In [1]:
```python
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
import sklearn.metrics as mc
```

In [2]:
```python
saldata= pd.read_csv('Salary_Data.csv')
x= saldata.iloc[:,:-1].values
y=saldata.iloc[:,1].values
```

In [3]:
```python
xtrain,xtest,ytrain,ytest= train_test_split(x,y,test_size=1/3,random_state=0)
```

In [4]:
```python
lr= LinearRegression()
lr.fit(xtrain,ytrain)
pred=lr.predict(xtest)
pred
```

Out[4]:
```
array([ 40835.10590871, 123079.39940819,  65134.55626083,  63265.36777221,
       115602.64545369, 108125.8914992 , 116537.23969801,  64199.96201652,
        76349.68719258, 100649.1375447 ])
```

In [5]:
```python
plt.scatter(x,y,color="green",marker="*")
plt.scatter(xtest,pred,color="blue",marker="*")
plt.plot(xtest,pred,color="red")
```

Out[5]:
```
[<matplotlib.lines.Line2D at 0x152aa3d5b50>]
```
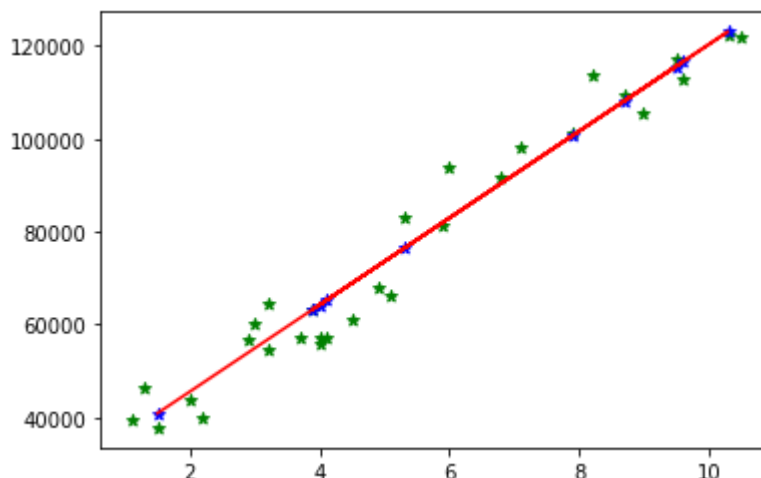

In [6]:
```python
#root mean squared error
mc.r2_score(ytest,pred)
```

Out[6]:
```
0.9749154407708353
```