3.ASSIGNMENT 3 ON SIMPLE REGRESSION

```
In [2]:  import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
         from sklearn.model_selection import train_test_split
         from sklearn.linear_model import LinearRegression
         from sklearn.metrics import mean_squared_error,r2_score
```

Load salary data

```
In [7]:  df=pd.read_csv('Salary_Dataset.csv')
```

Display first 5 rows of dataframe

```
In [8]:  display(df.head())
```

	Unnamed: 0	YearsExperience	Salary
0	0	1.2	39344.0
1	1	1.4	46206.0
2	2	1.6	37732.0
3	3	2.1	43526.0
4	4	2.3	39892.0

Display Dataset information

```
In [9]:  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 3 columns):
Unnamed: 0        30 non-null int64
YearsExperience   30 non-null float64
Salary            30 non-null float64
dtypes: float64(2), int64(1)
memory usage: 800.0 bytes
```

statistics summary of dataset

In [10]: `df.describe()`

Out[10]:

	Unnamed: 0	YearsExperience	Salary
count	30.000000	30.000000	30.000000
mean	14.500000	5.413333	76004.000000
std	8.803408	2.837888	27414.429785
min	0.000000	1.200000	37732.000000
25%	7.250000	3.300000	56721.750000
50%	14.500000	4.800000	65238.000000
75%	21.750000	7.800000	100545.750000
max	29.000000	10.600000	122392.000000

Data visualization

In [12]:
```
plt.scatter(x=df['YearsExperience'],y=df['Salary'])
plt.xlabel("YearsExperience")
plt.ylabel("Salary")
plt.title("Salary vs Experience")
plt.show()
```


prparing x and y

In [13]:
```
x=df[['YearsExperience']]
y=df['Salary']
```

splitting data into traning and test sets

```
In [14]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=
         42)
```

Model Creation

```
In [15]: model=LinearRegression()
```

Model Training

```
In [16]: model.fit(x_train,y_train)
```

```
Out[16]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

Make predictions(Testing)

```
In [17]: y_pred=model.predict(x_test)
```

## Model evaluation Parameters

```
In [18]: mse=mean_squared_error(y_test,y_pred)
         r2=r2_score(y_test,y_pred)
         print(f"Mean Squared Error: {mse}")
         print(f"R-Squared score: {r2}")
```

```
Mean Squared Error: 49830096.855908334
R-Squared score: 0.9024461774180498
```

plot Regression line

```
In [19]: plt.scatter(x_test,y_test,color='blue',label='Actual')
         plt.plot(x_test,y_pred,color='red',linewidth=2,label='Predicted')
         plt.xlabel("'YearsExperience")
         plt.ylabel("Salary")
         plt.title("Linear Regression:Salary vs Experience")
         plt.legend()
         plt.show()
```
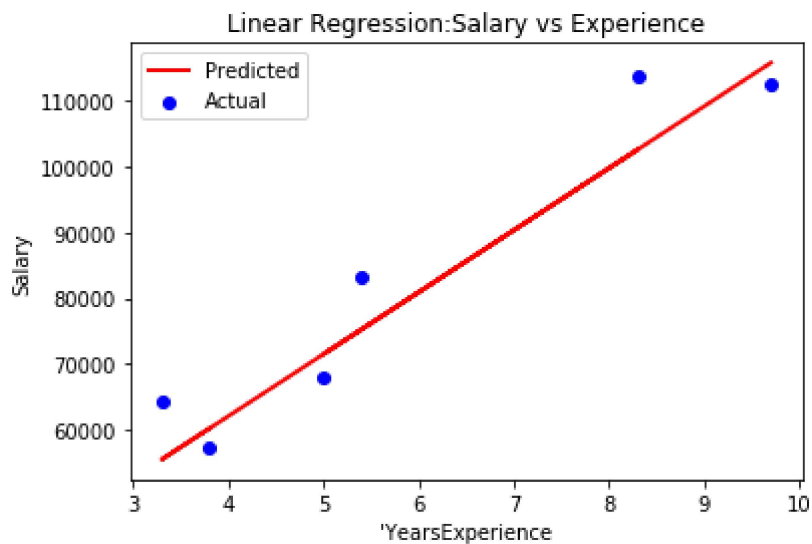

Display model coefficients

```
In [20]: print(f"Intercept:{model.intercept_}")
         print(f"coefficient:{model.coef_[0]}")
```

```
Intercept:24380.20147947369
coefficient:9423.81532303098
```