```
In [15]: import pandas as pd
         import matplotlib.pyplot as plt
         import numpy as np
```

```
In [16]: df = pd.read_csv('./Datasets/exp_1_A.csv')
         df.head()
```

Out[16]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

```
In [17]: df.describe()
```

Out[17]:

	YearsExperience	Salary
count	30.000000	30.000000
mean	5.313333	76003.000000
std	2.837888	27414.429785
min	1.100000	37731.000000
25%	3.200000	56720.750000
50%	4.700000	65237.000000
75%	7.700000	100544.750000
max	10.500000	122391.000000

```
In [18]: from sklearn.linear_model import LinearRegression
         from sklearn.model_selection import train_test_split
```

```
In [19]: x = df[['YearsExperience']]
         y = df['Salary']
```

```
In [20]: x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_st
         model = LinearRegression()
         model.fit(x_train,y_train)
         y_pred = model.predict(x_test)
         train_score = model.score(x_train,y_train)
         test_score = model.score(x_test,y_test)
         print(f"Train Score: {round(train_score*100,2)}")
         print(f"Test Score: {round(test_score*100,2)}")
```

```
Train Score: 94.12
Test Score: 98.82
```

In [21]:
```python
plt.scatter(x_train,y_train)
plt.plot(x_train, model.predict(x_train),color='red')
plt.show()
```
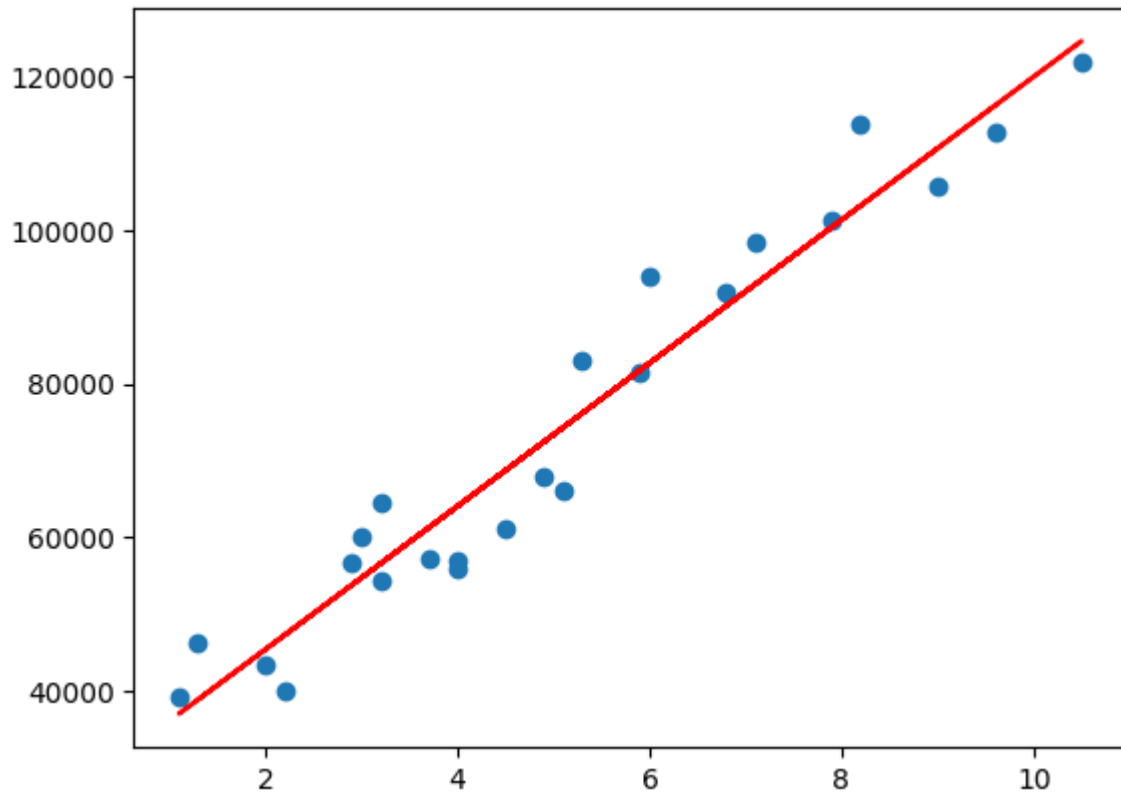

In [22]:
```python
plt.scatter(x_test,y_test)
plt.plot(x_test, model.predict(x_test),color='red')
plt.show()
```