```
In [44]: import pandas as pd
         import matplotlib.pyplot as plt
         from sklearn.linear_model import LinearRegression
         from sklearn.metrics import mean_squared_error
         from sklearn.model_selection import train_test_split

         df = pd.read_csv('Salary Data.csv')
         df.head(10)
```

Out[44]:

	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
5	2.9	56642.0
6	3.0	60150.0
7	3.2	54445.0
8	3.2	64445.0
9	3.7	57189.0

```
In [38]: X = df['YearsExperience'].values.reshape(-1,1)
         y = df['Salary'].values.astype(int)
```

```
In [39]: X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=35, test_size = 0.
```
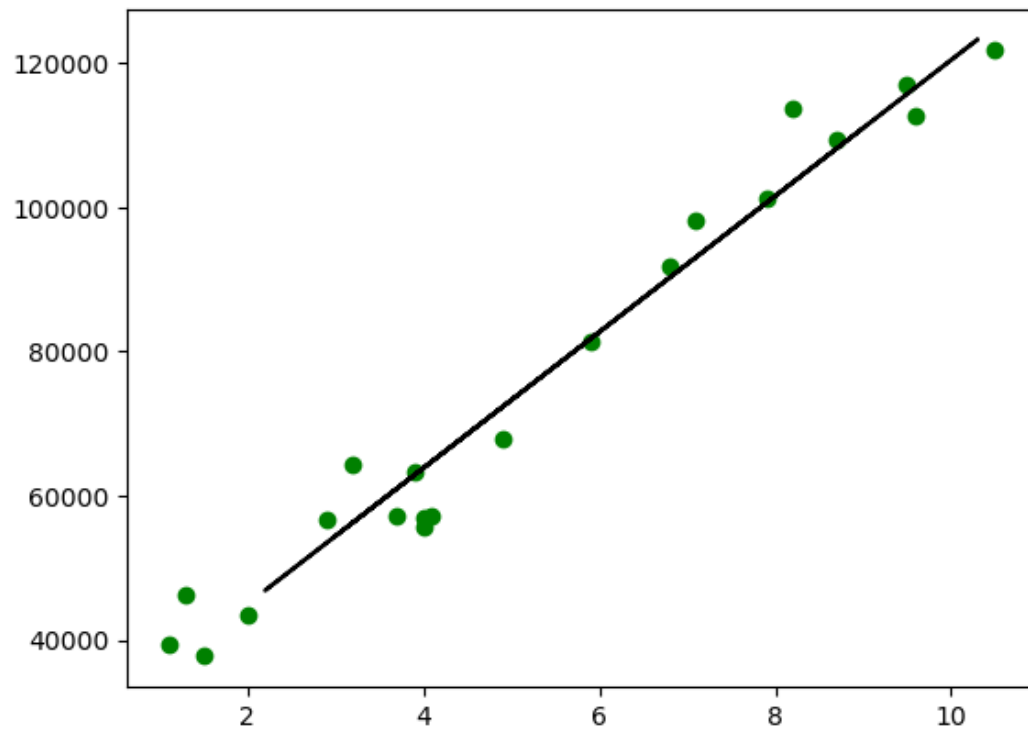
```
In [40]: model = LinearRegression()
         model.fit(X_train, y_train)

         predictions = model.predict(X_test)

         mse = mean_squared_error(y_test, predictions)
         print(f"Mean Square Error on Testing Data : {mse:.2f}")
```

```
Mean Square Error on Testing Data : 45845116.62
```

In [43]:
```python
plt.scatter(X_train, y_train, color='g')
plt.plot(X_test,predictions, color='k')
plt.show()
```


In [ ]: