Practical - 13

AIM

Based on multiple features/variables perform Linear Regression.

PROBLEM

Based on the experience, test score and interview score of an employee predict the salary of the employee.

CODE & OUTPUT

In [1]:
```
import numpy as np
import pandas as pd
from sklearn import linear_model
```

In [2]:
```
df = pd.read_csv("salaries.csv")
print(df)
```

```
  experience  test_score  interview_score  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 [3]:
```
df.experience = df.experience.map({'two': 2, 'three': 3, 'five': 5, 'seven': 7, 'ten': 10,
print(df)
```

```
   experience  test_score  interview_score  salary
0         NaN         8.0                9   50000
1         NaN         8.0                6   45000
2         5.0         6.0                7   60000
3         2.0        10.0               10   65000
4         7.0         9.0                6   70000
5         3.0         7.0               10   62000
6        10.0         NaN                7   72000
7        11.0         7.0                8   80000
```

In [4]:
```
df.experience = df.experience.fillna(0)
print(df)
```

```
   experience  test_score  interview_score  salary
0         0.0         8.0                9   50000
1         0.0         8.0                6   45000
2         5.0         6.0                7   60000
3         2.0        10.0               10   65000
4         7.0         9.0                6   70000
5         3.0         7.0               10   62000
6        10.0         NaN                7   72000
7        11.0         7.0                8   80000
```

In [6]:
```
df.test_score = df.test_score.fillna(df.test_score.median())
print(df)
```

```
   experience  test_score  interview_score  salary
0         0.0         8.0                9   50000
1         0.0         8.0                6   45000
2         5.0         6.0                7   60000
3         2.0        10.0               10   65000
4         7.0         9.0                6   70000
5         3.0         7.0               10   62000
6        10.0         8.0                7   72000
7        11.0         7.0                8   80000
```

In [7]:
```
model = linear_model.LinearRegression()
model.fit(df.drop("salary", axis="columns"), df.salary)
```

Out[7]:
```
LinearRegression()
```

In [8]:
```
model.predict([[2, 9, 6]])
```

Out[8]:
```
array([53205.96797671])
```

In [9]:
```
model.predict([[12, 10, 10]])
```

Out[9]:
```
array([92002.18340611])
```

In [10]:
```
model.intercept_
```

Out[10]:
```
17737.263464337688
```

In [11]:
```
model.coef_
```

Out[11]:
```
array([2812.95487627, 1845.70596798, 2205.24017467])
```

In [12]:
```
2*2812.95487627+9*1845.70596798+6*2205.24017467+17737.263464337688
```

Out[12]:
```
53205.96797671769
```

In [13]:
```
12*2812.95487627+10*1845.70596798+10*2205.24017467+17737.263464337688
```

Out[13]:
```
92002.18340607767
```