Regression and essential concepts

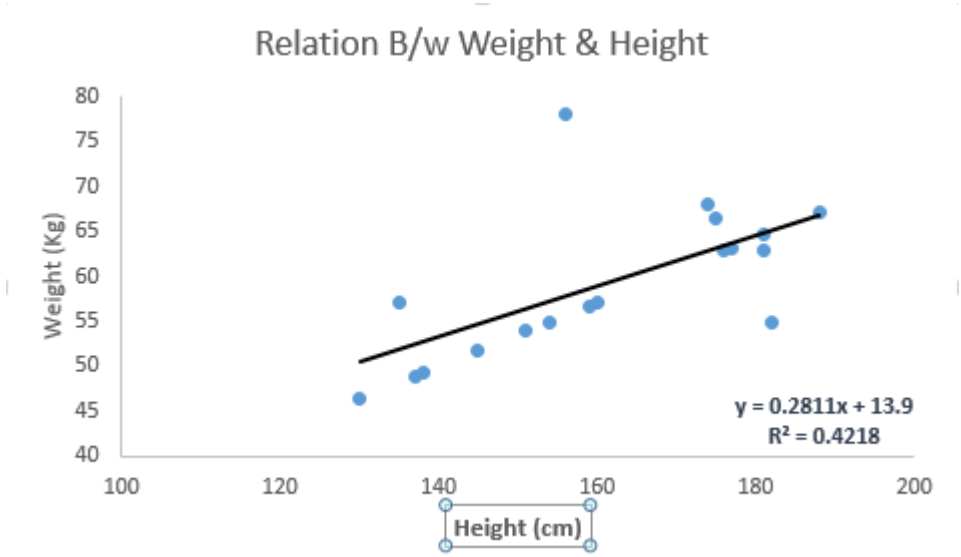
Introduction

- Statistical measurement to determine the strength of the relationship between one dependent variable (usually denoted by Y) and a series of other changing variables (known as independent variables).
    - Linear regression: $Y = a + bX + u$

- Y = the variable that you are trying to predict (dependent variable).
- X = the variable that you are using to predict Y (independent variable).
- a = the intercept.
- b = the slope.

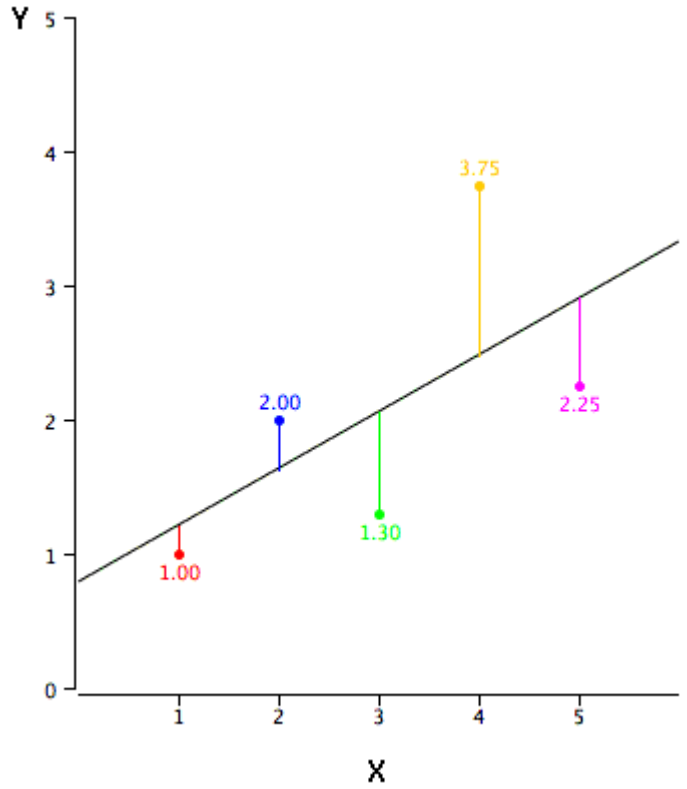
What is Regression Analysis?

- Investigates the relationship between a dependent (target) and independent variable (s) (predictor).
- This technique is used for forecasting and finding the causal effect relationship between the variables.
- For example, relationship between rash driving and number of road accidents by a driver is best studied through regression.



How to obtain best fit line (Value of a and b)?

- Easily accomplished by Least Square Method.
- Most common method used for fitting a regression line.
- Calculates the best-fit line for the observed data by minimizing the sum of the squares.



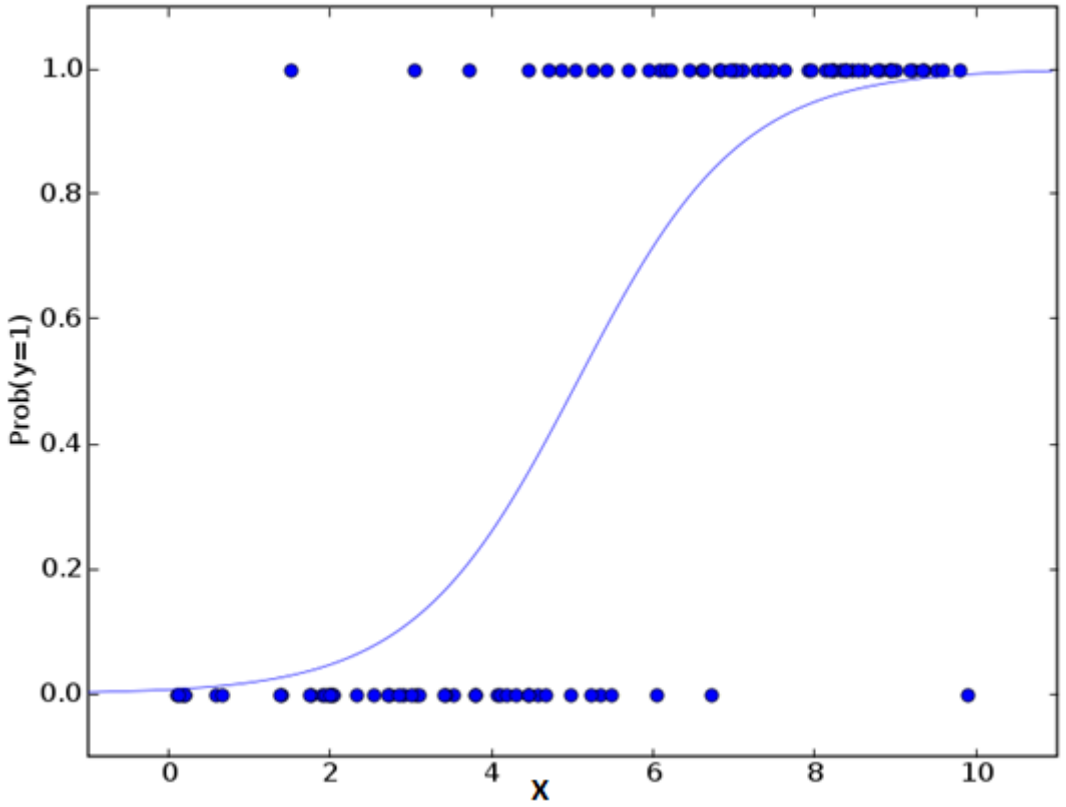
$$\min_{w} ||Xw - y||_2{}^2$$

Applet

- Refer to below URL for a live example:
  - http://www.rossmanchance.com/applets/RegShuffle.htm

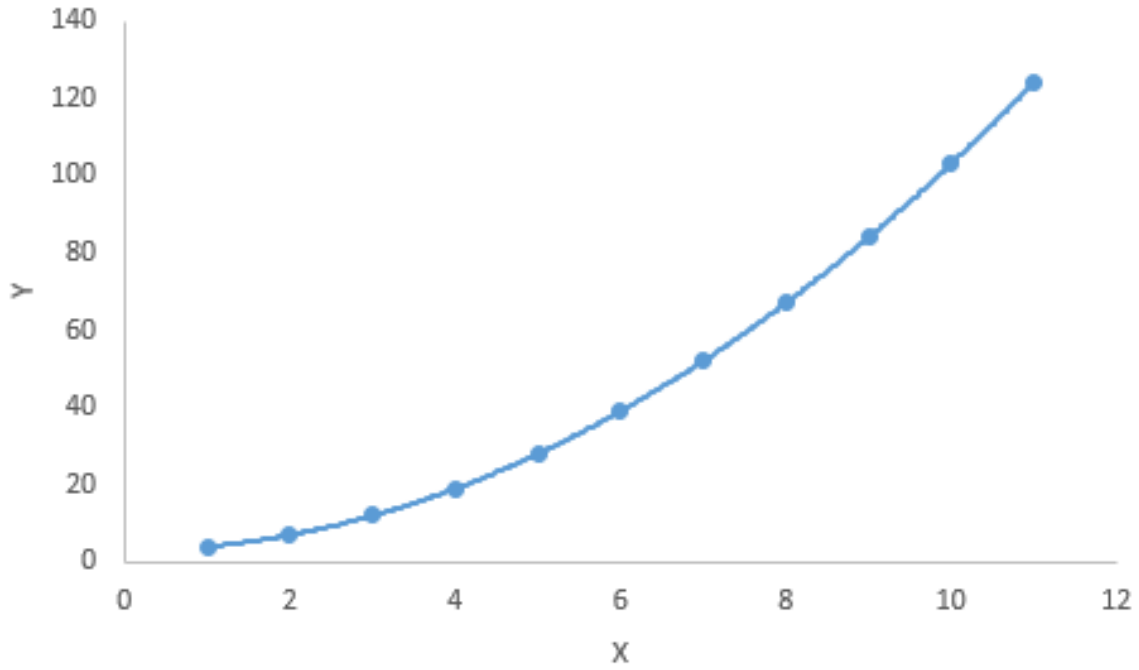
Logistic Regression

- To find the probability of event=Success and event=Failure.
- We should use logistic regression when the dependent variable is binary (0/ 1, True/ False, Yes/ No) in nature.

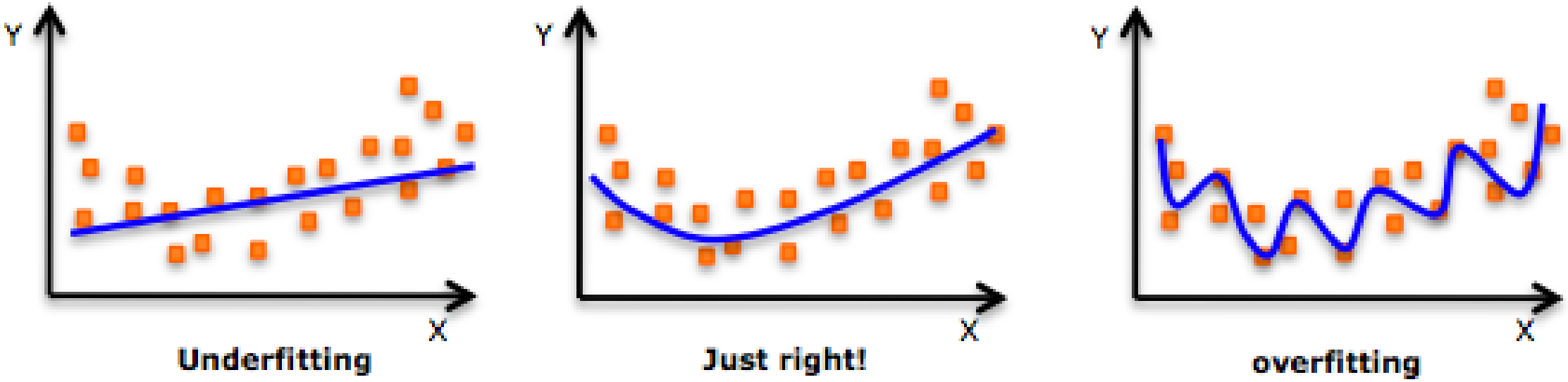


Polynomial Regression

- A regression equation is a polynomial regression equation if the power of independent variable is more than 1. The equation below represents a polynomial equation:
  - y=a+b*x^2
- In this regression technique, the best fit line is not a straight line. It is rather a curve that fits into the data points.



Important Points



Thanks