Linear Regression

Cong Cuong PHAM

November 29, 2016

Abstract

This document introduces some fundamental notions of Linear Regression.

1 Introduction

1.1 What is Linear Regression?

1800s: Francis Galton, was studying the relationship between parents and their children. He investigated the relationship between the heights of fathers and their sons. He discovered that a man's son tended to be roughly as tall as his father. However Galto's breakthrough was that the son's height tended to be closer to the overall overage height of all people. Galton called this phenomenon **regression**, as in "A father's son's height tends to regress (or drift towards) the mean (average) height".

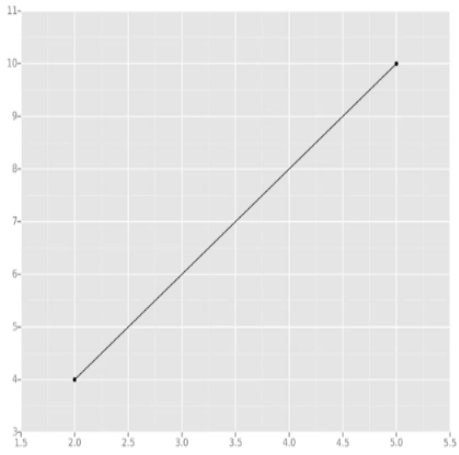

Figure 1: Two point

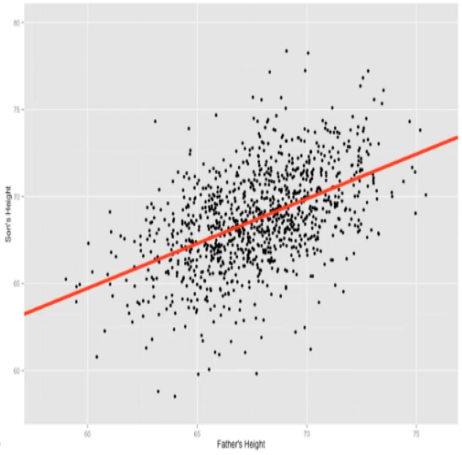

Figure 2: Multiple points

The goal of linear regression is to **minimize the vertical distance** between all the data point and the fitted line. In determining the **best line**, we are attempting to minimize the distance between **all** the points and their distance to the fitted line 2.

There are lots of different ways to minimize the vertical distance (sum of squared errors, sum of absolute errors, etc), but all these methods have a general gold of minimizing this value. The residuals for an observation if the different between the observation (the y-value) and the fitted line.

Example: Us the Least Squares Method which is fitted by **minimizing the sum of squares of the residuals**.

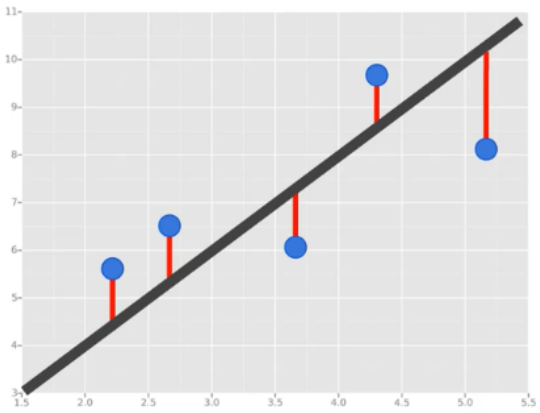

Figure 3: Choose the fitted line by minimizing the sum of square of the residuals.