Linear Regression

Boyan Dafov

February 25, 2020

I Solving the 1-D problem

First we are going to define and solve the linear regression problem for one independant and one dependant variable. We are looking for the "best fitting line" for random 2D data points. The equation we are looking to build looks like that:

$$\hat{y} = a * x + b$$

Where "y" and "x" are given for every data point. Lets the fine the "Error" function, assuming that we have "n" data points given :

$$E = \sum_{i=1}^{n} (\hat{y_i} - y_i)^2, \text{ where } \hat{y_i} \text{ is the prediction for } x_i$$

Now we are looking for the model, that gives us the minimum "Error function" result. First we are going to use that:

$$\hat{y_i} = a * x_i + b$$

Substitute it in the "Error function" above and get :

$$E = \sum_{i=1}^{n} (a * x_i + b - y_i)^2, \text{ where we are given } y_i \text{ and } x_i \text{ for every i}$$

To find the values for "a" and "b", we are going to take the derivatives of the function with respect first to "a" and than to "b":

II Solving the multidimensional problem