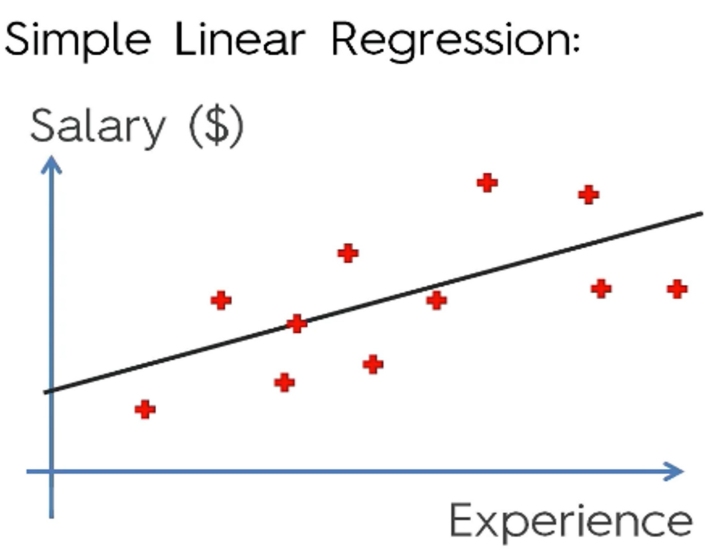
Step 2: Simple Linear Regression Intuition

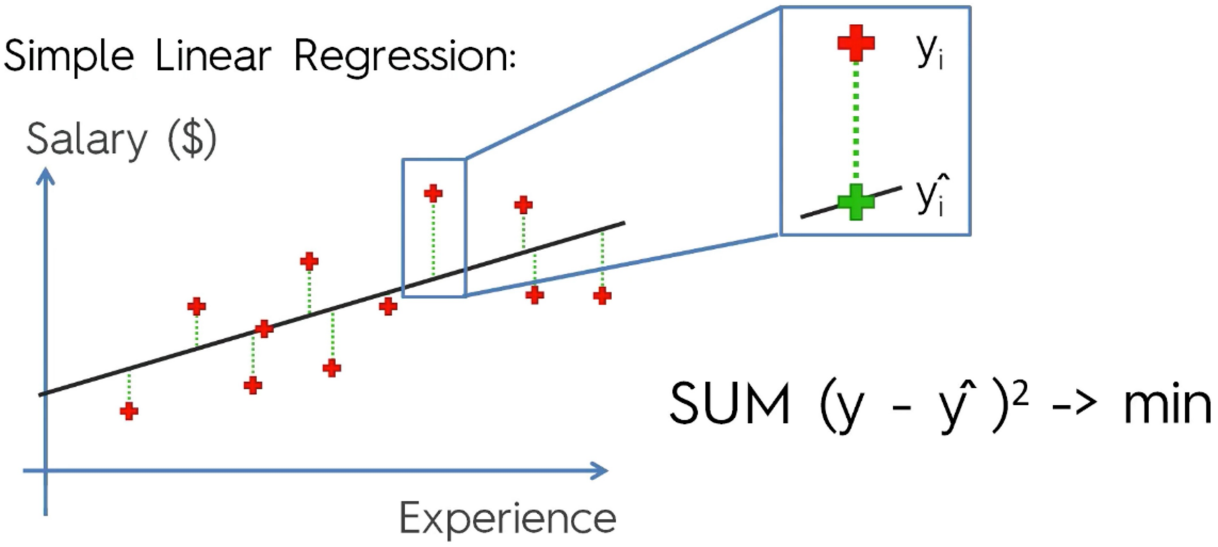
Thursday, March 31, 2022 5:58 PM

Ordinary Least Squares



Red cross is actual observation that we have in data.

Ordinary Least Squares



- Red cross is actual observation that we have in data and we have got the trend line which represents the best fitting line or the simple linear regression model.
- Lets take an example:

i) In above zoomed figure we can see that red cross is the person's salary in actual data. The model line or the black line at the bottom, it actually tells us where that should be sitting in salary according to the model in terms of salary. According

to the model, it should be lower, it should be somewhere where green cross is.

ii) The red cross is called y_i and that is actual observation.

iii) The green cross is called $\hat{y}_i$ and that is modeled observation or value.

iv) The green line is the difference between what he is actually earning(y_i) and what he should be earning or what he is modeled to be earning($\hat{y}_i$).

Simply, it is the difference between observed and the modelled.

v) Now, to get the best fitting line, what is done is you take each one of the green line or those distances and square them and then you take the sum of those squares. Once, you have the sum of the squares, we have to find the minimum. So, basically what a simple regression does is it draws a lots of these line(trend line). Linear Regression draws all possible trend lines through our data and counts the sum of those squares every single time and it records the sum temporarily imn file or something then it finds the minimum sum of squares. And that will be the best fitting line and that is called Ordinary Least Squares method.

So, that's how a Simple Linear Regression works.