OCTOBER 16, 2018 / #MATHEMATICS

Machine learning: an introduction to mean squared error and regression lines

by Moshe Binieli

Introduction

This article will deal with the statistical method **mean squared error**, and I'll describe the relationship of this method to the **regression line**.

The example consists of points on the Cartesian axis. We will define a mathematical function that will give us the straight line that passes best between all points on the Cartesian axis.

And in this way, we will learn the connection between these two methods, and how the result of their connection looks together.

General explanation

This is the definition from Wikipedia:

> In statistics, the mean squared error (MSE) of an estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors — that is, the average squared difference between the estimated values and what is estimated. MSE is a risk function, corresponding to the expected value of the squared error loss. The fact that MSE is almost always strictly positive (and not zero) is because of randomness or because the estimator does not account for information that could produce a more accurate estimate.

The structure of the article

The structure of the article

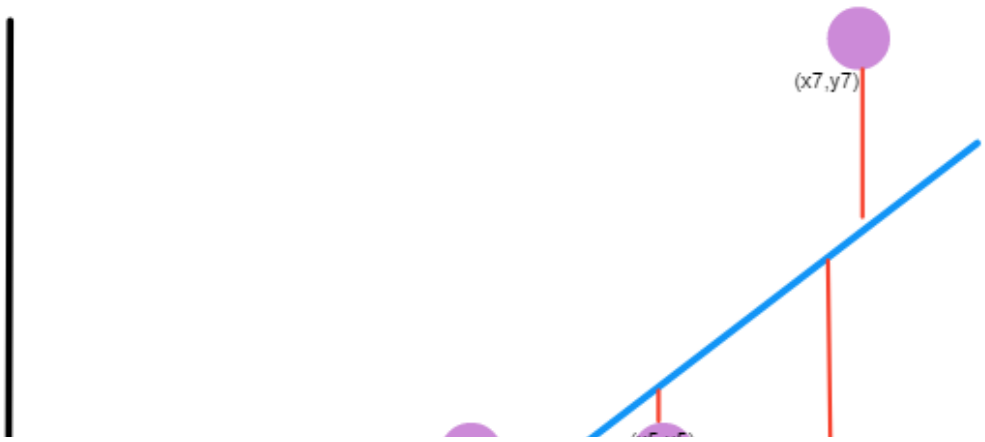
- Get a feel for the idea, graph visualization, mean squared error equation.
- The mathematical part which contains algebraic manipulations and a derivative of two-variable functions for finding a minimum. This section is **for those who want to understand how** we get the mathematical formulas later, you can skip it if that doesn't interest you.
- An explanation of the mathematical formulae we received and the role of each variable in the formula.
- Examples

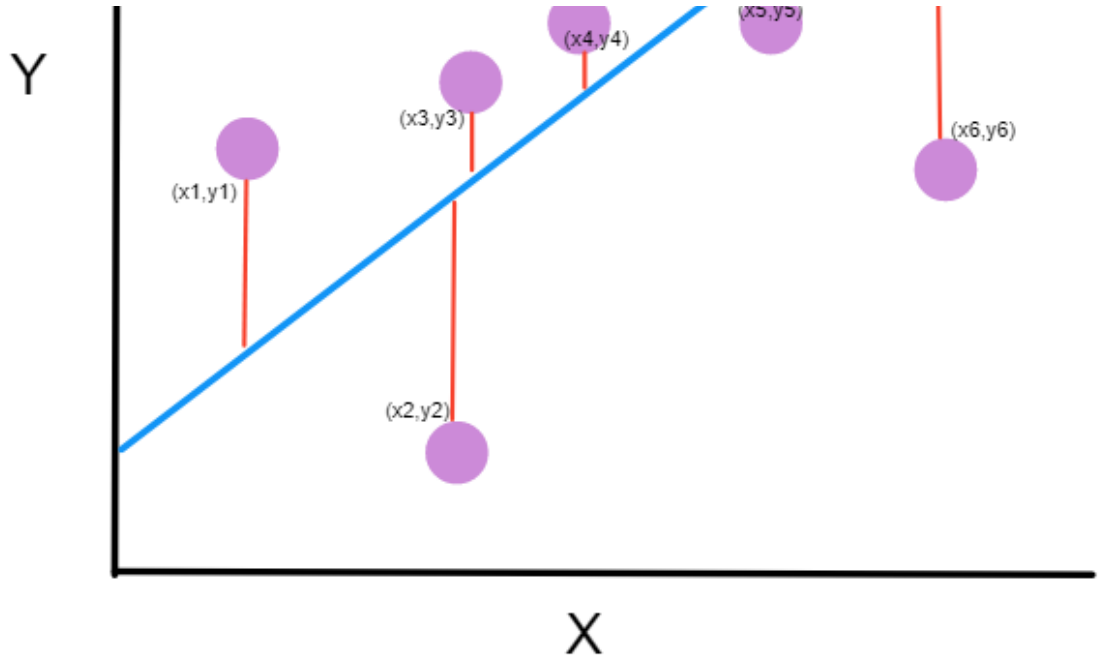
Get a feel for the idea

Let's say we have seven points, and our goal is to find a line that **minimizes** the squared distances to these different points.

Let's try to understand that.

I will take an example and I will draw a line between the points. Of course, my drawing isn't the best, but it's just for demonstration purposes.





Points on a simple graph.

You might be asking yourself, what is this graph?

- the **purple dots** are the points on the graph. Each point has an x-coordinate and a y-coordinate.
- The **blue line** is our prediction line. This is a line that passes through all the points and fits them in the best way. This line contains the predicted points.
- The **red line** between each purple point and the prediction line are the **errors.** Each error is the distance from the point to its predicted point.

You should remember this equation from your school days, ***y=Mx+B***, where **M** is the slope of the line and **B** is y-intercept of the line.

We want to find M (slope) and B (y-intercept) that **minimizes** the