

scratch

December 4, 2023

```
[7]: %matplotlib inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
plt.rcParams['figure.figsize']=(20.0, 10.0)

#Reading data
data=pd.read_csv(r"C:\Users\student\Desktop\ML\headbrain.csv")
print(data.shape)
data.head()
```

(237, 4)

```
[7]:
```

	Gender	Age Range	Head Size(cm ³)	Brain Weight(grams)
0	1	1	4512	1530
1	1	1	3738	1297
2	1	1	4261	1335
3	1	1	3777	1282
4	1	1	4177	1590

```
[11]: #collecting X and Y
X = data['Head Size(cm3)'].values
Y = data['Brain Weight(grams)'].values
```

```
[14]: # Mean X and Y
mean_x = np.mean(X)
mean_y = np.mean(Y)

#Total number of values
m= len(X)

# Using the formula to calculate b1 and b2
numerator = 0
denominator = 0
for i in range(m):
    numerator += (X[i] - mean_x) * (Y[i] - mean_y)
    denominator += (X[i] - mean_x) ** 2
b1 = numerator / denominator
```

```
b0 = mean_y - (b1 * mean_x)
```

```
# print coefficients
```

```
print(b1, b0)
```

0.26342933948939945 325.57342104944223

```
[18]: # plotting values and regression line
```

```
max_x = np.max(X) + 100
```

```
min_x = np.min(X) - 100
```

```
# Calculating line values x and y
```

```
x = np.linspace(min_x, max_x, 1000)
```

```
y = b0 + b1 * x
```

```
# Plotting Line
```

```
plt.plot(x,y, color = '#58b970', label = 'Regreesion line')
```

```
# Plotting scatter points
```

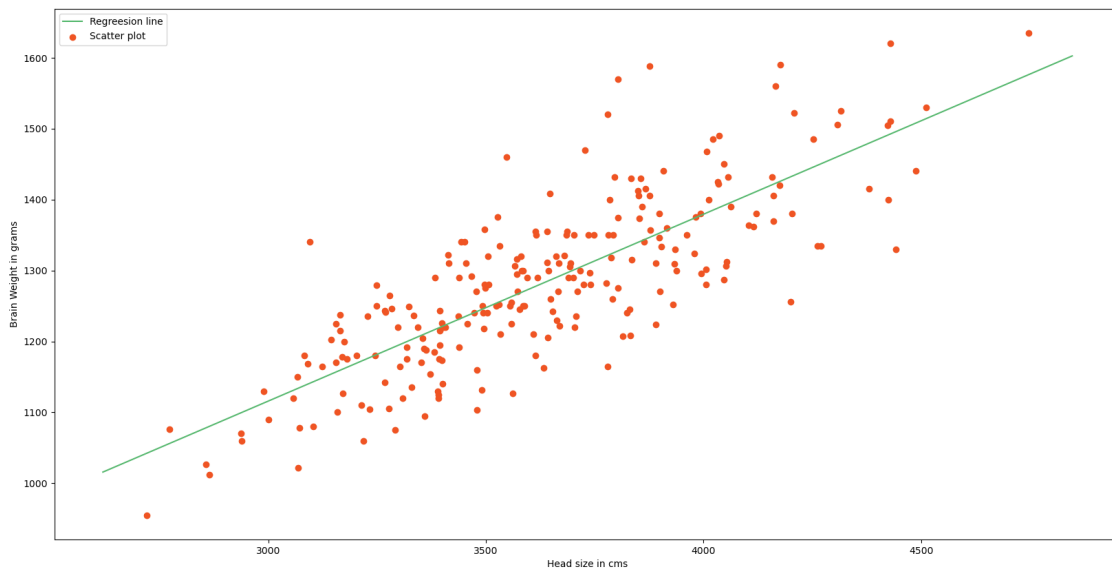
```
plt.scatter(X,Y,c='#ef5423' , label = 'Scatter plot')
```

```
plt.xlabel('Head size in cms')
```

```
plt.ylabel('Brain Weight in grams')
```

```
plt.legend()
```

```
plt.show()
```



```
[27]: ss_t = 0
```

```
ss_r = 0
```

```
for i in range(m):
    y_pred = b0 + b1 * X[i]
    ss_t += (Y[i] - mean_y)**2 # Corrected calculation for ss_t
    ss_r += (Y[i] - y_pred)**2

r2 = 1 - (ss_r/ss_t)
print(r2)
```

0.6393117199570003