

```
In [4]: import pandas as pd
        from sklearn.linear_model import LinearRegression
        import numpy as np
```

```
In [2]: df=pd.read_csv("f:/dataset/regression/salary_1_variable.csv")
        X=df.iloc[:, :-1].values
        y=df.iloc[:, -1].values
```

```
In [3]: model=LinearRegression()
        model.fit(X,y)
        print(model.coef_)
        print(model.intercept_)
```

```
[9449.96232146]
25792.20019866871
```

```
In [28]: xmean=np.mean(X)
        ymean=np.mean(y)
        xymean=np.mean(X[:,0]*y)
        xmeansqr=xmean**2
        xsqrmean=np.mean(X**2)
```

```
In [29]: m=((xmean*ymean)-xymean)/(xmeansqr-xsqrmean)
        print(m)
```

```
9449.962321455081
```

```
In [31]: c=ymean-m*xmean
        print(c)
```

```
25792.200198668666
```

```
In [32]: model.predict([[20]])
```

```
Out[32]: array([214791.44662777])
```

```
In [33]: pred=m*20+c
        print(pred)
```

```
214791.4466277703
```

```
In [34]: pred=model.predict(X)
```

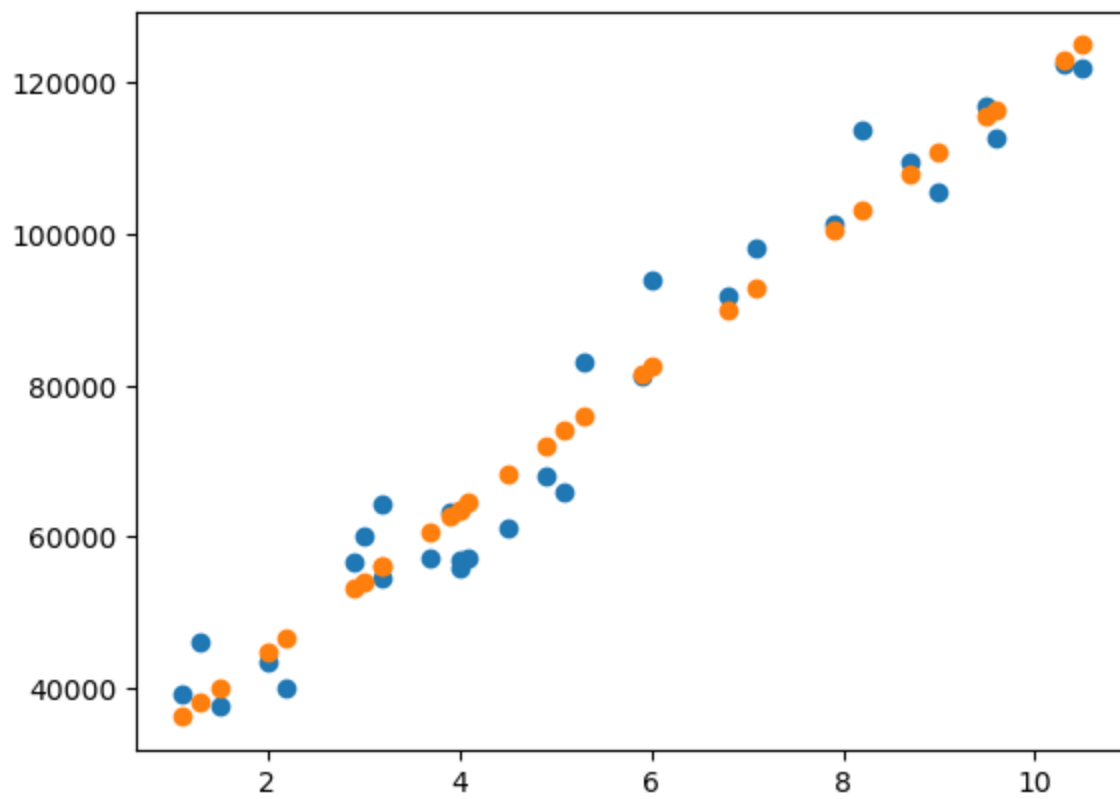
```
In [37]: from sklearn.metrics import r2_score
```

```
In [38]: r2_score(y,pred)
```

```
Out[38]: 0.9569566641435086
```

```
In [42]: import matplotlib.pyplot as plt
```

```
In [50]: plt.scatter(X[:,0],y)
        plt.scatter(X[:,0],pred)
        plt.show()
```



```
In [52]: df['pred']=pred
```

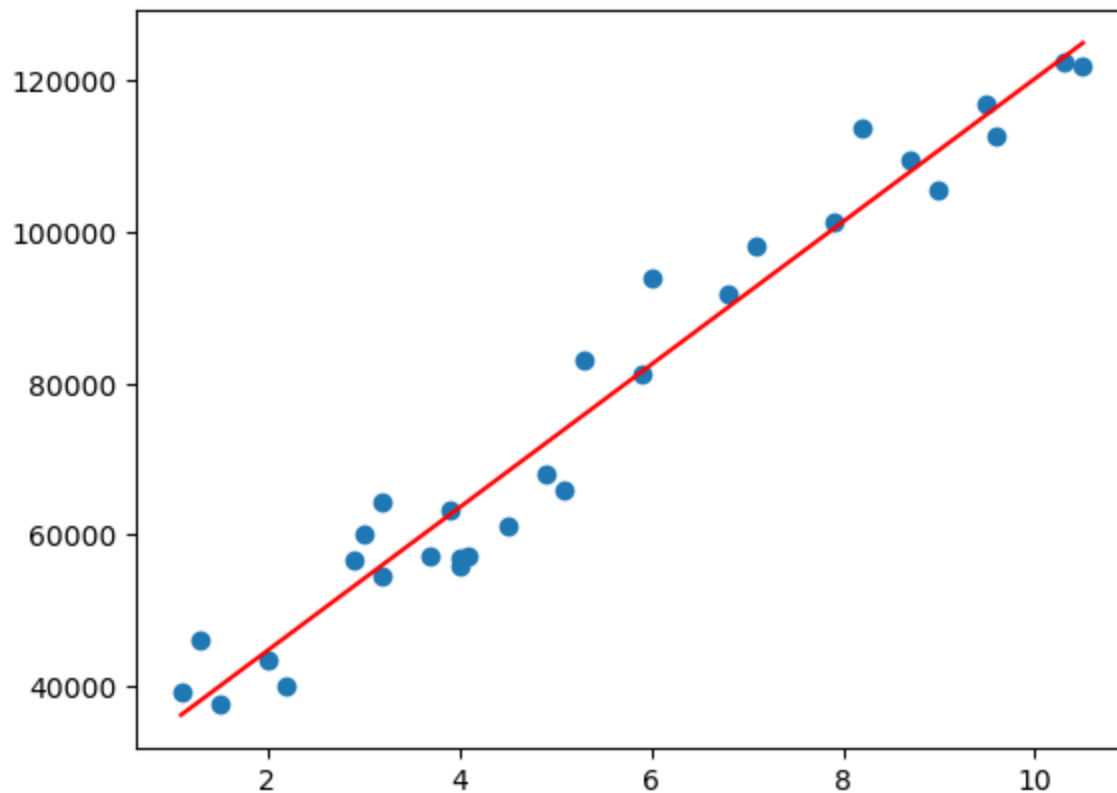
```
In [53]: df
```

```
Out[53]:
```

	YearsExperience	Salary	pred
0	1.1	39343	36187.158752
1	1.3	46205	38077.151217
2	1.5	37731	39967.143681
3	2.0	43525	44692.124842
4	2.2	39891	46582.117306
5	2.9	56642	53197.090931
6	3.0	60150	54142.087163
7	3.2	54445	56032.079627
8	3.2	64445	56032.079627
9	3.7	57189	60757.060788
10	3.9	63218	62647.053252
11	4.0	55794	63592.049484
12	4.0	56957	63592.049484
13	4.1	57081	64537.045717
14	4.5	61111	68317.030645
15	4.9	67938	72097.015574
16	5.1	66029	73987.008038
17	5.3	83088	75877.000502
18	5.9	81363	81546.977895

19	6.0	93940	82491.974127
20	6.8	91738	90051.943985
21	7.1	98273	92886.932681
22	7.9	101302	100446.902538
23	8.2	113812	103281.891235
24	8.7	109431	108006.872395
25	9.0	105582	110841.861092
26	9.5	116969	115566.842252
27	9.6	112635	116511.838485
28	10.3	122391	123126.812110
29	10.5	121872	125016.804574

```
In [56]: plt.scatter(X[:,0],y)
plt.plot(X[:,0],pred,c='r')
plt.show()
```



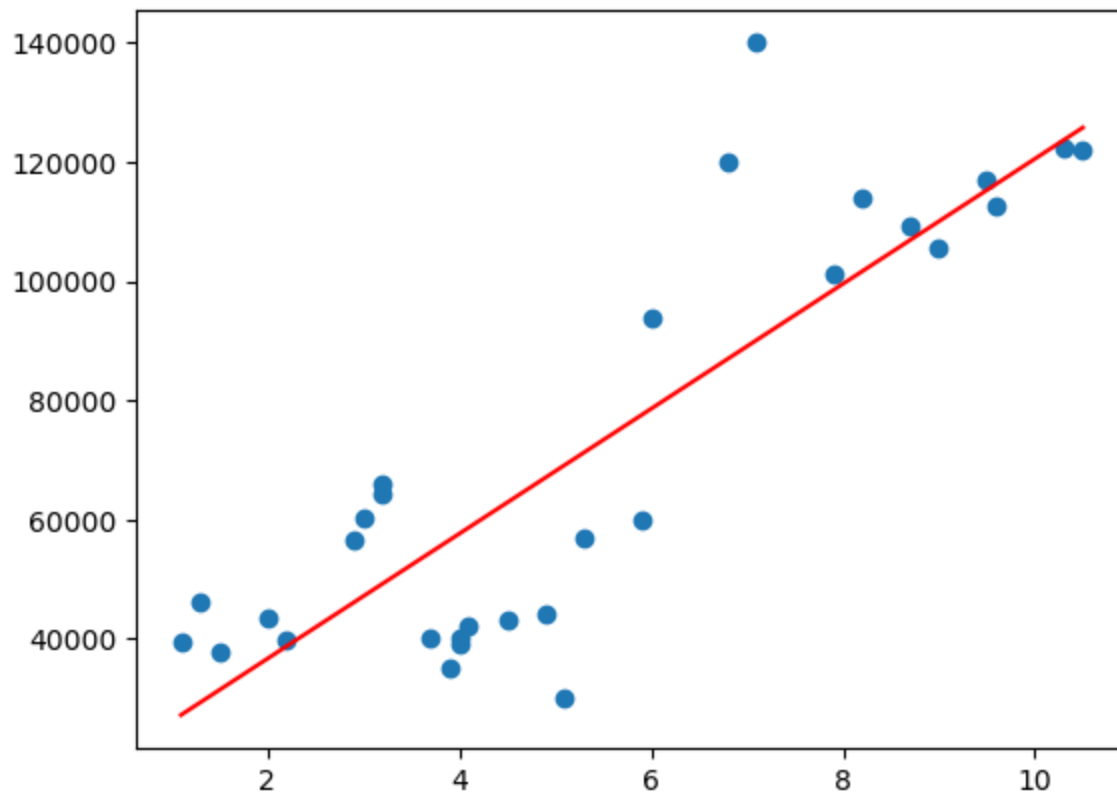
```
In [57]: r2_score(y,pred)
```

```
Out[57]: 0.9569566641435086
```

```
In [58]: df=pd.read_csv("f:/dataset/regression/salary_poly.csv")
X=df.iloc[:, :-1].values
y=df.iloc[:, -1].values
model=LinearRegression()
model.fit(X,y)
pred=model.predict(X)

plt.scatter(X[:,0],y)
```

```
plt.plot(X[:,0],pred,c='r')
plt.show()
```



```
In [59]: r2_score(y,pred)
```

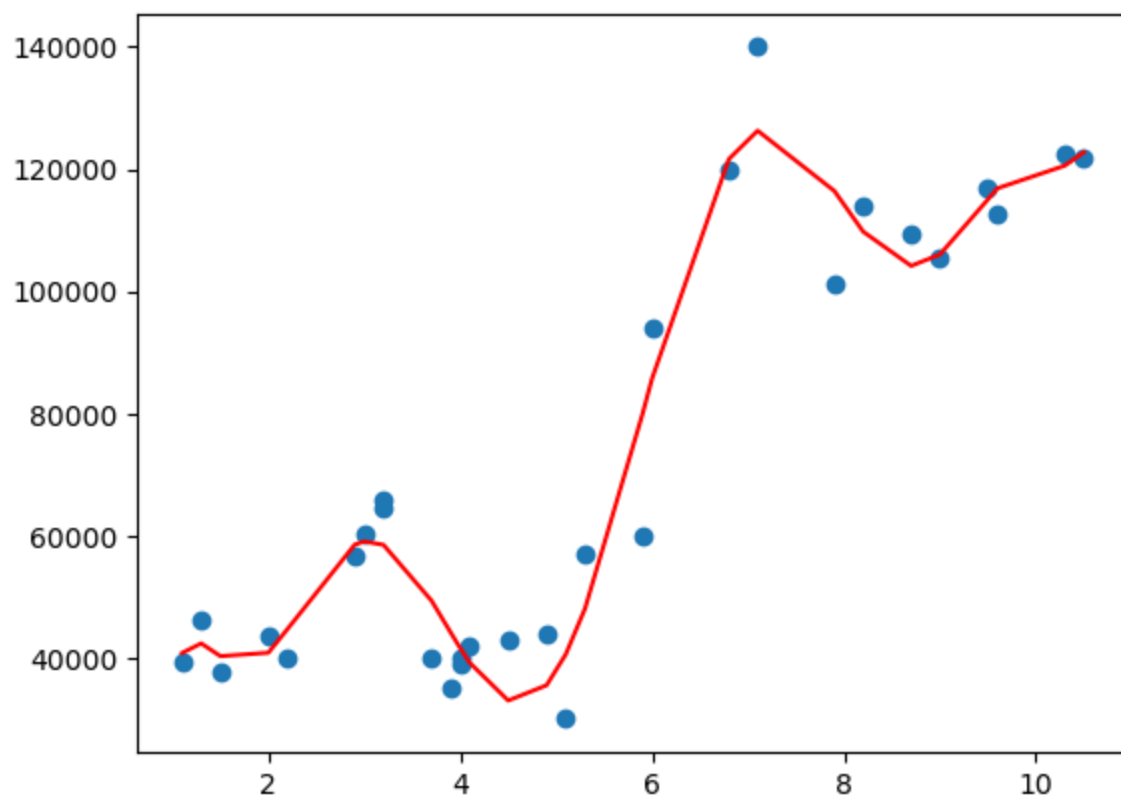
```
Out[59]: 0.7213773452799213
```

```
In [60]: from sklearn.preprocessing import PolynomialFeatures
```

```
In [81]: pf=PolynomialFeatures(degree=9)
X_new=pf.fit_transform(X)

model=LinearRegression()
model.fit(X_new,y)
pred=model.predict(X_new)

plt.scatter(X[:,0],y)
plt.plot(X[:,0],pred,c='r')
plt.show()
print(r2_score(y,pred))
```



0.953491624488293

In []: