exercise02

November 14, 2024

Task: - Predict the percapita income of canada in 2020 using linear regression.

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

df = pd.read_csv("canada_per_capita_income.csv")
df.head()
```

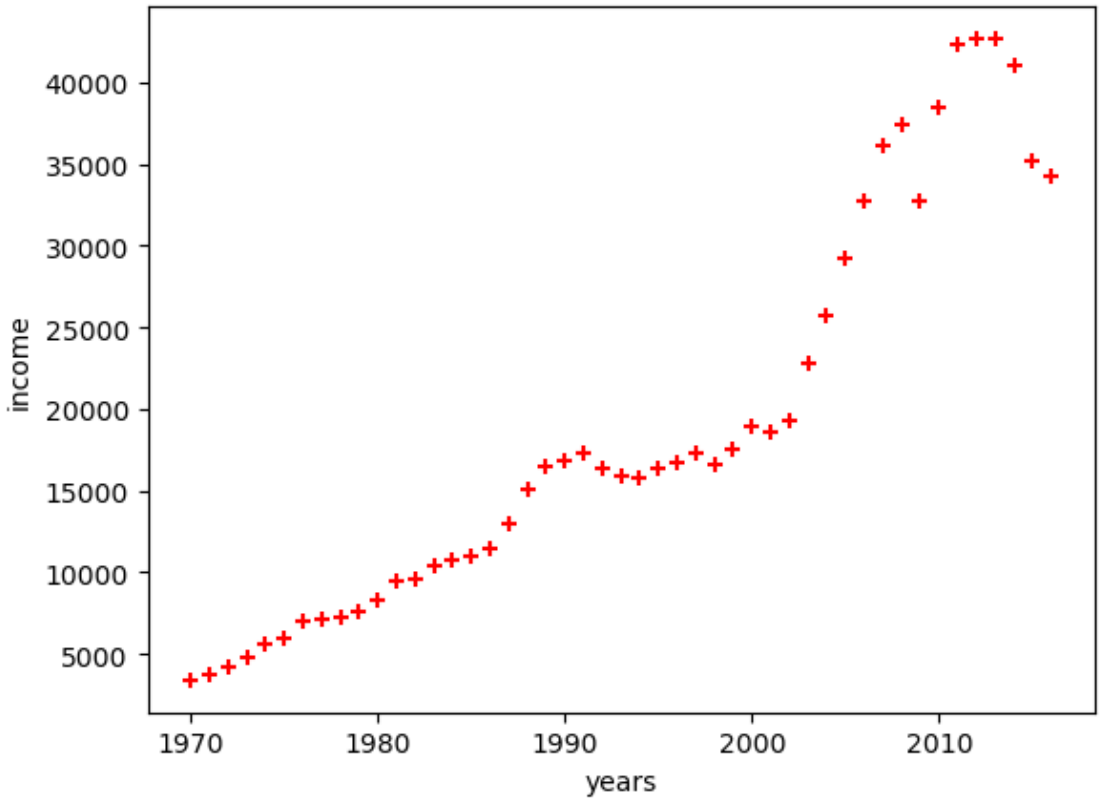
[2]:

	year	per capita income (US$)
0	1970	3399.299037
1	1971	3768.297935
2	1972	4251.175484
3	1973	4804.463248
4	1974	5576.514583

[3]:
```
# Rename the column for easier access
df = df.rename(columns={'per capita income (US$)': 'income'})

# Then plot
plt.xlabel('years')
plt.ylabel('income')
plt.scatter(df['year'], df['income'], color='red', marker='+')
```

[3]:
```
<matplotlib.collections.PathCollection at 0x1c2624cccb0>
```



[4]:
```
new_df = df.drop('income',axis='columns')
new_df.head(10)
```

[4]:
```
   year
0  1970
1  1971
2  1972
3  1973
4  1974
5  1975
6  1976
7  1977
8  1978
9  1979
```

[5]:
```
income = df.income
income
```

[5]:
```
0       3399.299037
1       3768.297935
2       4251.175484
```

```
3      4804.463248
4      5576.514583
5      5998.144346
6      7062.131392
7      7100.126170
8      7247.967035
9      7602.912681
10     8355.968120
11     9434.390652
12     9619.438377
13    10416.536590
14    10790.328720
15    11018.955850
16    11482.891530
17    12974.806620
18    15080.283450
19    16426.725480
20    16838.673200
21    17266.097690
22    16412.083090
23    15875.586730
24    15755.820270
25    16369.317250
26    16699.826680
27    17310.757750
28    16622.671870
29    17581.024140
30    18987.382410
31    18601.397240
32    19232.175560
33    22739.426280
34    25719.147150
35    29198.055690
36    32738.262900
37    36144.481220
38    37446.486090
39    32755.176820
40    38420.522890
41    42334.711210
42    42665.255970
43    42676.468370
44    41039.893600
45    35175.188980
46    34229.193630
Name: income, dtype: float64
```

[6]:
```
# Create linear regression object
reg = linear_model.LinearRegression()
reg.fit(new_df,income)
```

[6]:
```
LinearRegression()
```

[7]:
```
reg.predict([[2020]])
```

```
d:\Projects\Anaconda_Installed\Lib\site-packages\sklearn\base.py:493:
UserWarning: X does not have valid feature names, but LinearRegression was
fitted with feature names
  warnings.warn(
```

[7]:
```
array([41288.69409442])
```