Nandini Bajaj 18CY20020

Lab Assignment 3

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.metrics import r2_score as r2
```

```
df = pd.read_csv("/content/Dataset-Lab3.xlsx - Sheet1 (1).csv")
df
```

	Compounds	w (Xi)	logP (Xi)	pLC50 (expt.) (Yi)
0	1	3.68	5.70	6.38
1	2	3.02	4.75	5.85
2	3	2.68	4.16	5.00
3	4	2.63	4.04	4.89
4	5	2.33	3.57	4.56
5	6	2.36	3.57	4.30
6	7	2.27	3.45	4.40
7	8	1.95	2.56	3.77
8	9	1.92	3.01	3.89
9	10	1.87	3.35	4.33
10	11	1.40	3.14	4.21
11	12	1.44	3.14	3.82
12	13	1.42	3.09	3.48
13	14	1.53	2.64	3.32
14	15	1.60	2.14	3.40

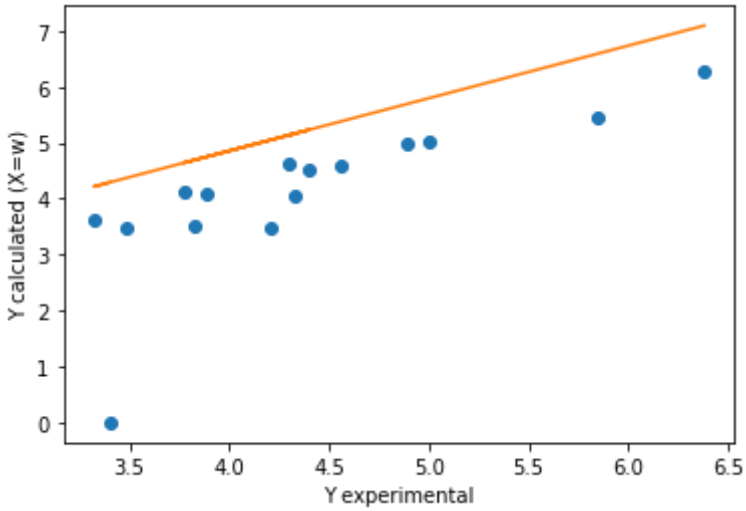
```
arr_x =  df['w (Xi)']
arr_y =  df['pLC50 (expt.) (Yi)']
y1 = np.zeros(15)
cpx=[]
cpy=[]
for i in range(14):
  cpx = list(arr_x)
  cpy = list(arr_y)

  np.delete(cpx, i)
  np.delete(cpy, i)

```

```
  m, b = np.polyfit(cpx, cpy, 1)
  y1[i] = b + (m*arr_x[i])
```

```
plt.plot(arr_y, y1, 'o')
plt.plot(arr_y, m*arr_y + b)
plt.xlabel("Y experimental")
plt.ylabel("Y calculated (X=w)")
one = r2(y1,arr_y)
```




```
one
```

```
0.5235898331343869
```

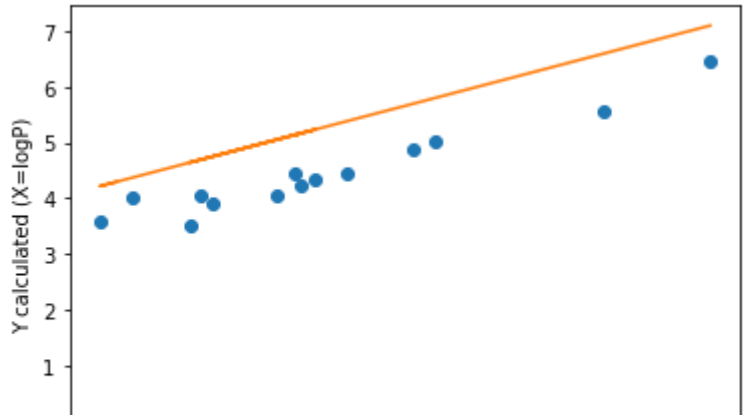
```
arr_x =  df['logP (Xi)']
arr_y =  df['pLC50 (expt.) (Yi)']
y2 = np.zeros(15)
cpx=[]
cpy=[]
for i in range(14):
  cpx = list(arr_x)
  cpy = list(arr_y)

  np.delete(cpx, i)
  np.delete(cpy, i)

  m, b = np.polyfit(cpx, cpy, 1)
  y2[i] = b + (m*arr_x[i])

```

```
plt.plot(arr_y, y2, 'o')
plt.plot(arr_y, m*arr_y + b)
plt.xlabel("Y experimental")
plt.ylabel("Y calculated (X=logP)")
two = r2(y2, arr_y)
```




```
1 two
```

```
0.5465214787794805
```

```
1 if(one > two):
2   print("The descriptor w works better than logP")
3 else:
4   print("The descriptor logP works better than w")
```

```
The descriptor logP works better than w
```

```
1 y1
```

```
array([6.26357439, 5.45347108, 5.03614513, 4.97477367, 4.60654489,
       4.64336777, 4.53289914, 4.14012177, 4.1032989 , 4.04192743,
       3.46503568, 3.51413285, 3.48958427, 3.62460149])
```

```
1 y2
```

```
array([6.45321604, 5.56022633, 5.00563272, 4.89283401, 4.4510391 ,
       4.4510391 , 4.3382404 , 3.50165004, 3.92464517, 4.24424149,
       4.04684376, 4.04684376, 3.9998443 , 3.57684918])
```

```
1   df['Y calc(with X=w)'] = y1
2   df['Y calc(with X=logP)'] = y2
```

```
1   df
2
```

	Compounds	w (Xi)	logP (Xi)	pLC50 (expt.) (Yi)	Y calc(with X=w)	Y calc(with X=logP)
0	1	3.68	5.70	6.38	6.263574	6.453216
1	2	3.02	4.75	5.85	5.453471	5.560226
2	3	2.68	4.16	5.00	5.036145	5.005633
3	4	2.63	4.04	4.89	4.974774	4.892834
4	5	2.33	3.57	4.56	4.606545	4.451039
5	6	2.36	3.57	4.30	4.643368	4.451039
6	7	2.27	3.45	4.40	4.532899	4.338240
7	8	1.95	2.56	3.77	4.140122	3.501650

1

9	10	1.87	3.35	4.33	4.041927	4.244241
10	11	1.40	3.14	4.21	3.465036	4.046844
11	12	1.44	3.14	3.82	3.514133	4.046844
12	13	1.42	3.09	3.48	3.489584	3.999844
13	14	1.53	2.64	3.32	3.624601	3.576849
14	15	1.60	2.14	3.40	0.000000	0.000000

✓ 0s completed at 4:19 PM