**LOGISTIC REGRESSION**

* We already know about linear regression. We have simple linear regression, and it has a very short formula , with one independent variable.
* We have also looked into multiple linear regression, which has many independent variable

.

Chart, scatter chart

Description automatically generated

Chart

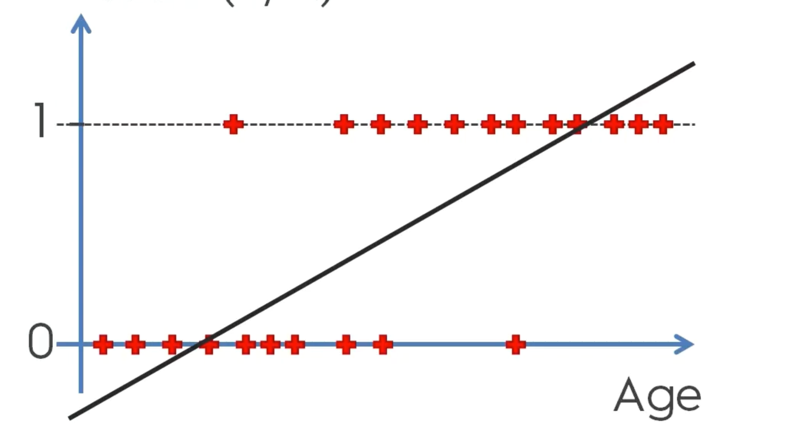
Description automatically generated

* So, for the above example let’s take an example – some customers were contacted through email and telephone about a product, and here you have the sample of some recently contacted customers, their actions – whether they took any action to our approach or not, and their age.
* And we can see here that the observations on the top are bit more to the right, and observations on the bottom are more to the left. Implying kind of that older people are more likely to act based on the offer, and younger people are more likely to ignore this.
* This is what it will look like if we try to implement linear regression on this set of data –

Chart, scatter chart

Description automatically generated

* And as we can see, it doesn’t look like the best method/approach to solve this problem.
* Let’s draw a horizontal line above.



* Instead of trying to predict the outcome action for any given person, knowing their age, how about we predict the probability?
* We will state a probability or a likelihood of that person taking up that offer.
* If we think about it that way, right away things start becoming clearer.
* Right away we can see that this chart is from 0 – 1, and probabilities are between 0 – 1. Basically, we can fit in probabilities between 0 – 1. And the fact that red observations are already either 0 or 1, and nowhere in between, that’s because we already know the result.
* But for something that we are predicting, it kind of makes more sense to say, instead of being sure about the outcome, the probability of the action to occur.
* When you think of it that way, the linear regression line, or at least the part that’s in the middle between 0 and 1, makes sense. Because that is basically telling you that anybody between those ages, for instance where it’s crossing the horizontal line for the first time – it might be 25 or 35, and the horizontal line it crosses at one – it might be 55, has a probability in increasing order of the person responding you, as the age increases.
* So, the part of the linear regression in the middle makes sense and we can do something about it. But the parts that don’t make sense at all are the ones at the top or the bottom. Because probability can never be less than 0 or greater than 1. So, what is it trying to tell us, or give us a hint about?
* It could be interpreted as people above that nominal age – 55, are more likely to take up the product. Anybody below 35, on the bottom left side are defiantly not taking up the product or the offer.

Chart, line chart

Description automatically generated

So, we can cut that regression line, and make it look like this.

* This would be very basic, but it would be an attempt at creating a model for this situation. So, we would still be able to use this to make predictions and assumptions that talks about the correlation between action and age of the customer.

**Scientific Approach for Logistic Regression**

Diagram

Description automatically generated

* We can use the above logistic regression model to predict probabilities about the occurrence of an event. The probability here is called p^ (p-hat).

Note - And anything with a hat in this section means something we are predicting. So, we are predicting the probability.

Why did we apply feature scaling in the implementation?

* Feature Scaling is not required for logistic regression.
* However, still applying it will improve training performance.
* Therefore, it will improve final predictions.
* So, for some models, like SVR, feature scaling is an absolute necessity. However, for other models like Logistic Regression it is not an absolute necessary. But still applying it will improve the training performance, and thus, the final predictions.
* You can implement the same without implementing feature scaling.
* AND ALWAYS FEATURE SCALE AFTER SPLITTING THE DATASET INTO TRAINING SET AND TEST SET TO AVOID INFORMATION LEAKAGE FROM TEST SET.