EXP 2

Aim: Simple Linear Regression in Python

Theory:

Simple linear regression is a linear regression with one independent variable, also called the explanatory variable, and one dependent variable, also called the response variable. In simple linear regression, the dependent variable is continuous.

Simple linear regression helps make predictions and understand relationships between one independent variable and one dependent variable. For example, you might want to know how a tree's height (independent variable) affects the number of leaves it has (dependent variable). By collecting data and fitting a simple linear regression model, you could predict the number of leaves based on the tree's height. This is the 'making predictions' part. But this approach also reveals how much the number of leaves changes, on average, as the tree grows taller, which is how simple linear regression is also used to understand relationships.

Simple linear regression equation

Let's take a look at the simple linear regression equation. We can start by first looking at the slope-intercept form of a straight line using notation that is common in geometry or algebra textbooks. That is, we will start at the beginning.

$$y = mx + b$$

Here

- m is the slope of the line
- b is the intercept

Numerical:

Hours studied(X)	Marks obtained(Y)
1	50
2	55
	60
4	65
5	70

We want best fit line for this data, which can be represented by eqn:
y= mx+c
where m is slope and b is intercept
S1: Formula for the slope(m) and intercept(b)

$$m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{\sum y - m(\sum x)}{n}$$

S2: Calculating the necessary sums

$x=[1,2,3,4,5]$

$y=[50,55,60,65,70]$

1. Sum of x values:

$\sum x=1+2+3+4+5=15$

2.Sum of y values:

$\sum y=50+55+60+65+70=300$

3.Sum of x^2 values:

$\sum x2=12+22+32+42+52=1+4+9+16+25=55$

4. Sum of xy values:

$\sum xy=(1\times50)+(2\times55)+(3\times60)+(4\times65)+(5\times70)=50+110+180+260+350=950$

S3: Calculation the slope and intercept

Therefore

Slope (m) = 250/50 =5

Intercept(b) = 225/5=45

S4:

Final eqn:

$y= 5x +45$

S5: Predictions

For x=1:

$y=5(1)+45=50$

For x=2:

$y=5(2)+45=55$

For x=3:

$y=5(3)+45=60$

For x=4:

$y=5(4)+45=65$

For x=5:

$y=5(5)+45=70$

Code:

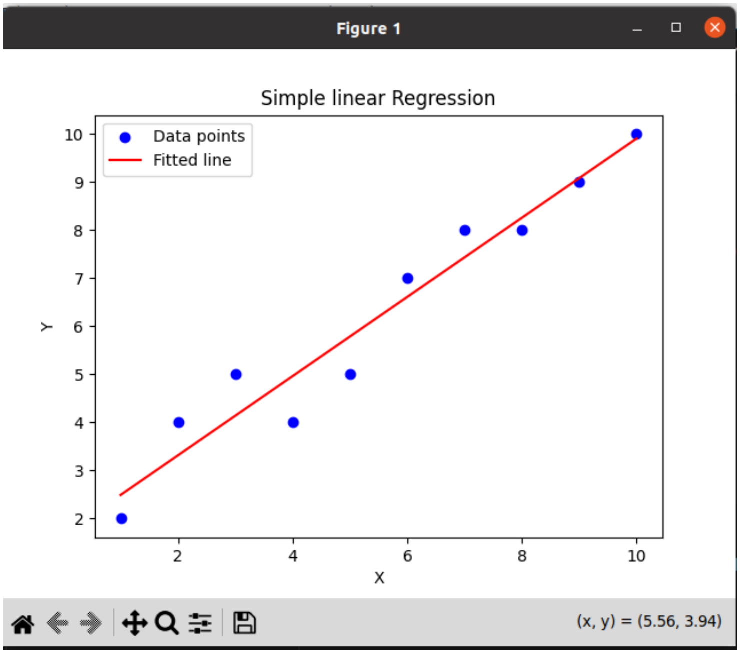
```
import numpy as np
import matplotlib.pyplot as plt
#defining the data X and Y
X= np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])
Y= np.array([2, 4, 5, 4, 5, 7, 8, 8, 9, 10])
#Calculating number of datapoints
n=len(X)
#Calculating the sum needed for formulas
```

```
sum_x=np.sum(X)
sum_y=np.sum(Y)
sum_xx=np.sum(X**2)
sum_xy=np.sum(X*Y)
#Applying slope(m) formula
m=(n*sum_xy-sum_x*sum_y)/(n*sum_xx-sum_x**2)
#Applying formular of intercept(b)
b=(sum_y-m*sum_x)/n
print(f"Slope(m):{m}")
print(f"Intercept (b):{b}")
#Make prediction for Y based on learned model (Y=mx+b)
Y_pred=m*X+b

#visualizing data adn fitted line
plt.scatter(X,Y,color='blue',label='Data points')
plt.plot(X,Y_pred,color='red',label='Fitted line')
plt.xlabel('X')
plt.ylabel('Y')
plt.legend()
plt.title('Simple linear Regression')
plt.show()
```

Output:

```
(.venv) a@computer-ThinkCentre:~/aman$ /home/a/aman/.venv/bin/python /home/a/aman/exp2.py
Slope(m):0.8242424242424242
Intercept (b):1.6666666666666672
```



Conclusion: In conclusion, Simple Linear Regression successfully identifies the relationship between two variables, predicting one based on the other using a straight-line equation. The model helps to understand the strength and direction of the relationship, offering valuable insights for making predictions. However, its effectiveness can be limited by the simplicity of the model and the nature of the data, especially in more complex scenarios.