In [2]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
%matplotlib inline
```

In [3]:

```
data = pd.read_csv('Salary_Data.csv')
data
```

Out[3]:

	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
10	3.9	63218.0
11	4.0	55794.0
12	4.0	56957.0
13	4.1	57081.0
14	4.5	61111.0
15	4.9	67938.0
16	5.1	66029.0
17	5.3	83088.0
18	5.9	81363.0
19	6.0	93940.0
20	6.8	91738.0
21	7.1	98273.0
22	7.9	101302.0
23	8.2	113812.0
24	8.7	109431.0
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0

	YearsExperience	Salary
28	10.3	122391.0
29	10.5	121872.0

In [4]:

```
data.head(7)
```

Out[4]:

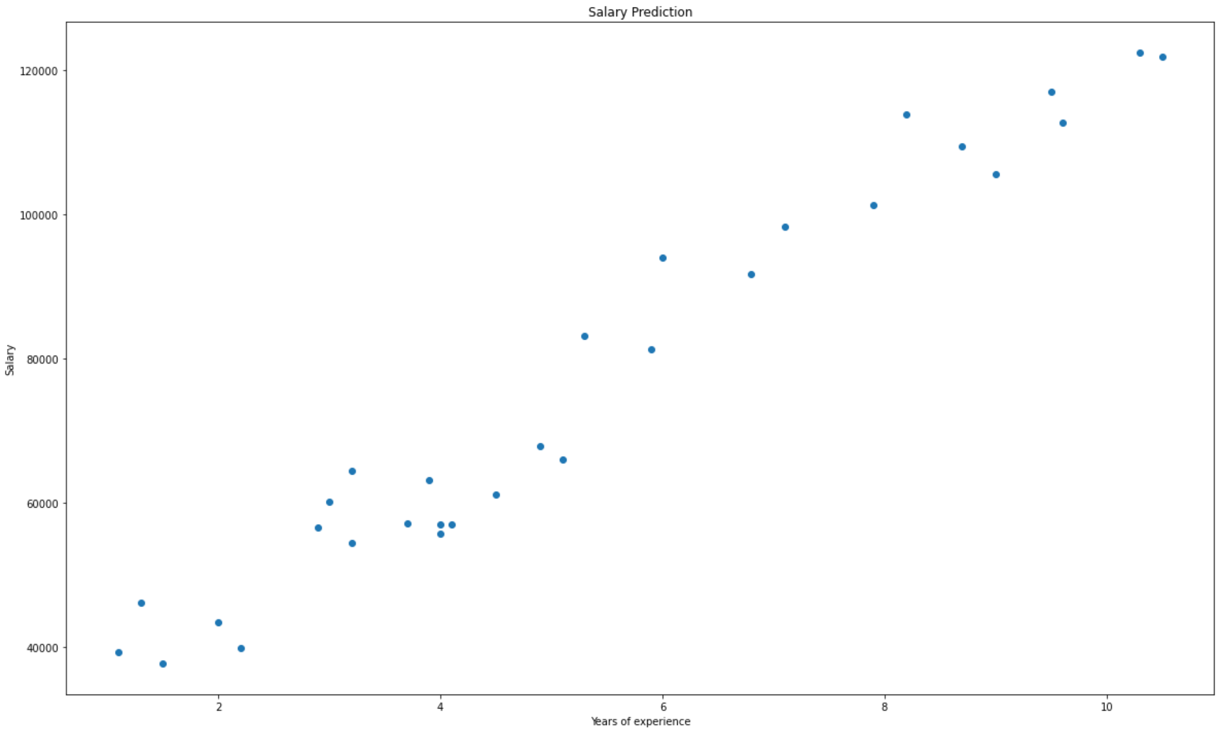
	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0
5	2.9	56642.0
6	3.0	60150.0

In [5]:

```
x = data['YearsExperience']
y = data['Salary']
```

In [6]:

```
plt.figure(figsize=(20,12))                #to set the size of the graph (optional)
plt.scatter(x,y)                           #built-in function in matplotlib to visu
plt.title("Salary Prediction")
plt.xlabel("Years of experience")
plt.ylabel("Salary")
plt.show()
```



In [7]:

```
regression_model = LinearRegression()
```

In [8]:
```
#reshapes the arrays to fit in the graph accordingly
x = np.array(x).reshape((len(x), 1))
#to minimize the residual sum of squares between the observed targets in the dataset
#and the targets predicted by the linear approximation
regression_model.fit(x, y)
```

Out[8]:
```
LinearRegression()
```

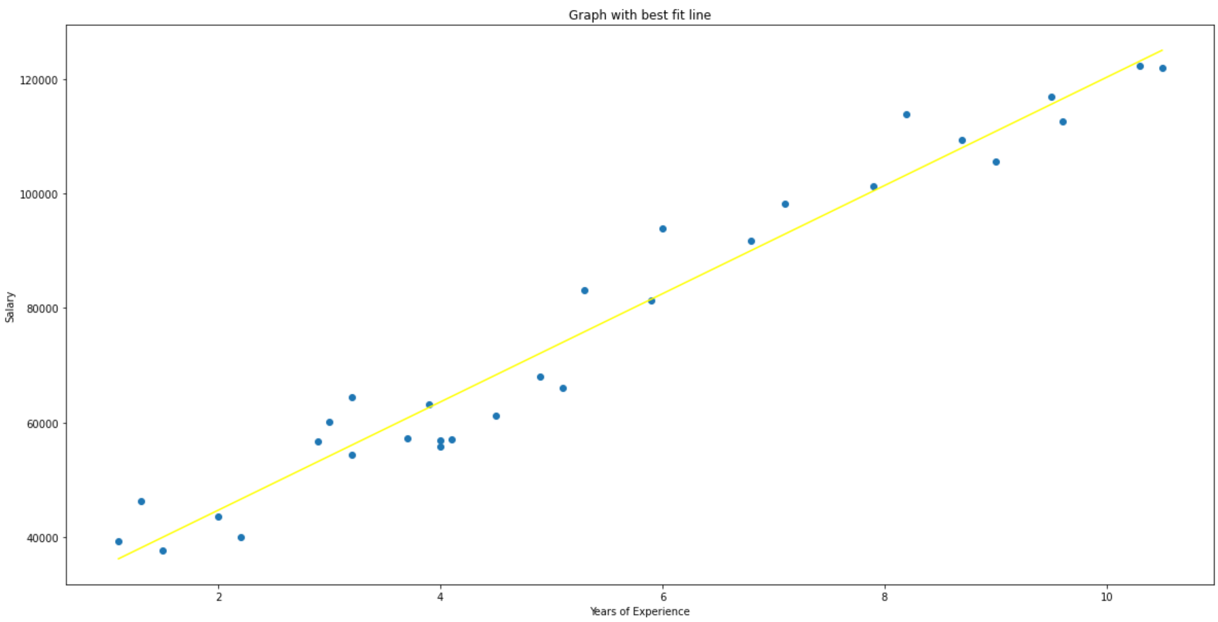
In [9]:
```
y_predicted = regression_model.predict(x)
y_predicted
```

Out[9]:
```
array([ 36187.15875227,  38077.15121656,  39967.14368085,  44692.12484158,
        46582.11730587,  53197.09093089,  54142.08716303,  56032.07962732,
        56032.07962732,  60757.06078805,  62647.05325234,  63592.04948449,
        63592.04948449,  64537.04571663,  68317.03064522,  72097.0155738 ,
        73987.00803809,  75877.00050238,  81546.97789525,  82491.9741274 ,
        90051.94398456,  92886.932681  , 100446.90253816, 103281.8912346 ,
       108006.87239533, 110841.86109176, 115566.84225249, 116511.83848464,
       123126.81210966, 125016.80457395])
```

In [10]:
```
mse = mean_squared_error(y, y_predicted)
r2 = r2_score(y, y_predicted)
print('Slope: ',regression_model.coef_)
print('Intercept: ', regression_model.intercept_)
print('Root mean squared error: ', mse)
print('R2 score: ',r2)
```

```
Slope:  [9449.96232146]
Intercept:  25792.20019866871
Root mean squared error:  31270951.722280968
R2 score:  0.9569566641435086
```

In [11]:
```
plt.figure(figsize=(20,10))
plt.scatter(x,y)
plt.plot(x, y_predicted, color='yellow')
plt.title('Graph with best fit line')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```



In [ ]: