**Reflection after Class 10**

Linear regression is a type of analysis that helps us predict values based on a given dataset. It is a way of modeling the relationship between the inputs and their scalar outputs. To create a linear regression model, you need to graph what is commonly known as the “line of best fit.” To do this, you need to:

1. Plot the data from the dataset in the form of a scatterplot, making sure that the x-axis corresponds with the independent variable (which can be continuous or discrete) and the y-axis corresponds with the dependent variable (which must be continuous).
2. Draw a line across the graph, measure the distance of each point to the line, square the distances, and sum up the numbers.
3. Repeat Step 2 after rotating the line slightly, and continue to do so until it is possible to create a new graph, this one with the x-axis (independent variable) representing the different degrees of rotation of the line and the y-axis (dependent variable) representing the sum of the squares of the distances from each point to the line with the corresponding degrees of rotation (the sum of the squared residuals).
4. The line with the least sum is the line of best fit for the data in the scatterplot.

This method is known as “Least Squares,” and requires you to estimate two parameters— the y-intercept of the line of best fit and the slope of the line of best fit.

Using the graph with the line of best fit (derived by running through Steps 1-4 above), we can find the likelihood of a guess coming close to the line of best fit in a form known as R-squared. R-squared is a very important aspect of linear regression as it displays the variation of the data relative to the line of best fit and helps us understand the significance of this. In order to find the R-squared, you can do the following:

1. Draw a horizontal line on the scatterplot with the y-value of every point on the line being the average of the y-values of the data points.
2. Using the aforementioned horizontal line, sum the squared residuals and divide this sum by the sample size. This process provides you with the variation of the means.
3. Using the line of best fit, sum the squared residuals and divide this sum by the sample size. This process provides you with the variation of fit.
4. You can calculate the R-squared by evaluating the formula as follows: ((variation of the means) - (variation of fit)) / (variation of the means).

Again, by calculating the R-squared, you can get the R-squared, which shows you how well the line of best fit matches with the data.

Linear regression is exceedingly important as it allows us to predict the outcome of an event and also understand how likely it is that our prediction is accurate. It can be used to create a predictive model, or even to perceive the range of variation that the dependent variable in a dataset may go through. Some even say that it is a form of machine learning, with how useful it can be (and with how the prediction works). This idea is part of a (relatively) controversial debate between data scientists about what constitutes a form of machine learning.

In terms of the course, I would say that we’re going at a good pace. It’s quicker than it was at the beginning of the course, but it’s still engaging and isn’t hard to learn. I would say that I do wish we had more time, as I feel that we could benefit from more lab time and more lecture time. I would recommend that you lecture us while we do the lab (and if you need to play videos, you do it while we learn/after some time working on the lab), as I believe that would be more efficient. And again, I think this is more of an issue with the amount of time we have left. So far, we’ve covered a lot, and that’s likely due to your teaching ability. Thank you for being such a wonderful instructor.