LINEAR AND LOGISTIC REGRESSION

Lab2

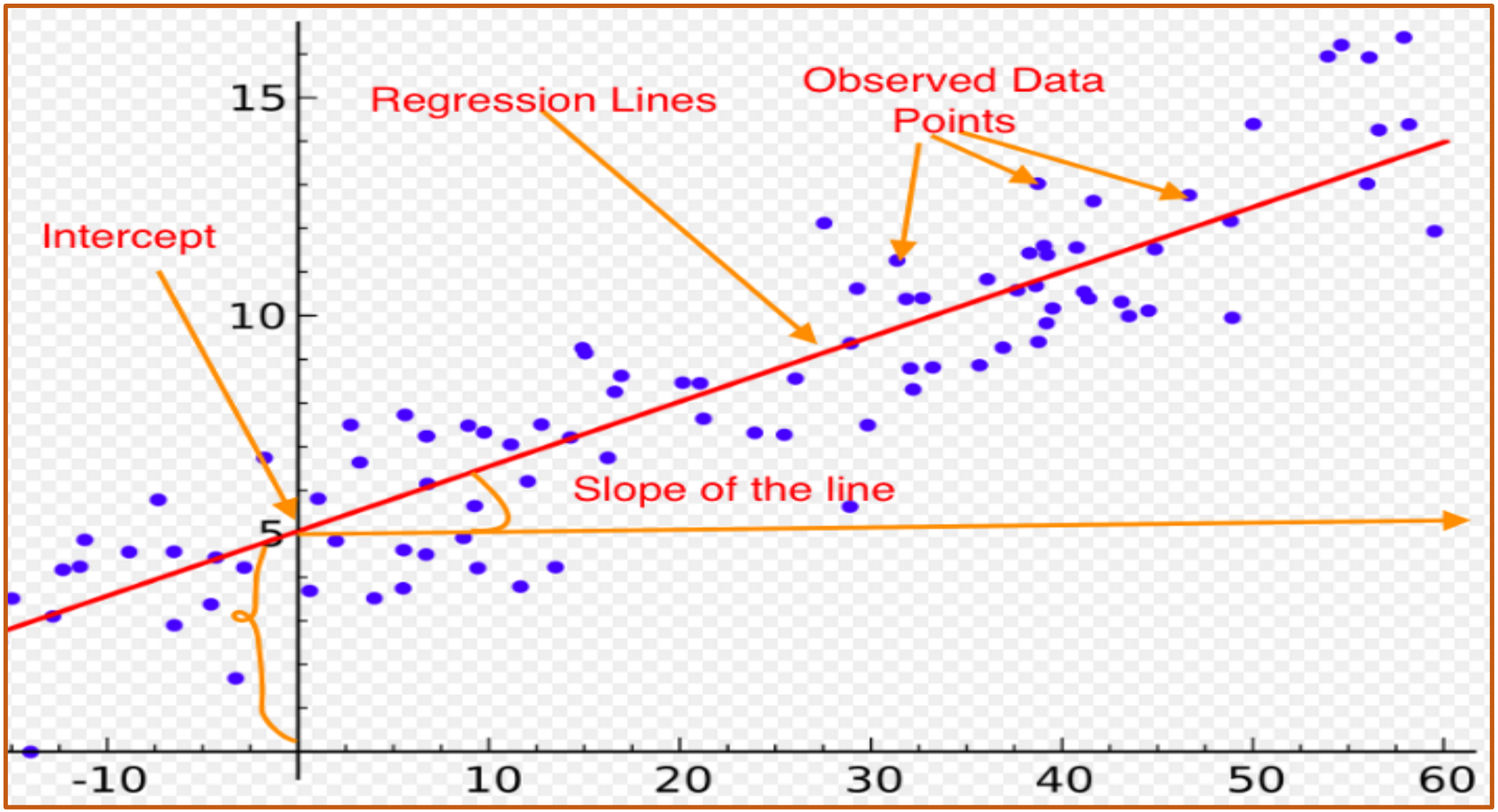
19-11-2021

REGRESSION

- Regression is used to describe the relationships between a set of independent variables and dependent variables.
- As an outcome of regression analysis, we get a mathematical equation often called a regression equation.
- Y=ax+b, b is the intercept and a is the slope
- By determining the values of “a” and “b” we can calculate the value of “y” for a given value of “x”.
- Primarily used for:
    I. Prediction and Forecasting
    II. Inferring relationships between the independent and dependent variables.

LINEAR REGRESSION

- Linear Regression is a regression analysis of dependent and independent variables when they exhibit a linear relationship.
- A supervised learning technique.
- **It assumes that the dependence of Y on X1, X2, . . . Xp is linear.**
- The core idea is to **obtain a line that best fits the data**.
- The Regression line is a straight line that best fits the data, such that the overall distance from the line to the points (variable values) plotted on a graph is the smallest.
- The formula for the best-fitting line (or regression line) is **y=ax+b** ,where a is commonly known as the *slope*, and b is commonly known as the *intercept,* "x" is an explanatory variable,"y" is a dependent variable
- Regression line attempts to define the predicted value of "y" (dependent variable) for a given value of "x" (independent variable).
- The best-fit regression line attempts to minimize the **sum of the squared distances** between the observed(actual) data points and the predicted ones.
- The intercept of regression lines helps us to estimate the value of "y" (dependent variable), having no effects of "x" (independent variable).


Regression Lines
Observed Data Points
Intercept
Slope of the line
15
10
5
-10
10
20
30
40
50
60

LOGISTIC REGRESSION

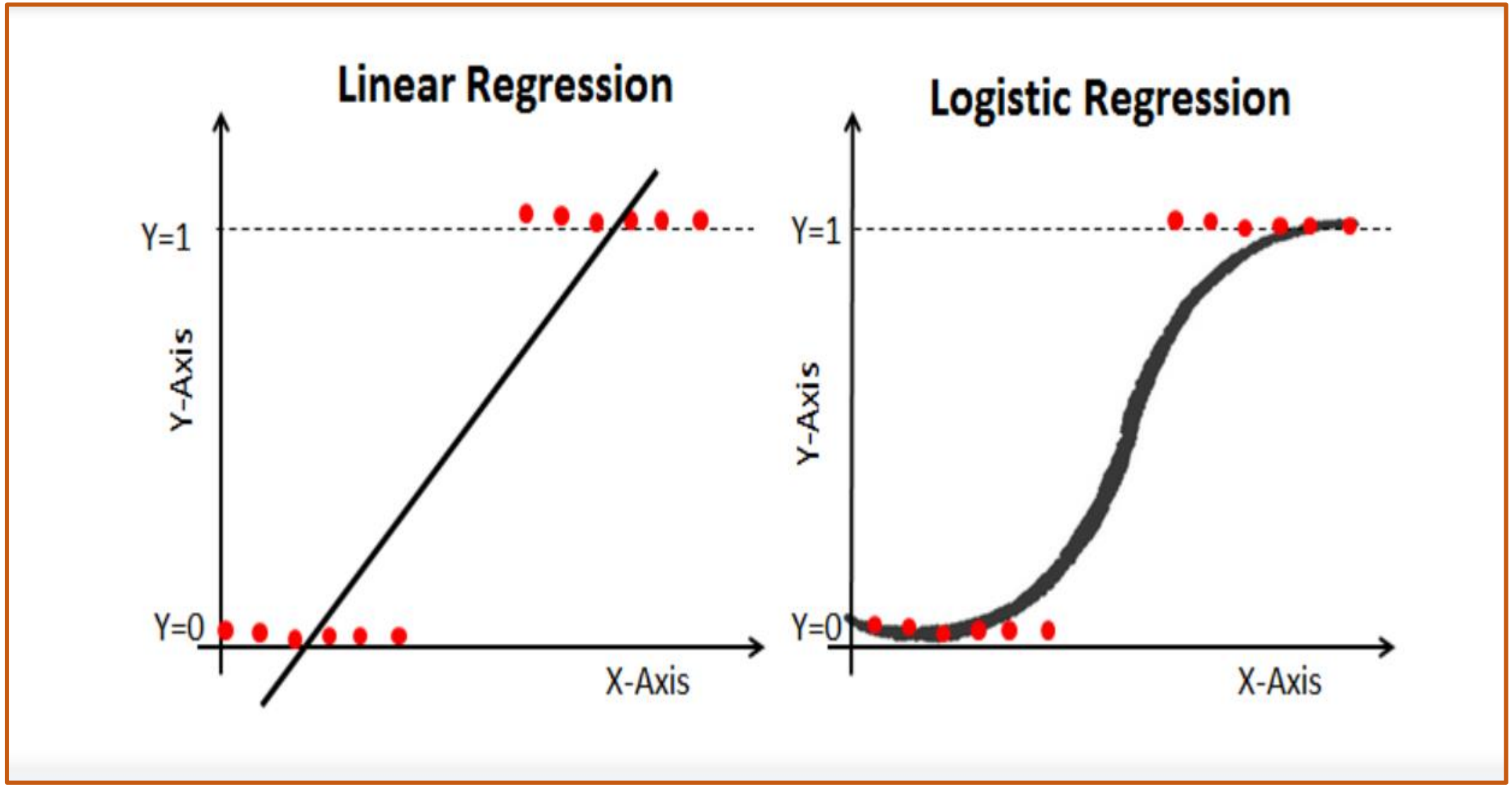
- It is an algorithm that can be used for regression as well as classification tasks but it is **widely used for classification tasks**.
- It is used to estimate the relationship between a dependent (target) variable and one or more independent variables.
- The output of the dependent variable is represented in discrete values such as 0 and 1.
- Logistic Regression, converts the straight best fit line in linear regression to an S-curve using the sigmoid function, which will always give values between 0 and 1.
- In logistic regression... **basically, you are performing linear regression but applying a sigmoid function for the predicted values.**

- **Sigmoid Function:**

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

- **Cost Function** :

$$\text{Log loss} = \frac{1}{N} \sum_{i=1}^{N} -(y_i * \log (\hat{Y}_i) + (1 - y_i) * \log (1 - \hat{Y}_i))$$


Linear Regression
Y=1
Y=0
Y-Axis
X-Axis
Logistic Regression
Y=1
Y=0
Y-Axis
X-Axis