In [11]:
```
import pandas as pd
import numpy as np
from sklearn import linear_model
import matplotlib.pyplot as plt
```

In [30]:
```
df=pd.read_csv('https://raw.githubusercontent.com/codebasics/py/master/ML/1_linear_reg/Exercise/canada_per_capita_income.csv')
```

In [31]:
```
df
```

Out[31]:

	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
5	1975	5998.144346
6	1976	7062.131392
7	1977	7100.126170
8	1978	7247.967035
9	1979	7602.912681
10	1980	8355.968120
11	1981	9434.390652
12	1982	9619.438377
13	1983	10416.536590
14	1984	10790.328720
15	1985	11018.955850
16	1986	11482.891530
17	1987	12974.806620
18	1988	15080.283450
19	1989	16426.725480
20	1990	16838.673200
21	1991	17266.097690
22	1992	16412.083090
23	1993	15875.586730
24	1994	15755.820270
25	1995	16369.317250
26	1996	16699.826680
27	1997	17310.757750
28	1998	16622.671870
29	1999	17581.024140
30	2000	18987.382410
31	2001	18601.397240
32	2002	19232.175560
33	2003	22739.426280
34	2004	25719.147150
35	2005	29198.055690
36	2006	32738.262900
37	2007	36144.481220
38	2008	37446.486090
39	2009	32755.176820
40	2010	38420.522890
41	2011	42334.711210
42	2012	42665.255970
43	2013	42676.468370
44	2014	41039.893600
45	2015	35175.188980
46	2016	34229.193630

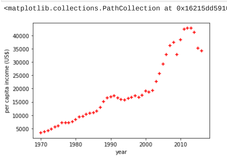
In [43]:
```
df1=df.rename({'per capita income (US$)':'income'},axis=1)
df1
```

Out[43]:

	year	income
0	1970	3399.299037
1	1971	3768.297935
2	1972	4251.175484
3	1973	4804.463248
4	1974	5576.514583
5	1975	5998.144346
6	1976	7062.131392
7	1977	7100.126170
8	1978	7247.967035
9	1979	7602.912681
10	1980	8355.968120
11	1981	9434.390652
12	1982	9619.438377
13	1983	10416.536590
14	1984	10790.328720
15	1985	11018.955850
16	1986	11482.891530
17	1987	12974.806620
18	1988	15080.283450
19	1989	16426.725480
20	1990	16838.673200
21	1991	17266.097690
22	1992	16412.083090
23	1993	15875.586730
24	1994	15755.820270
25	1995	16369.317250
26	1996	16699.826680
27	1997	17310.757750
28	1998	16622.671870
29	1999	17581.024140
30	2000	18987.382410
31	2001	18601.397240
32	2002	19232.175560
33	2003	22739.426280
34	2004	25719.147150
35	2005	29198.055690
36	2006	32738.262900
37	2007	36144.481220
38	2008	37446.486090
39	2009	32755.176820
40	2010	38420.522890
41	2011	42334.711210
42	2012	42665.255970
43	2013	42676.468370
44	2014	41039.893600
45	2015	35175.188980
46	2016	34229.193630

In [45]:
```
%matplotlib inline
plt.xlabel('year')
plt.ylabel('per capita income (US$)')
plt.scatter(df1.year,df1.income,color='red',marker='+')
```

Out[45]:
```
<matplotlib.collections.PathCollection at 0x16215dd5910>
```



In [46]:
```
new_df=df1.drop('income',axis='columns')
```

In [47]:
```
new_df
```

Out[47]:

	year
0	1970
1	1971
2	1972
3	1973
4	1974
5	1975
6	1976
7	1977
8	1978
9	1979
10	1980
11	1981
12	1982
13	1983
14	1984
15	1985
16	1986
17	1987
18	1988
19	1989
20	1990
21	1991
22	1992
23	1993
24	1994
25	1995
26	1996
27	1997
28	1998
29	1999
30	2000
31	2001
32	2002
33	2003
34	2004
35	2005
36	2006
37	2007
38	2008
39	2009
40	2010
41	2011
42	2012
43	2013
44	2014
45	2015
46	2016

In [50]:
```
model= linear_model.LinearRegression()
model.fit(new_df,df1.income)
```

Out[50]:
```
LinearRegression()
```

In [51]:
```
model.predict([[2020]])
```

Out[51]:
```
array([41288.69409442])
```

In [ ]:
```

```