In [16]:

```python
import pandas as pd
import matplotlib.pyplot as plt
```

In [21]:

```python
df = pd.read_csv('https://raw.githubusercontent.com/AP-State-Skill-Development-Corporation/
                  Datasets/master/Regression/Salary_Data.csv')
df = pd.read_csv('Salary_Data.csv')
```

In [22]:

```python
df.head()
```

Out[22]:

	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 [7]:

```python
df.columns
```

Out[7]:

```
Index(['YearsExperience', 'Salary'], dtype='object')
```

In [8]:

```python
df.size
```

Out[8]:

```
60
```

In [19]:

```python
df.shape
```

Out[19]:

```
(40, 2)
```

```
In [11]:
```

```python
df.max(), df.min()
```

```
Out[11]:
```

```
(YearsExperience        10.5
 Salary             122391.0
 dtype: float64,
 YearsExperience         1.1
 Salary              37731.0
 dtype: float64)
```

```
In [20]:
```

```python
df.isnull().sum()
```

```
Out[20]:
```

```
YearsExperience    0
Salary             2
dtype: int64
```

```
In [24]:
```

```python
df[df.duplicated()]
```

```
Out[24]:
```

	YearsExperience	Salary
30	7.1	98273.0
31	7.9	101302.0
32	7.1	98273.0
33	7.9	101302.0
34	7.1	98273.0
35	7.9	101302.0
36	7.1	98273.0
37	7.9	101302.0

```
In [25]:
```

```python
df.drop_duplicates(inplace = True)

df.dropna(inplace = True)
```

In [26]:

```python
df.isnull().sum()
```

Out[26]:

```
YearsExperience    0
Salary             0
dtype: int64
```
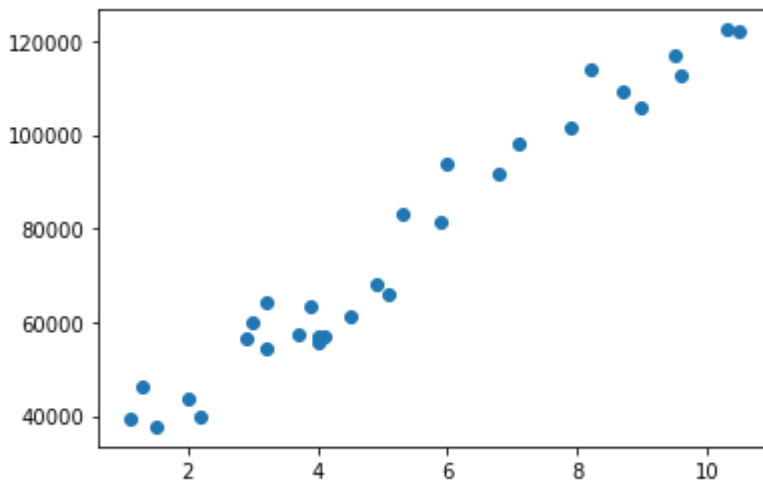
In [17]:

```python
plt.scatter(df['YearsExperience'], df['Salary'])
```

Out[17]:

```
<matplotlib.collections.PathCollection at 0x2d935097490>
```


- Numerical
- Linear
- +ve corealated

In [27]:

```python
from sklearn.linear_model import LinearRegression
```

Y = MX + C

salary = M YrsExpe + C

In [28]:

```python
lr_reg = LinearRegression()
```

```python
X = df['YearsExperience'].values.reshape(-1, 1)
y = df['Salary']
```

```python
lr_reg.fit(X, y)
```

```
LinearRegression()
```
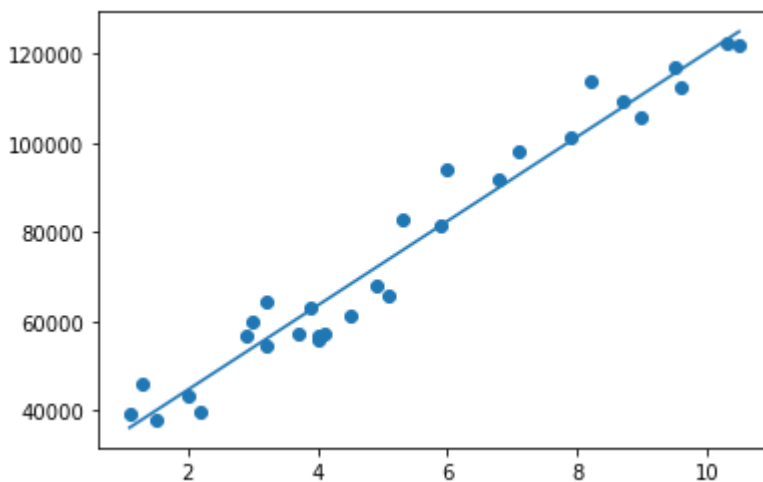
```python
lr_reg.predict([[11.1]])
```

```
array([130686.78196682])
```

```python
y_pred = lr_reg.predict(X)
plt.scatter(df['YearsExperience'], df['Salary'])
plt.plot(df['YearsExperience'], y_pred)
```

```
[<matplotlib.lines.Line2D at 0x2d943988250>]
```

```python
lr_reg.coef_
```

```
array([9449.96232146])
```

```
lr_reg.intercept_
```

```
25792.20019866871
```

Y = 25792.20019866871 * X + 9449.96232146

```
lr_reg.score(X, y)
```

```
0.9569566641435086
```