

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

```
In [2]: df = pd.read_csv("china_sales.csv")
df.head()
```

Out[2]:

	Year	value
0	1960	5.918412e+10
1	1961	4.955705e+10
2	1962	4.668518e+10
3	1963	5.009730e+10
4	1964	5.906225e+10

```
In [7]: x1 = df['Year'].values
x2 = df['value'].values

x1 = x1.reshape(-1,1)
#x2 = x2.reshape(-1,1)
```

```
In [5]: print(x1)
```

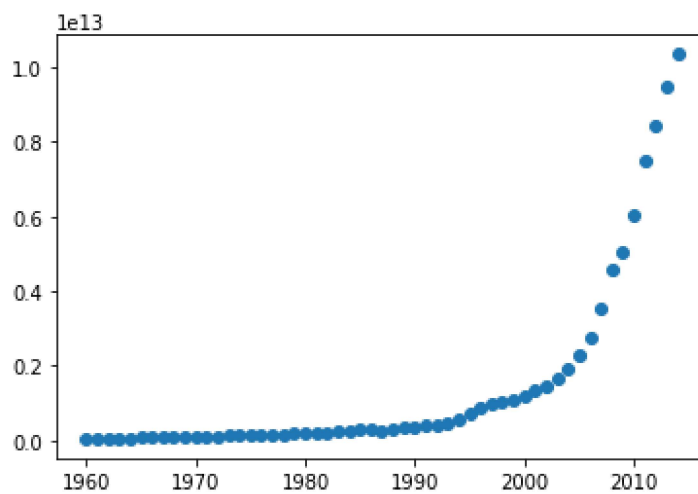
```
0      1960
1      1961
2      1962
3      1963
4      1964
5      1965
6      1966
7      1967
8      1968
9      1969
10     1970
11     1971
12     1972
13     1973
14     1974
15     1975
16     1976
17     1977
18     1978
19     1979
20     1980
21     1981
22     1982
23     1983
24     1984
25     1985
26     1986
27     1987
28     1988
29     1989
30     1990
31     1991
32     1992
33     1993
34     1994
35     1995
36     1996
37     1997
38     1998
39     1999
40     2000
41     2001
42     2002
43     2003
44     2004
45     2005
46     2006
47     2007
48     2008
49     2009
50     2010
51     2011
52     2012
53     2013
54     2014
```

```
Name: Year, dtype: int64
```

```
In [8]: print(x2)
```

```
[5.91841165e+10 4.95570502e+10 4.66851785e+10 5.00973033e+10
 5.90622549e+10 6.97091531e+10 7.58794348e+10 7.20570286e+10
 6.99934979e+10 7.87188205e+10 9.15062113e+10 9.85620238e+10
 1.12160000e+11 1.36770000e+11 1.42255000e+11 1.61162000e+11
 1.51628000e+11 1.72349000e+11 1.48382000e+11 1.76857000e+11
 1.89650000e+11 1.94369000e+11 2.03550000e+11 2.28950000e+11
 2.58082000e+11 3.07480000e+11 2.98806000e+11 2.71350000e+11
 3.10722000e+11 3.45957000e+11 3.58973000e+11 3.81455000e+11
 4.24934000e+11 4.42875000e+11 5.62261000e+11 7.32032000e+11
 8.60844000e+11 9.58159000e+11 1.02528000e+12 1.08945000e+12
 1.20526000e+12 1.33223000e+12 1.46191000e+12 1.64993000e+12
 1.94175000e+12 2.26860000e+12 2.72978000e+12 3.52309000e+12
 4.55843000e+12 5.05942000e+12 6.03966000e+12 7.49243000e+12
 8.46162000e+12 9.49060000e+12 1.03548000e+13]
```

```
In [9]: plt.scatter(x1, x2);
```



```
In [10]: from sklearn.pipeline import Pipeline
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import Ridge

def get_preds_ridge(x1, x2, alpha):
    model = Pipeline([
        ('poly_features', PolynomialFeatures(degree = 5)),
        ('ridge', Ridge(alpha=alpha))
    ])
    model.fit(x1, x2)
    return model.predict(x1)
```

```

In [11]: alphas = [0, 20, 200]
cs = ['r', 'g', 'b']

plt.figure(figsize=(10,6))
plt.plot(x1, x2, 'b+', label='Datapoints')

for alpha, c in zip(alphas, cs):
    preds = get_preds_ridge(x1, x2, alpha)
    #plot
    plt.plot(sorted(x1[:,0]), preds[np.argsort(x1[:,0])], c, label = 'Alpha: {}'.format(alpha))

plt.legend()

```

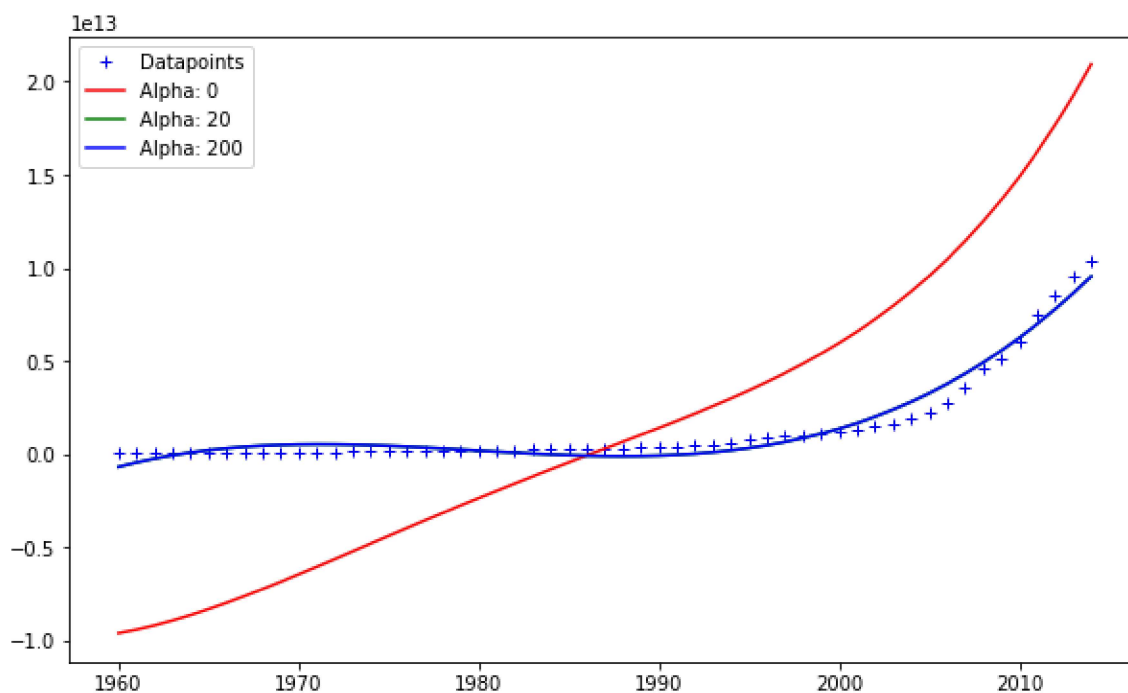
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\linear_model\ridge.py:125: LinAlgWarning: Ill-conditioned matrix (rcond=2.37095e-31): result may not be accurate.

overwrite_a=True).T

C:\ProgramData\Anaconda3\lib\site-packages\sklearn\linear_model\ridge.py:125: LinAlgWarning: Ill-conditioned matrix (rcond=2.37057e-30): result may not be accurate.

overwrite_a=True).T

Out[11]: <matplotlib.legend.Legend at 0x209feb3b7b8>



In []: