", axis="columns")
print(new_df)

price = df.price
print(price)
```

```
   area
0  2600
1  3000
2  3200
3  3600
4  4000
0    550000
1    565000
2    610000
3    680000
4    725000
Name: price, dtype: int64
```

In [5]:

```
reg = linear_model.LinearRegression()
reg.fit(new_df, price)
reg.predict([[3300]])
```

Out[5]:

```
array([628715.75342466])
```

In [6]:

```
reg.coef_
reg.intercept_

reg.predict([[5000]])

area_df = pd.read_csv("areas.csv")
print(area_df.head(3))

p = reg.predict(area_df)
print(p)
```

```
   area
0  1000
1  1500
2  2300
[ 316404.10958904  384297.94520548  492928.08219178  661304.79452055
  740061.64383562  799808.21917808  926090.75342466  650441.78082192
  825607.87671233  492928.08219178 1402705.47945205 1348390.4109589
 1144708.90410959]
```

In [7]:

```
area_df["prices"] = p
print(area_df)

area_df.to_csv("prediction.csv")
```

```
    area        prices
0   1000  3.164041e+05
1   1500  3.842979e+05
2   2300  4.929281e+05
3   3540  6.613048e+05
4   4120  7.400616e+05
5   4560  7.998082e+05
6   5490  9.260908e+05
7   3460  6.504418e+05
8   4750  8.256079e+05
9   2300  4.929281e+05
10  9000  1.402705e+06
11  8600  1.348390e+06
12  7100  1.144709e+06
```