In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from word2number import w2n
df = pd.read_csv("hiring.csv")
df
```

Out[1]:

| | experience | test_score(out of 10) | interview_score(out of 10) | salary($) |
|---|---|---|---|---|
| **0** | NaN | 8.0 | 9 | 50000 |
| **1** | NaN | 8.0 | 6 | 45000 |
| **2** | five | 6.0 | 7 | 60000 |
| **3** | two | 10.0 | 10 | 65000 |
| **4** | seven | 9.0 | 6 | 70000 |
| **5** | three | 7.0 | 10 | 62000 |
| **6** | ten | NaN | 7 | 72000 |
| **7** | eleven | 7.0 | 8 | 80000 |

In [2]:
```
#Assuming the experience be zero instead of NAN
df["experience"].fillna("zero",inplace=True)
```

In [3]:
```
df["experience"] = df["experience"].apply(w2n.word_to_num)    #Convert
 the words to number
df
```

Out[3]:

| | experience | test_score(out of 10) | interview_score(out of 10) | salary($) |
|---|---|---|---|---|
| **0** | 0 | 8.0 | 9 | 50000 |
| **1** | 0 | 8.0 | 6 | 45000 |

| | experience | test_score(out of 10) | interview_score(out of 10) | salary($) |
|---|---|---|---|---|
| **2** | 5 | 6.0 | 7 | 60000 |
| **3** | 2 | 10.0 | 10 | 65000 |
| **4** | 7 | 9.0 | 6 | 70000 |
| **5** | 3 | 7.0 | 10 | 62000 |
| **6** | 10 | NaN | 7 | 72000 |
| **7** | 11 | 7.0 | 8 | 80000 |

In [4]:
```
df["test_score(out of 10)"].fillna(df["test_score(out of 10)"].mean(),i
nplace=True)
df
```

Out[4]:

| | experience | test_score(out of 10) | interview_score(out of 10) | salary($) |
|---|---|---|---|---|
| **0** | 0 | 8.000000 | 9 | 50000 |
| **1** | 0 | 8.000000 | 6 | 45000 |
| **2** | 5 | 6.000000 | 7 | 60000 |
| **3** | 2 | 10.000000 | 10 | 65000 |
| **4** | 7 | 9.000000 | 6 | 70000 |
| **5** | 3 | 7.000000 | 10 | 62000 |
| **6** | 10 | 7.857143 | 7 | 72000 |
| **7** | 11 | 7.000000 | 8 | 80000 |

In [5]:
```
x = df.drop("salary($)",axis=1)
y = df["salary($)"]
riss = LinearRegression()
riss.fit(x,y)
sal_pred = riss.predict(x)
```

In [9]:
```
a = riss.intercept_
print("The intercept is",a)
```

```
The intercept is 17237.330313727172
```

In [10]:
```
b = riss.coef_
print("The coeffs are",b)
```

```
The coeffs are [2827.63404314 1912.93803053 2196.9753141 ]
```

In [11]:
```
# Prediction for a candidate with 2 years of with 9 marks in test score
 and 6 marks in interview score
riss.predict([[2,9,6]])
```

Out[11]:
```
array([53290.89255945])
```

In [12]:
```
# Verification of calaculated salary
sal = 17237.330313727172 + 2827.63404314 * 2 + 1912.93803053*9 + 2196.9
753141*6
print(sal)
```

```
53290.89255937718
```

In [ ]:
```

```