



```
#11-02-23
import numpy as np
import matplotlib.pyplot as mtp
import pandas as pd

from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

data_set= pd.read_csv('/content/drive/MyDrive/Personal/Studies/MSC Data Science Material/SEM2/ML/Practical/data_set/Salary_Data.csv')
```

data_set

	YearsExperience	Salary
0	1.1	39343
1	1.3	46205
2	1.5	37731
3	2.0	43525
4	2.2	39891
5	2.9	56642
6	3.0	60150
7	3.2	54445
8	3.2	64445
9	3.7	57189
10	3.9	63218
11	4.0	55794
12	4.0	56957
13	4.1	57081
14	4.5	61111
15	4.9	67938
16	5.1	66029
17	5.3	83088
18	5.9	81363
19	6.0	93940
20	6.8	91738
21	7.1	98273
22	7.9	101302
23	8.2	113812
24	8.7	109431
25	9.0	105582
26	9.5	116969
27	9.6	112635
28	10.3	122391
29	10.5	121872

```
x=data_set.iloc[:, :-1].values
```

```
x
array([[ 1.1],
       [ 1.3],
       [ 1.5],
       [ 2. ],
       [ 2.2],
       [ 2.9],
       [ 3. ],
       [ 3.2],
```

```
[ 3.2],
[ 3.7],
[ 3.9],
[ 4. ],
[ 4. ],
[ 4.1],
[ 4.5],
[ 4.9],
[ 5.1],
[ 5.3],
[ 5.9],
[ 6. ],
[ 6.8],
[ 7.1],
[ 7.9],
[ 8.2],
[ 8.7],
[ 9. ],
[ 9.5],
[ 9.6],
[10.3],
[10.5]])
```

```
y=data_set.iloc[:, 1].values
```

```
y
```

```
array([ 39343, 46205, 37731, 43525, 39891, 56642, 60150, 54445,
        64445, 57189, 63218, 55794, 56957, 57081, 61111, 67938,
        66029, 83088, 81363, 93940, 91738, 98273, 101302, 113812,
        109431, 105582, 116969, 112635, 122391, 121872])
```

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test= train_test_split(x,y, test_size=1/3, random_state=0)
```

```
x_train
```

```
array([[ 2.9],
[ 5.1],
[ 3.2],
[ 4.5],
[ 8.2],
[ 6.8],
[ 1.3],
[10.5],
[ 3. ],
[ 2.2],
[ 5.9],
[ 6. ],
[ 3.7],
[ 3.2],
[ 9. ],
[ 2. ],
[ 1.1],
[ 7.1],
[ 4.9],
[ 4. ]])
```

```
x_test
```

```
array([[ 1.5],
[10.3],
[ 4.1],
[ 3.9],
[ 9.5],
[ 8.7],
[ 9.6],
[ 4. ],
[ 5.3],
[ 7.9]])
```

```
y_train
```

```
array([ 56642, 66029, 64445, 61111, 113812, 91738, 46205, 121872,
        60150, 39891, 81363, 93940, 57189, 54445, 105582, 43525,
        39343, 98273, 67938, 56957])
```

```
y_test
```

```
array([ 37731, 122391, 57081, 63218, 116969, 109431, 112635, 55794,
        83088, 101302])
```

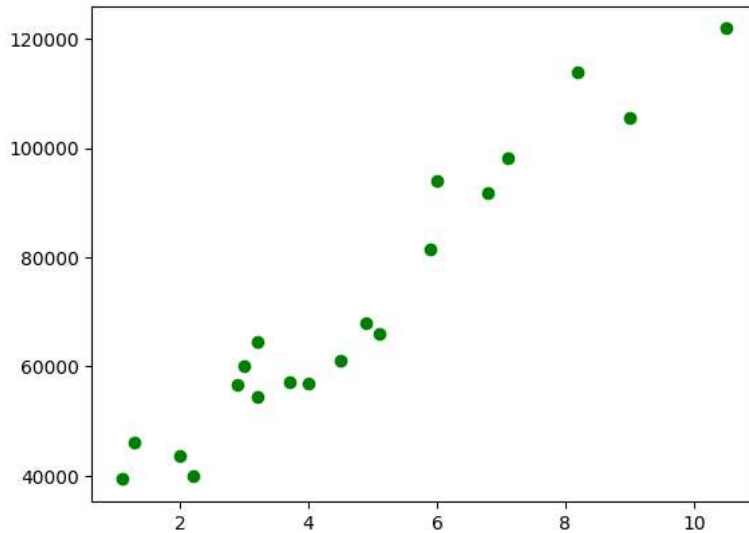
```
from sklearn.linear_model import LinearRegression
regressor=LinearRegression() #regressor is a variable that has all properties of LinearRegression alg
regressor.fit(x_train,y_train)
```

```
▼ LinearRegression
LinearRegression()
```

```
y_pred=regressor.predict(x_test)
```

```
mtp.scatter(x_train, y_train, color="green")
```

```
<matplotlib.collections.PathCollection at 0x7fed352beee0>
```



```
mtp.scatter(x_train, y_train, color="green")
mtp.plot(x_test, y_pred, color="red")
mtp.title("Salary vs Experience (training data)")
mtp.xlabel("Years of Experience")
mtp.ylabel("Salary (in rupees)")
mtp.show()
```



✓ 0s completed at 10:49 PM

● ✕