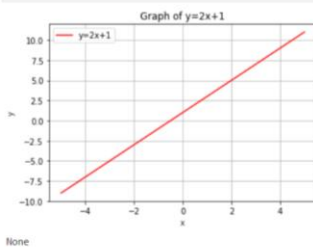


1. Using a graph to illustrate slope and intercept, define basic linear regression.

Ans: The equation $y=mx+c$ represents a straight line graphically, where m is its slope/gradient and c its intercept. In this tutorial, you will learn how to plot $y=mx+b$ in Python with Matplotlib.

```
In [2]: import matplotlib.pyplot as plt
import numpy as np
x = np.linspace(-5,5,100)
y = 2*x+1
plt.plot(x, y, '-r', label='y=2x+1')
plt.title('Graph of y=2x+1')
plt.xlabel('x', color='#1C2833')
plt.ylabel('y', color='#1C2833')
plt.legend(loc='upper left')
plt.grid()
display(plt.show())
```



2. In a graph, explain the terms rise, run, and slope.

Ans: The slope of a line measures the steepness of the line. Most of you are probably familiar with associating slope with "Rise Over Run". Rise means how many units you move up or down from point to point. On the graph that would be a change in the y values. Run means how far left or right you move from point to point.

3. Use a graph to demonstrate slope, linear positive slope, and linear negative slope, as well as the different conditions that contribute to the slope.

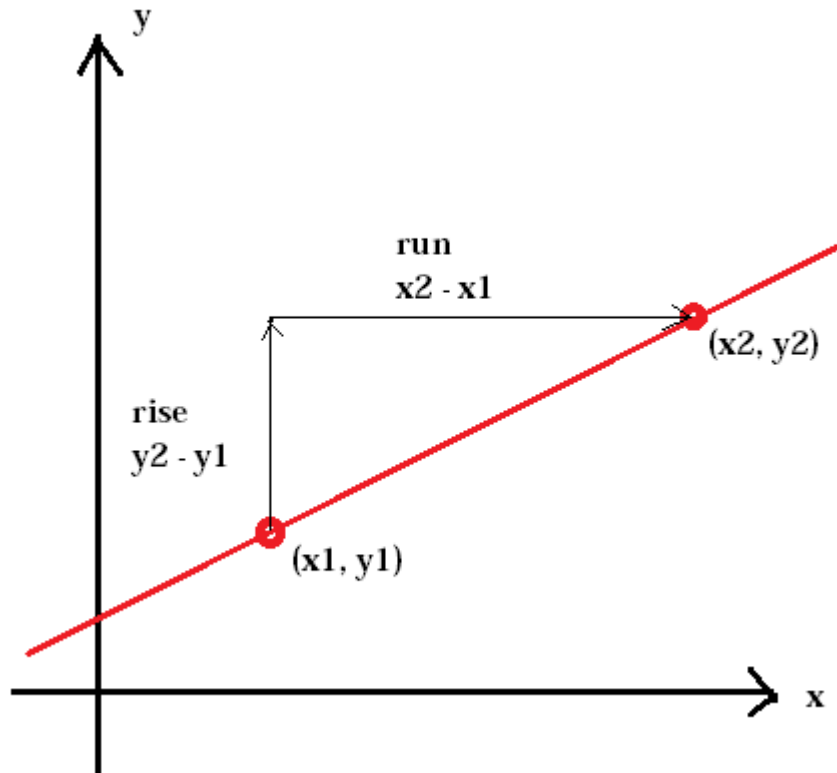
Ans: The steepness of a hill is called a slope. The same goes for the steepness of a line. The slope is defined as the ratio of the vertical change between two points, the rise, to the horizontal change between the same two points, the run.

$$\text{Slope} = \text{Rise/Run} = \text{Change in Y/change in X}$$

The slope of a line is usually represented by the letter m . (x_1, y_1) represents the first point whereas (x_2, y_2) represents the second point.

$$M = \frac{Y_2 - Y_1}{X_2 - X_1}$$

It is important to keep the x - and y -coordinates in the same order in both the numerator and the denominator otherwise you will get the wrong slope.



4. Use a graph to demonstrate curve linear negative slope and curve linear positive slope.

Ans: Image result for a graph to demonstrate curve linear negative slope and curve linear positive slope. If the signs are different then the answer is negative! If the slope is negative you can plot your next point by going down and right OR up and left. If the slope is positive you can plot your next point by going up and right OR down and left.

5. Use a graph to show the maximum and low points of curves.

Ans: To find the maximum/minimum of a curve you must first differentiate the function and then equate it to zero. This gives you one coordinate. To find the other you must resubstitute the one already found into the original function.

6. Use the formulas for a and b to explain ordinary least squares.

Ans: This best line is the Least Squares Regression Line (abbreviated as LSRL). This is true where \hat{y} is the predicted y-value given x, a is the y intercept, b and is the slope. For every x-value, the Least Squares Regression Line makes a predicted y-value that is close to the observed y-value, but usually slightly off.

7. Provide a step-by-step explanation of the OLS algorithm.

Ans: Ordinary Least Square Method :

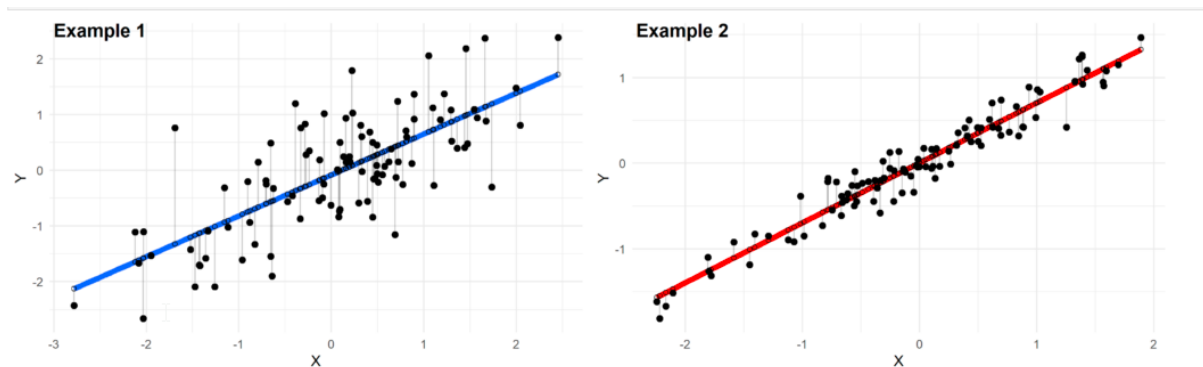
- Set a difference between dependent variable and its estimation:
- Square the difference:
- Take summation for all data.

- To get the parameters that make the sum of square difference become minimum, take partial derivative for each parameter and equate it with zero.

8. What is the regression's standard error? To represent the same, make a graph.

Ans: The standard error of the regression (S), also known as the standard error of the estimate, represents the average distance that the observed values fall from the regression line.

Conveniently, it tells you how wrong the regression model is on average using the units of the response variable.



9. Provide an example of multiple linear regression.

Ans: Multiple Linear Regression is one of the important regression algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable. Example: Prediction of CO2 emission based on engine size and number of cylinders in a car.

10. Describe the regression analysis assumptions and the BLUE principle.

Ans: There are four assumptions associated with a linear regression model:

- **Linearity:** The relationship between X and the mean of Y is linear.
- **Homoscedasticity:** The variance of residual is the same for any value of X.
- **Independence:** Observations are independent of each other.
- **BLUE:** is an acronym for the following: Best Linear Unbiased Estimator. In this context, the definition of "best" refers to the minimum variance or the narrowest sampling distribution.

11. Describe two major issues with regression analysis.

Ans: It involves very lengthy and complicated procedure of calculations and analysis. It cannot be used in case of qualitative phenomenon viz. honesty, crime etc.

The overall idea of regression is to examine two things:

1. Does a set of predictor variables do a good job in predicting an outcome (dependent) variable?
2. Which variables in particular are significant predictors of the outcome variable, and in what way do they—indicated by the magnitude and sign of the beta

12. How can the linear regression model's accuracy be improved?

Ans: 8 Methods to Boost the Accuracy of a Model :

- Add more data. Having more data is always a good idea.
- Treat missing and Outlier values.
- Feature Engineering.
- Feature Selection.
- Multiple algorithms.
- Algorithm Tuning.
- Ensemble methods.

13. Using an example, describe the polynomial regression model in detail.

Ans: In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an n th degree polynomial in x . For this reason, polynomial regression is considered to be a special case of multiple linear regression.

Polynomial regression is one of the machine learning algorithms used for making predictions. For example, it is widely applied to predict the spread rate of COVID-19 and other infectious diseases.

14. Provide a detailed explanation of logistic regression.

Ans: Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set. Based on historical data about earlier outcomes involving the same input criteria, it then scores new cases on their probability of falling into a particular outcome category.

15. What are the logistic regression assumptions?

Ans: Basic assumptions that must be met for logistic regression include independence of errors, linearity in the logit for continuous variables, absence of multicollinearity, and lack of strongly influential outliers.

16. Go through the details of maximum likelihood estimation.?

Ans: Maximum likelihood estimation is a method that determines values for the parameters of a model. The parameter values, are found such that they maximise the likelihood that the process described by the model produced the data that were, actually observed."