Simple Linear Regression

Aim: Simple Linear Regression

Experiment no.: 5

In [1]:
```
#Name: Prapti Pramod Ugale
#Roll no.: 73
#Sec: A
#Subject: Data Science and Statistics (Lab 1)
#Date: 25/07/2023
```

In [2]:
```
import pandas as pd
```

In [3]:
```
from matplotlib import pyplot as plt
```

In [4]:
```
import numpy as np
```

In [5]:
```
import os
```

In [6]:
```
os.getcwd()
```

Out[6]:
```
'C:\\Users\\hp\\Desktop\\DSS Practicals'
```

In [7]:
```
os.chdir('C:\\Users\\HP\\Desktop')
```

In [8]:
```
df=pd.read_csv("Salary_dataset.csv")
```

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

Out[9]:

	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

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

Out[10]:

	Unnamed: 0	YearsExperience	Salary
25	25	9.1	105583.0
26	26	9.6	116970.0
27	27	9.7	112636.0
28	28	10.4	122392.0
29	29	10.6	121873.0

In [11]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   Unnamed: 0       30 non-null     int64  
 1   YearsExperience  30 non-null     float64
 2   Salary           30 non-null     float64
dtypes: float64(2), int64(1)
memory usage: 852.0 bytes
```

In [12]:
```
df.describe()
```

Out[12]:

	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

In [13]:
```
df.shape
```

Out[13]:
```
(30, 3)
```

In [14]:
```
df.size
```

Out[14]:
```
90
```

In [15]:
```
df.ndim
```

Out[15]:
```
2
```

In [16]:
```
df.isnull().sum()
```

Out[16]:
```
Unnamed: 0          0
YearsExperience     0
Salary              0
dtype: int64
```

In [17]:
```
df.head()
```

Out[17]:

	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

In [18]:
```
df.columns
```

Out[18]:
```
Index(['Unnamed: 0', 'YearsExperience', 'Salary'], dtype='object')
```

In [19]:
```
a=(1,2,3,4,5,6,7,8,9,10)
```

In [20]:
```
a[0]
```

Out[20]:
```
1
```

In [21]:
```
a[-1]
```

Out[21]:
```
10
```

In [22]:
```
a[9]
```

Out[22]:
```
10
```

In [23]:
```
a[-10]
```

Out[23]:
```
1
```

In [24]:
```
df.loc[4,'Salary']
```

Out[24]:
```
39892.0
```

In [25]:
```
df.head()
```

Out[25]:

	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

In [26]:
```
df.loc[29]
```

Out[26]:
```
Unnamed: 0             29.0
YearsExperience        10.6
Salary             121873.0
Name: 29, dtype: float64
```

In [27]:
```
df.head(30)
```

Out[27]:

	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
5	5	3.0	56643.0
6	6	3.1	60151.0
7	7	3.3	54446.0
8	8	3.3	64446.0
9	9	3.8	57190.0
10	10	4.0	63219.0
11	11	4.1	55795.0
12	12	4.1	56958.0
13	13	4.2	57082.0
14	14	4.6	61112.0
15	15	5.0	67939.0
16	16	5.2	66030.0
17	17	5.4	83089.0
18	18	6.0	81364.0
19	19	6.1	93941.0
20	20	6.9	91739.0
21	21	7.2	98274.0
22	22	8.0	101303.0
23	23	8.3	113813.0
24	24	8.8	109432.0
25	25	9.1	105583.0
26	26	9.6	116970.0
27	27	9.7	112636.0
28	28	10.4	122392.0
29	29	10.6	121873.0

In [28]:
```
df.loc[4]
```

Out[28]:
```
Unnamed: 0             4.0
YearsExperience        2.3
Salary             39892.0
Name: 4, dtype: float64
```

In [29]:
```
a=(1,2,3,4,5,6,7,8,9,10)
```

In [30]:
```
a[1:4]
```

Out[30]:
```
(2, 3, 4)
```

In [31]:
```
df.iloc[1,2]
```

Out[31]:
```
46206.0
```

In [32]:
```
df.head()
```

Out[32]:

	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

In [33]:
```
df.loc[1,'Salary']
```

Out[33]:
```
46206.0
```

In [34]:
```
x=df.iloc[:,:-1].values
```

In [35]:
```
y=df.iloc[:,-1].values
```

In [36]:
```
print(x)
```

```
[[ 0.   1.2]
 [ 1.   1.4]
 [ 2.   1.6]
 [ 3.   2.1]
 [ 4.   2.3]
 [ 5.   3. ]
 [ 6.   3.1]
 [ 7.   3.3]
 [ 8.   3.3]
 [ 9.   3.8]
 [10.   4. ]
 [11.   4.1]
 [12.   4.1]
 [13.   4.2]
 [14.   4.6]
 [15.   5. ]
 [16.   5.2]
 [17.   5.4]
 [18.   6. ]
 [19.   6.1]
 [20.   6.9]
 [21.   7.2]
 [22.   8. ]
 [23.   8.3]
 [24.   8.8]
 [25.   9.1]
 [26.   9.6]
 [27.   9.7]
 [28.  10.4]
 [29.  10.6]]
```

In [37]:
```
print(y)
```

```
[ 39344.  46206.  37732.  43526.  39892.  56643.  60151.  54446.  64446.
  57190.  63219.  55795.  56958.  57082.  61112.  67939.  66030.  83089.
  81364.  93941.  91739.  98274. 101303. 113813. 109432. 105583. 116970.
 112636. 122392. 121873.]
```

In [38]:
```
a=(1,2,3,4,5,6,7,8,9,10)
```

In [39]:
```
a[:2]
```

Out[39]: (1, 2)

In [40]:
```
a[2:]
```

Out[40]: (3, 4, 5, 6, 7, 8, 9, 10)

In [41]:
```
a[1:6:2]
```

Out[41]: (2, 4, 6)

In [42]:
```
a[1:6:1]
```

Out[42]: (2, 3, 4, 5, 6)

In [43]:
```
print(x)
```

```
[[ 0.   1.2]
 [ 1.   1.4]
 [ 2.   1.6]
 [ 3.   2.1]
 [ 4.   2.3]
 [ 5.   3. ]
 [ 6.   3.1]
 [ 7.   3.3]
 [ 8.   3.3]
 [ 9.   3.8]
 [10.   4. ]
 [11.   4.1]
 [12.   4.1]
 [13.   4.2]
 [14.   4.6]
 [15.   5. ]
 [16.   5.2]
 [17.   5.4]
 [18.   6. ]
 [19.   6.1]
 [20.   6.9]
 [21.   7.2]
 [22.   8. ]
 [23.   8.3]
 [24.   8.8]
 [25.   9.1]
 [26.   9.6]
 [27.   9.7]
 [28.  10.4]
 [29.  10.6]]
```

In [44]:
```
print(y)
```

```
[ 39344.  46206.  37732.  43526.  39892.  56643.  60151.  54446.  64446.
  57190.  63219.  55795.  56958.  57082.  61112.  67939.  66030.  83089.
  81364.  93941.  91739.  98274. 101303. 113813. 109432. 105583. 116970.
 112636. 122392. 121873.]
```