

Exercise

Predict Canada's per capita income in year 2020. There is an exercise folder here on github download that and you will find `canada_per_capita_income.csv` file. Using this build a regression model and predict the per capita income fo Canadian citizens in year 2020.

Answer 41288.69409442

```
#Required imports
```

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import linear_model
```

```
# Reading csv file to dataframe
```

```
In [2]: df = pd.read_csv('./data/canada_per_capita_income.csv')
df.head()
```

```
Out[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

```
# Last 5 rows
```

```
In [3]: df.tail()
```

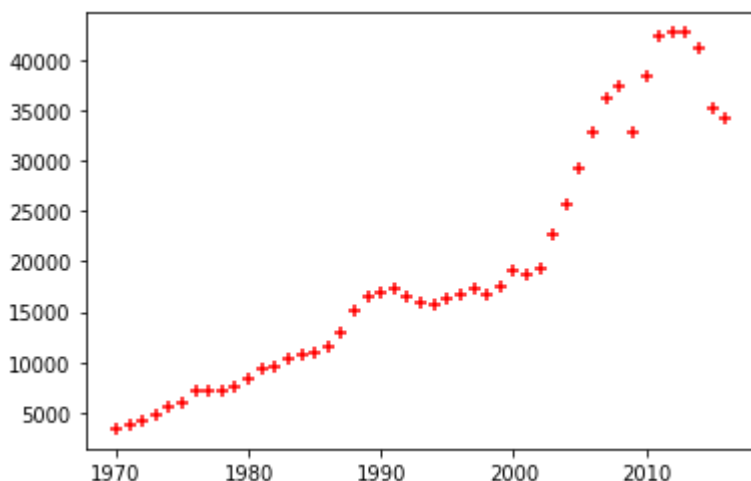
```
Out[3]:
```

	year	per capita income (US\$)
42	2012	42665.25597
43	2013	42676.46837
44	2014	41039.89360
45	2015	35175.18898
46	2016	34229.19363

```
# Scatter plot for the dataset
```

```
In [4]: %matplotlib inline
plt.xlabel = "year"
plt.ylabel = "per capita income (US$)"
plt.scatter(df.year, df['per capita income (US$)'], color='red', marker='+')
```

Out[4]: <matplotlib.collections.PathCollection at 0x178e865f790>



Preparing data for linear regression,

```
In [5]: x_df = df.drop('per capita income (US$)', axis='columns')
x_df.head()
```

```
Out[5]:   year
0  1970
1  1971
2  1972
3  1973
4  1974
```

```
In [6]: y_df = df['per capita income (US$)']
y_df.head()
```

```
Out[6]: 0    3399.299037
1    3768.297935
2    4251.175484
3    4804.463248
4    5576.514583
Name: per capita income (US$), dtype: float64
```

Applying Linear Regression

```
model = linear_model.LinearRegression()
model.fit(x_df, y_df)
```

Out[7]:

```
LinearRegression()
```

In [8]:

```
m = model.coef_
c = model.intercept_
print('Coefficient, m = ', m)
print('Intercept, c = ', c)
Coefficient, m = [828.46507522]
Intercept, c = -1632210.7578554575
```

In [9]:

```
plt.scatter(df.year,
            df['per capita income (US$)'],
            color='red',
            marker='+',
```

```

label='per capita income (US$)')

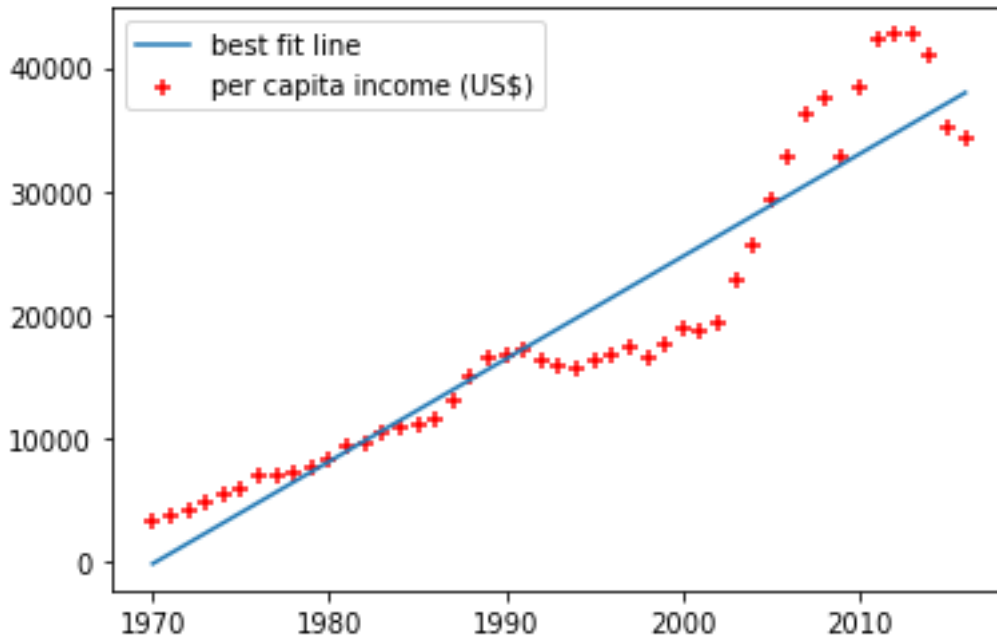
# best fit line
y = [(m*i+c) for i in df.year]
plt.plot(df.year, y, label='best fit line')

plt.legend()

<matplotlib.legend.Legend at 0x2a30f12cd90>

```

Out[9]:



Predicting 'per capita income for Canadian citizens in year 2020'

In [10]:

```

ans = model.predict([[2020]])
print('The per capita income for Canadian citizens in year 2020 will be,
\nUS${}'.format(ans))
The per capita income for Canadian citizens in year 2020 will be,
US$[41288.69409442]
Calculating y for y = m*x +c,

```

In [11]:

```

y = m*2020 + c
print('y = m*x +c =', y)
y = m*x +c = [41288.69409442]
Here, we can see that y = ans = 41288.69409442.

```