**LOGISTIC REGRESSION**

Linear Regression

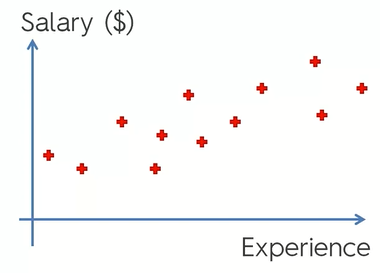
- Simple

y = b0 + b1\*x1

Multiple

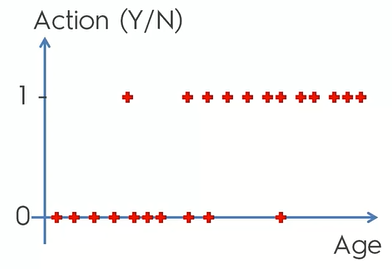
y = b0 + b1\*x1 + ... + bn\*xn

When we have this kind of scattered plot we use linear regression



Now, your company hired you as a data scientist. They send e-mail offers to customers like a proposal to buy some products. It may be a clothing store, grocery store or anything. They send out an offer in e-mail to lots of customers to purchase some product.

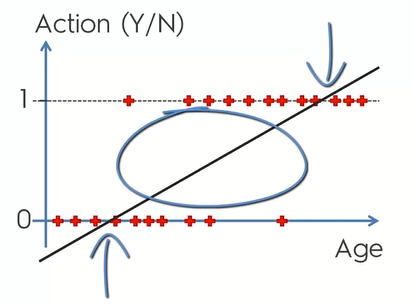
Here is a sample of customers they contacted recently. You got their age and also have a variable whether they take up an action or not(i.e respond to the e-mail or not, purchased or not).

There is some kind of co-relation. Older people are more likely to take action on this offer, younger people ignore it.

Can we model it somehow?

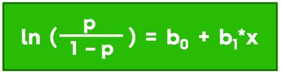
Linear regression does not seem like the best approach.

Also instead of predicting exactly whether they are going to take up an action or not, we state a probability for that person taking up the offer.

By linear regression model we see that the linear regression or atleast part of it that is in the middle between 0 and 1 means that anybody between those ages there is a probability that they will take up the offer and the probability increasing as we move to the right. But the part above and below does not make sense as probability cannot be lass than 0 and more than 1. Therefore we modify the graph.

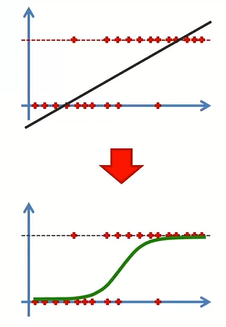
Actual scientific approach:

When we apply sigmoid function(p = 1/(1-e-y) ) to simple linear equation(y = b0  + b1\*x), then we get the following equation:



This is the formula for logistic regression.

This will convert our graph to this:



Based on the observations qnd the logistic regression equation we get the graph and is the best fitting line that can fit the dataset.

We can use this logistic regression to predict probabilities.

Let us take 4 random values and project these values on curve. Anything that falls on this curve below 0.5 line will be projected downwards on to the zero line.

=> If the predicted probability of taking up the offer is less than 50% then we are just going to say that the probability is the person is not going to take up this offer.

(if above 0.5 line then projected on 1 line and the person is going to take up the offer.)

