In [ ]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [ ]:

```
ds = pd.read_csv("S&P 500 Historical Data.csv")
df = pd.DataFrame(ds)
arr = np.array(df)
Date = np.flip(np.transpose(arr[:,np.where(np.array(df.columns) == 'Date')[0]])[0])
Date = pd.to_datetime(Date)
prices = np.flip(np.transpose(arr[:,np.where(np.array(df.columns) == 'Price')[0]])[0]
prices = prices.astype(str)
prices = np.char.replace(prices,',','').astype(float)
```

In [ ]:

```
### Get x dataset and corresponding y final value
### this regression uses the past __ many days as features when making the next day's
### set the number of instances
percentTrain = 0.4
## set the number of days back you will look in the linear regression
days = 50
### now build the combined array with data of x|y
ind = np.random.randint(0,len(prices)-days - 1)
data = np.reshape(prices[ind:ind + days + 1],(1,days + 1))
for _ in range(int(percentTrain * len(prices))):
    ind = np.random.randint(0,len(prices)-days - 1)
    add_data = np.reshape(prices[ind:ind + days + 1],(1,days + 1))
    data = np.concatenate((data,add_data))
previous = data[:,0:days]
final = np.reshape(data[:,days],(previous.shape[0],1))
print(final.shape)
```

```
(227, 1)
```

In [ ]:

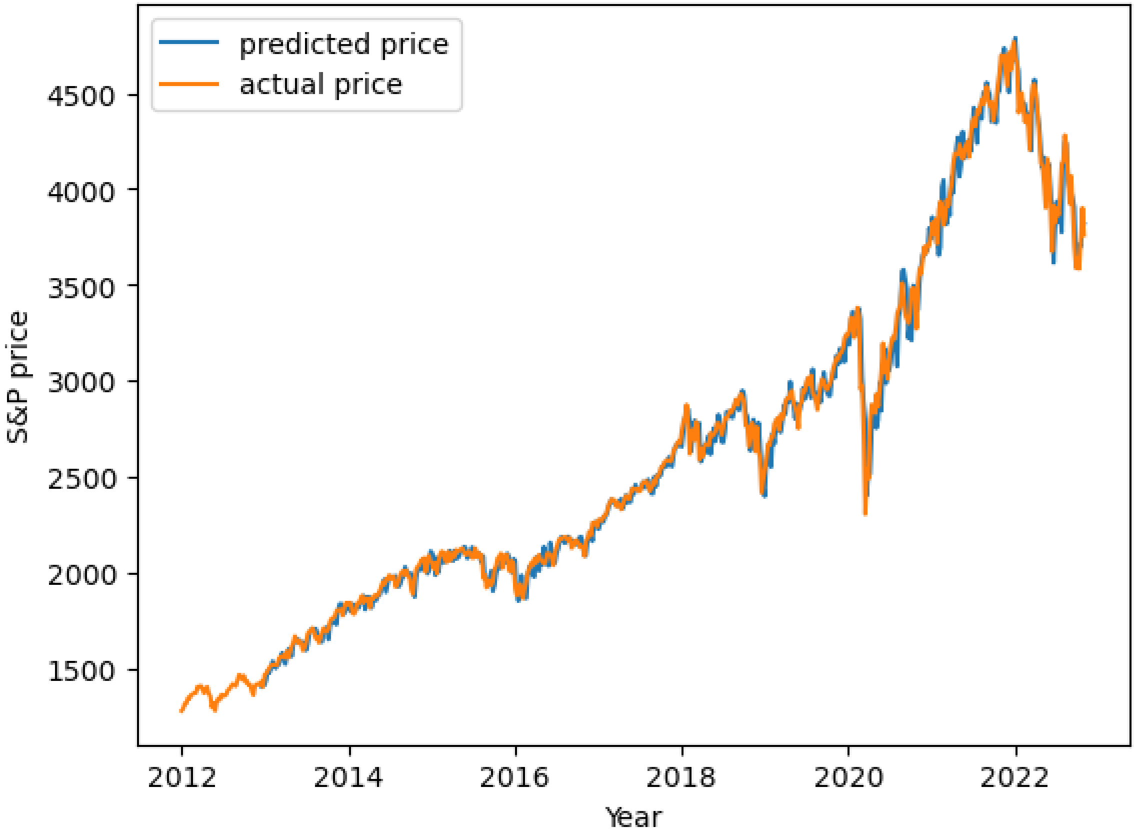
```
from regression import Regression
linReg = Regression()
a = linReg.linear_fit_closed(previous, final)
### make array with previous 'days' prices to make predictions
ind = 0
data = np.reshape(prices[ind:ind + days],(1,days))
for ind in range(1,int(len(prices)-days)):
    add_data = np.reshape(prices[ind:ind + days],(1,days))
    data = np.concatenate((data,add_data))
print(a.shape)
pred = linReg.predict(data,a)
print(pred.shape)
```

```
(50, 1)
(516, 1)
```

In [ ]:

```
plt.plot(Date[days:], pred, label = 'predicted price')
plt.plot(Date,prices, label = 'actual price')
plt.legend()
plt.xlabel("Year")
plt.ylabel("S&P price")
print("rmse:")
print(linReg.rmse(prices[days:],pred))
```

```
rmse:
1227.9173562715464
```


predicted price
actual price
4500
4000
3500
3000
2500
2000
1500
S&P price
2012
2014
2016
2018
2020
2022
Year