

# Linear Regression - 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

```
In [ ]: #Required imports

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import linear_model
```

```
In [ ]: # Reading csv file to dataframe

df = pd.read_csv('./data/canada_per_capita_income.csv')
df.head()
```

```
Out[ ]:   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
```

```
In [ ]: # Last 5 rows

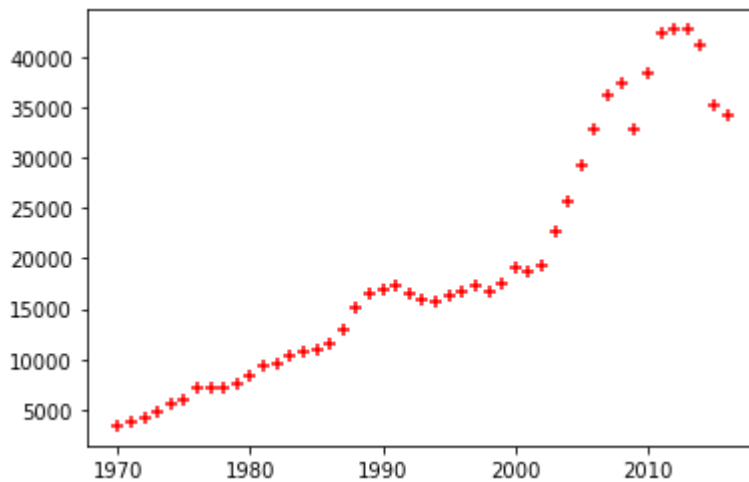
df.tail()
```

```
Out[ ]:   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
```

```
In [ ]: # Scatter plot for the dataset

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

Out[ ]: <matplotlib.collections.PathCollection at 0x2a30ee33070>



## Preparing data for linear regression,

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

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

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

```
Out[ ]: 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

```
In [ ]: model = linear_model.LinearRegression()
        model.fit(x_df, y_df)
```

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

```
In [ ]: m = model.coef_
        c = model.intercept_
        print('Coefficient, m = ', m)
        print('Intercept, c = ', c)
```

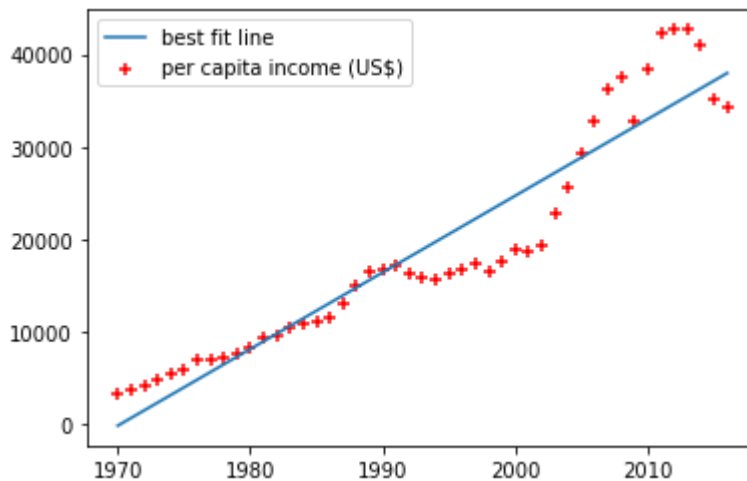
```
Coefficient, m = [828.46507522]
Intercept, c = -1632210.7578554575
```

```
In [ ]: 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()
```

Out[ ]: <matplotlib.legend.Legend at 0x2a30f12cd90>



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

```
In [ ]: ans = model.predict([[2020]])
print('The per capita income for Canadian citizens in year 2020 will be, \nUS${}'.format
```

The per capita income for Canadian citizens in year 2020 will be,  
US\$[41288.69409442]

Calculating  $y$  for  $y = m \cdot x + c$ ,

```
In [ ]: y = m*2020 + c
print('y = m*x + c =', y)
```

$y = m \cdot x + c = [41288.69409442]$

Here, we can see that  $y = \text{ans} = 41288.69409442$ .