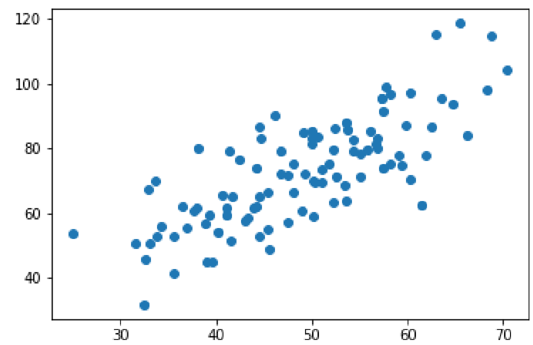
Linear Regression



Let us consider that we have this dataset (x_i, y_i) and we need to fit the best fit straight line between this data points.

So, how to do that?

Let's see.

We'll first take a random straight line $y = mx + c$. Then, we can calculate some kind of loss/cost that signifies sum of the distances between our assumed line and the point the taken dataset constitutes of. Based on this loss function, we'll try to minimize this loss and eventually, as a result, move our assumed line in the direction of best fit. After doing this for a large number of times, we'll get our best fit straight line that will serve as the linear regression model for the given dataset.

That's, err, a lot of words, let's see the math more clearly.

Let us first take the assumed line/hypothesis as:

$$h(x, m, c) = mx + c$$

We talked about a Loss/Cost function. We'll use Mean Squared Error (MSE) for this situation. The cost function will be defined as:

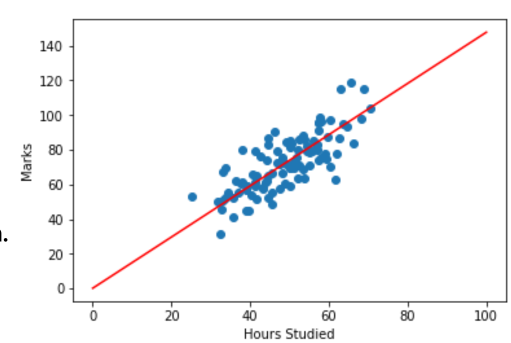
$$L(m, c) = \frac{1}{N} \sum_{i=1}^{N} (\, h(x_i, m, c) - y_i\,)^2$$

As the data given is fixed, (x_i, y_i) are constants. To minimize the Loss, we need to minimize it with respect to m and c individually. Clearly,

$$\frac{\partial L(m, c)}{\partial m} = \frac{2}{N} \sum_{i=1}^{N} (mx_i + c - y_i)x_i$$

And,

$$\frac{\partial L(m, c)}{\partial c} = \frac{2}{N} \sum_{i=1}^{N} (mx_i + c - y_i)$$



A clever observation – To go to the minimum of the loss w.r.t. either c or m, we need to subtract this $\partial L/\partial c$ or $\partial L/\partial m$ term from the already taken c or m.

There is a concept of learning rate, which is the constant you'll need to multiply $\partial L/\partial c$ or $\partial L/\partial m$ term by, so that we don't overshoot the minima.

And that's how you'll get your new m and c values. You can plot that and get your line of best fit!