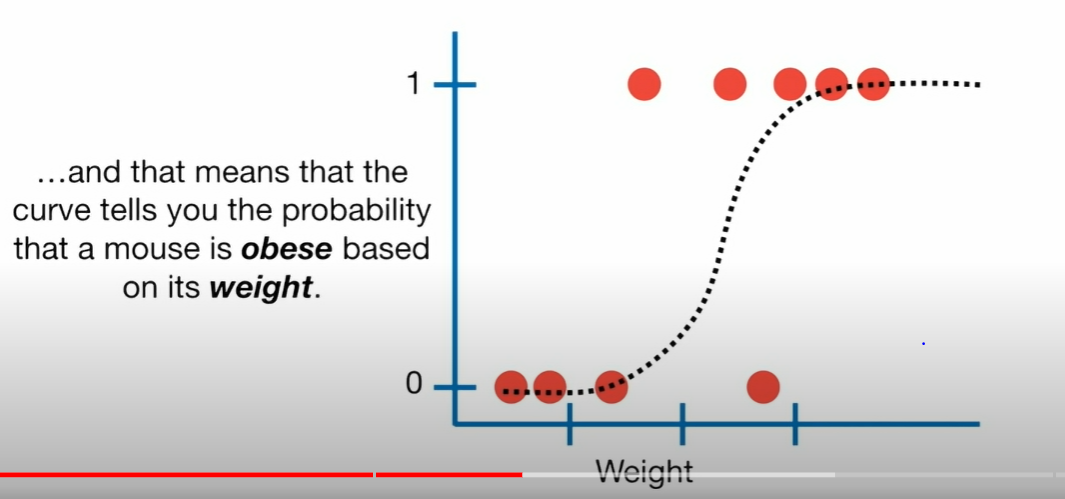
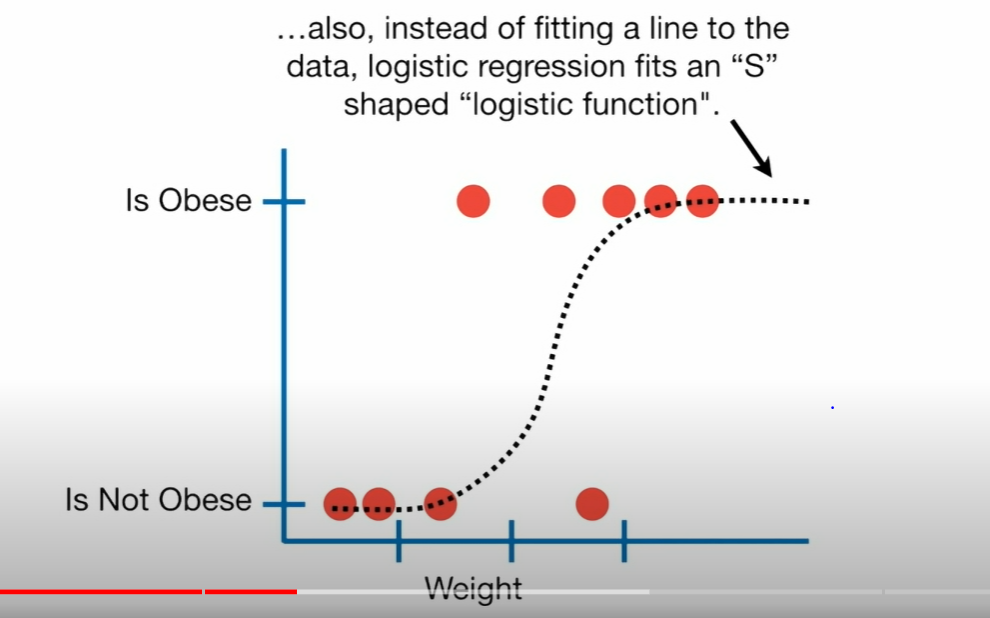
* Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
* Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, **it gives the probabilistic values which lie between 0 and 1**. This value of prob = p is calculated using **the Maximum Likelihood concept .**
* Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas **Logistic regression is used for solving the classification problems**.
* **In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).**



**FIG: S** shaped curve represents probability. Here in Fig ii, Pt. 0 = denotes Is Not Obese and Pt.1 = denotes Is Obese and rest of the pts in between denote Probabilities

* The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.

EXAMPLE: Cross pt. in Fig(i) below, denotes that there is a small prob. Of mouse being Obese and a large prob of it being a Not Obese . Slly in Fig (ii) below , an Intermediate weighing mouse (shown with a cross pt) will have a a 50 % chance of being a obese and Not obese.

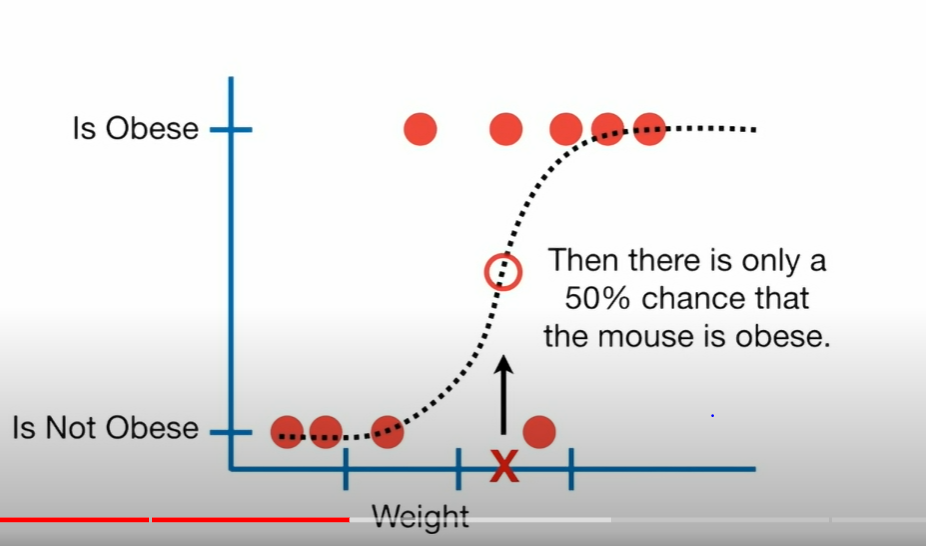
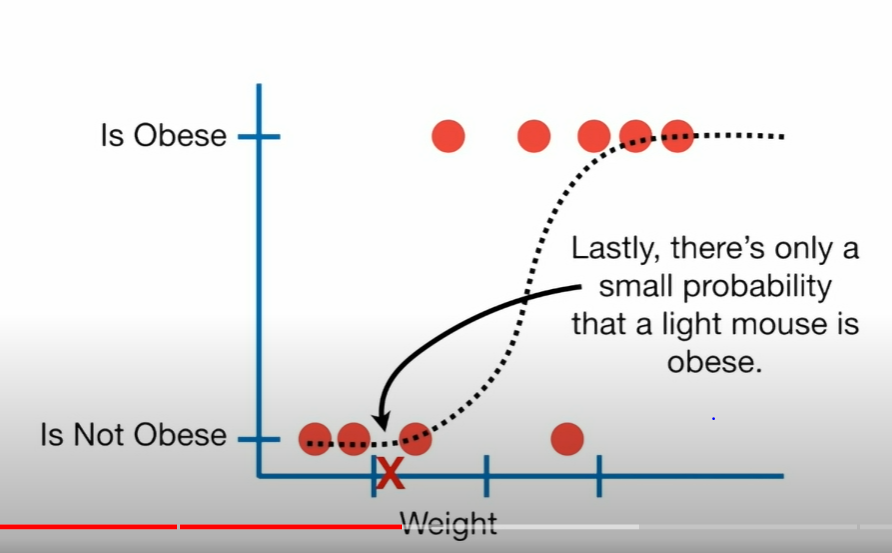


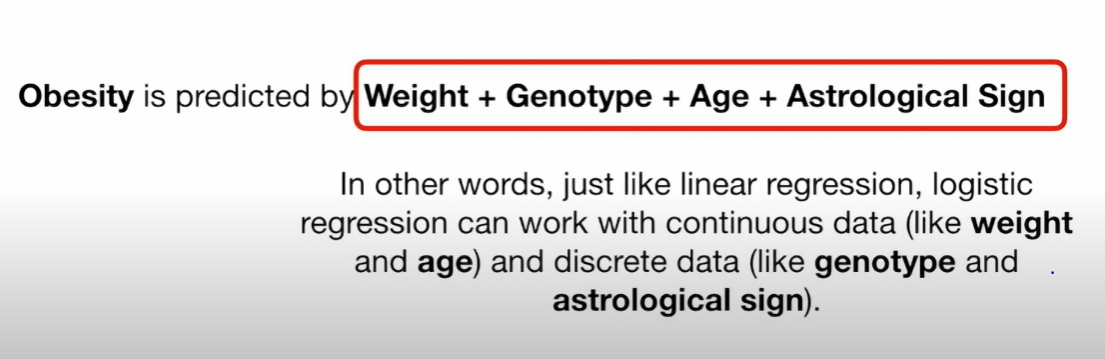
Fig: (i) Fig(ii)

* **We usually use Logistic Regression for Classification even though it is capable of giving us an exact Prob. of any event.** In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

Ex in above case: Usually if Prob of Obese is greater than 50% (or Prob of Not Obese is less than 50%), then mouse is classified as Obese. And if its vice versa , then it is classified as Not Obese.



* Just like Linear regression, Logistic Regression can be modelled using one or Mutiple dependent variables . ex:



* Here also we can test to see if the dependent variable’s effect on the prediction is significant or Not .i.e. do we need to consider that variable in our model or Not. The test we use to figure this out is called **Wald’s Test. (recall that in Linear regression we were using p-value test to determine this).**
* **We don’t have any concept of least squared and R-squared in Logistic Regression.(explained later)**
* **To find the best S curve** (best fit sigmoid Fn) , we use the concept of **Maximum Likelihood.**

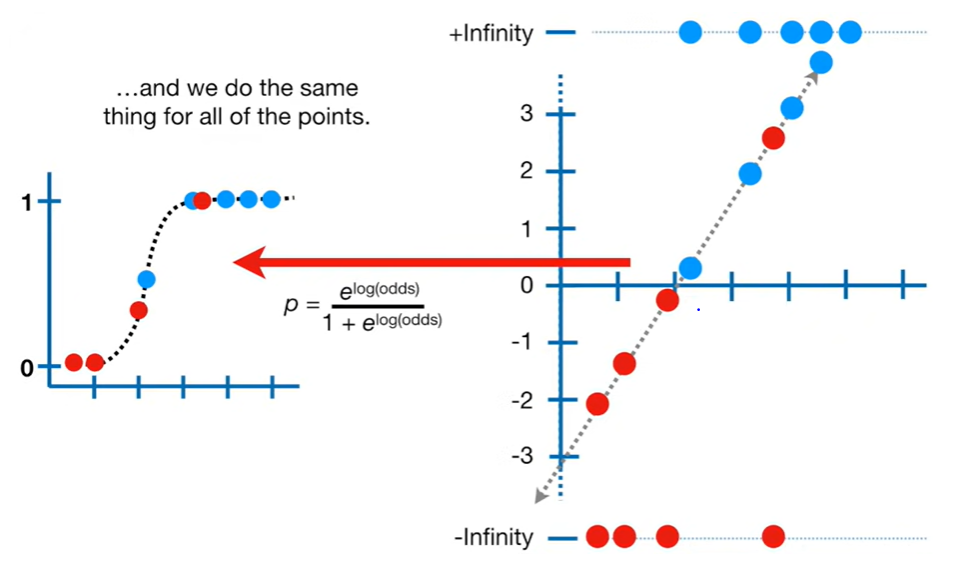
Type of Logistic Regression:

On the basis of the categories, Logistic Regression can be classified into three types:

* **Binomial or Binary Classification:** In binomial Logistic regression, there can be only two possible types of outcomes, such as 0 or 1, Pass or Fail, etc.
* **Multinomial or Multi class Classification:** In multinomial Logistic regression, there can be 3 or more possible unordered types of outcomes, such as "cat", "dogs","sheep" , or, low, medium, high.

1. **Logistic Regression for Continuous Variables :**

**This type of Logistic regression is closely related to Linear Regression i.e. they are solved by first modelling a Linear curve and then fitting it to a S- shaped curve.**



**Fig: Observe that 1st we start with any line**

**(candidate line) and then we fit it to the S-curve .Then we calculate its Maximum Probability We keep on doing this until we have found the highest value for Maximum Probability and hence found the best fit line and best fit S-curve**

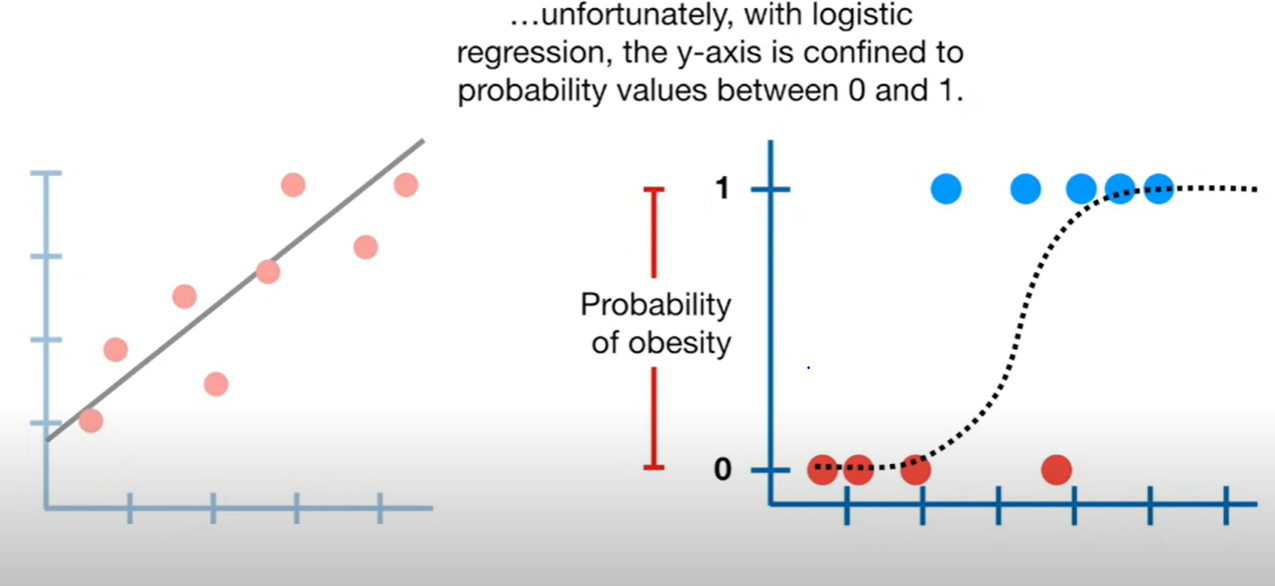
g(z)=1/1+e−z where g(z) is our Sigmoid Function ranges from (0,1)

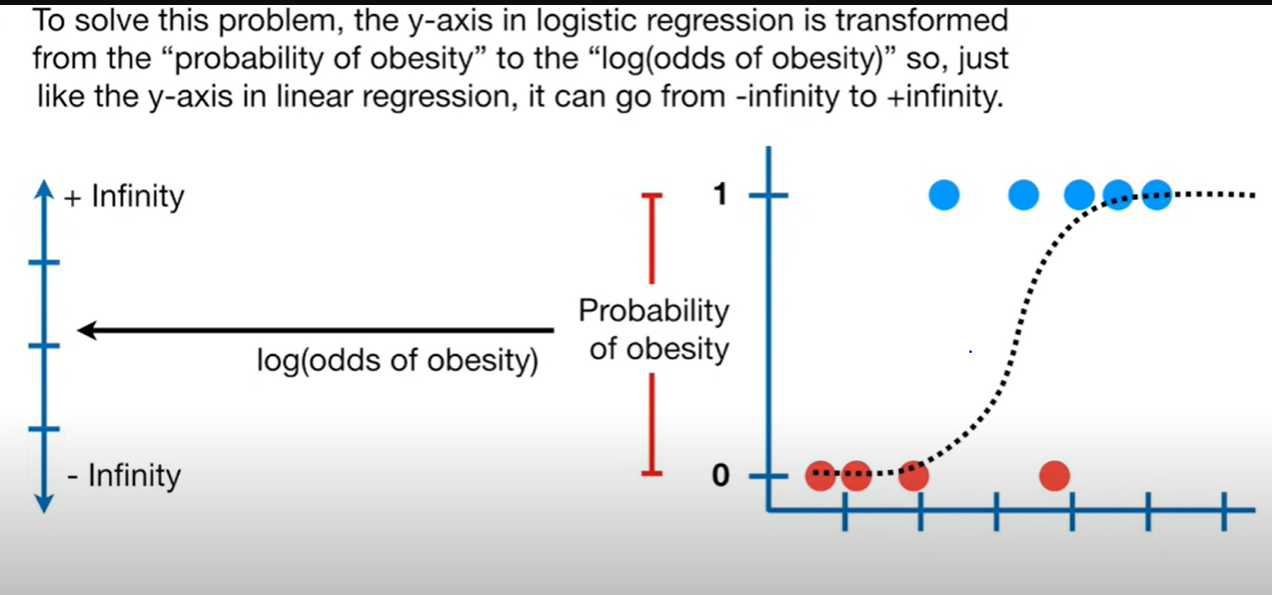
e = Euler’s Number = 2.7182

wkt g(z) is a number between 0 and 1 .



TRANSFORMATION FROM NON-LINEAR SIGMOID TO LINEAR CURVE





To carry out the transform of the sigmoid fn to a linear fn , we calculate the “log of odds” or “logit” value for each data set. Maths is explained below.

Let y = g(z) = probability of any outcome

In probability , definition of “odds” = Prob of event happening / Prob of event not happening

= y/(1-y) .

* In Logistic Regression y can be between 0 and 1 only, so for this let's divide the above equation by (1-y):

Logistic Regression in Machine Learning

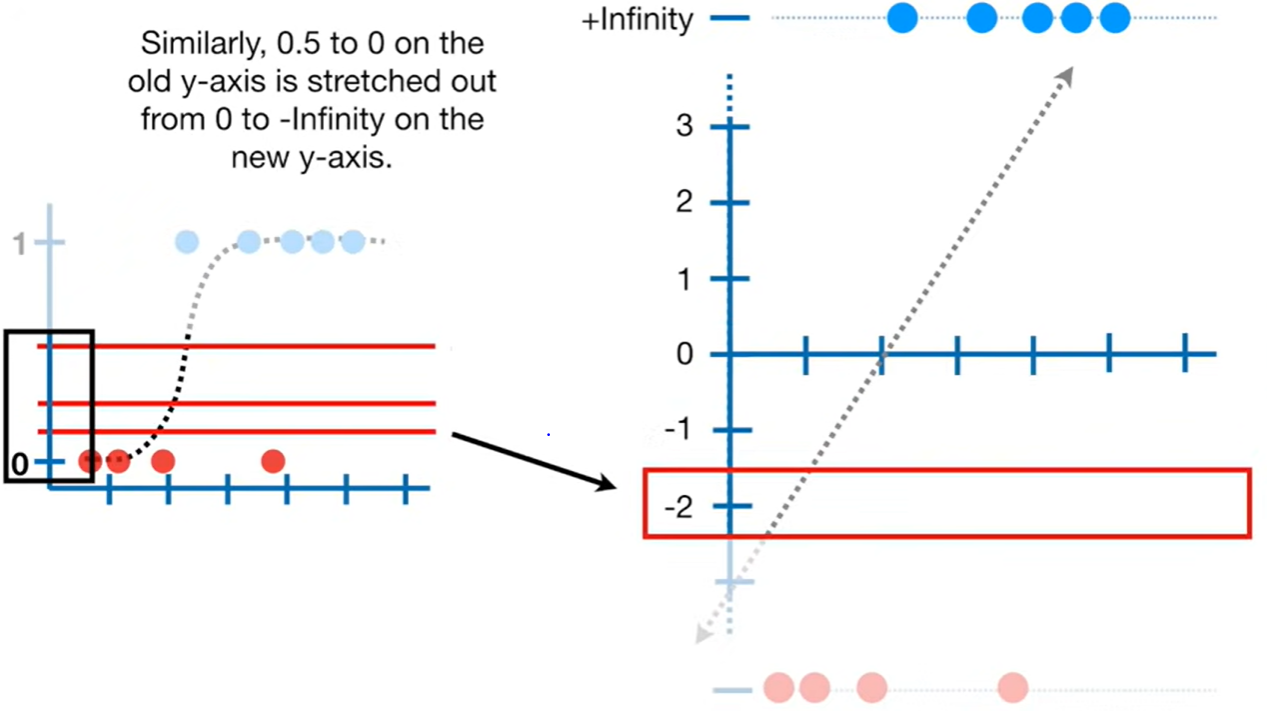
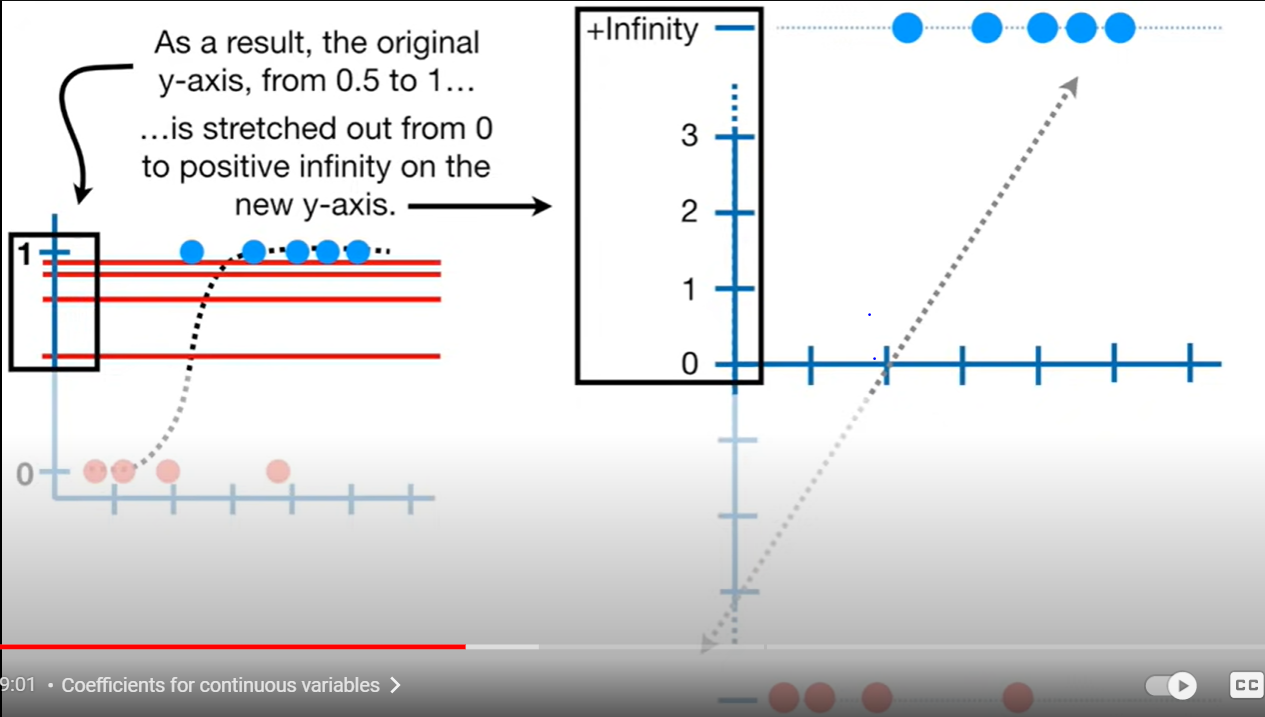
* But we need range between -[infinity] to +[infinity], then take logarithm of the equation to get the **logit** value or log(odds) value.

Logistic Regression in Machine Learning

**The above equation is the final equation for Logistic Regression**.

OBSERVE:

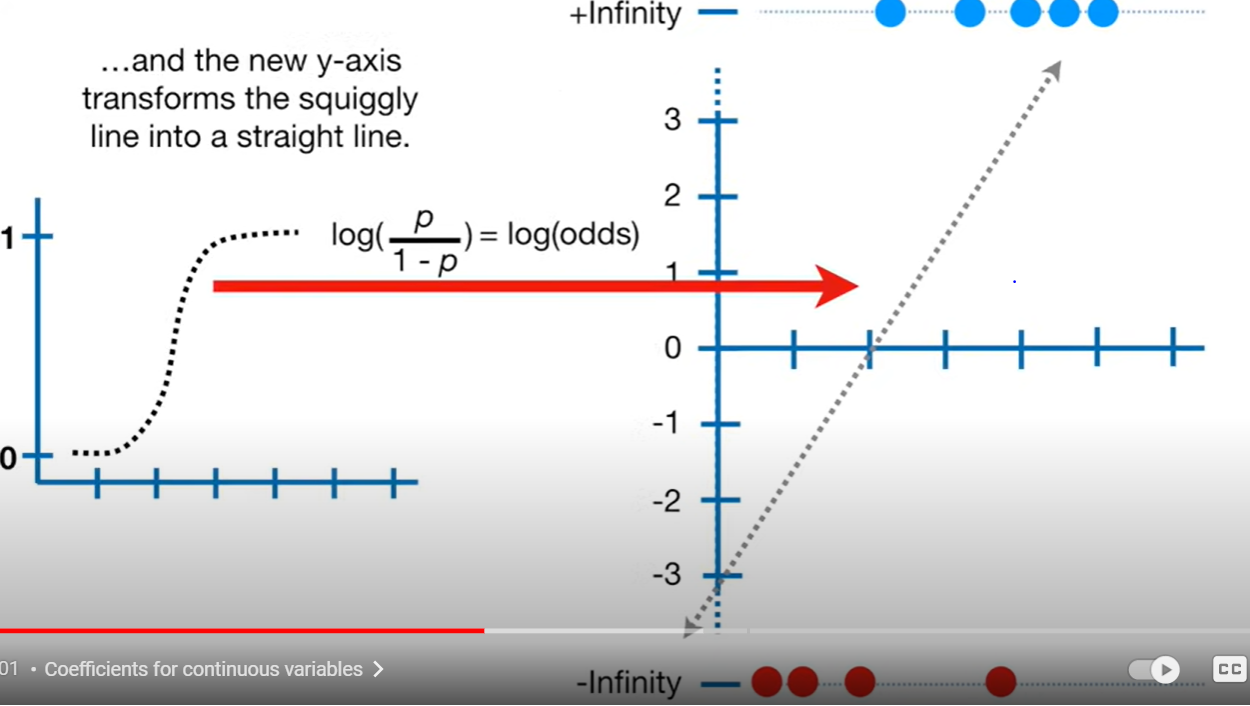
1. **The R.H.S. is a linear term and therefore a plot between “logit“ v/s x will be Linear.**
2. **After the transformation , the original pts (on S curve) from 0.5 to 1 will be stretched from 0 to +infinity on linear curve .**
3. **Also, after the transformation , the original pts (on S curve) from 0 to 0.5 will be stretched from -infinity to 0 on linear curve .**



1. Pts lying on 0 and 1(in red and blue dots) on the S curve will be represented at (-infinity) and (+infinity) respectively. (see above fig).

say at y = 1 , log(1/(1-1)) = log 1 /0 = log(+infinity) = +infinity

Slly say at y = 0, log(0/(1-0)) = log 0 = - infinity



1. Even though our Logistic Regression is done based on Sigmoid curve, **the Coeff of the model (b0,b1,b2……) is taken from this Linear Curve (or logit curve)**

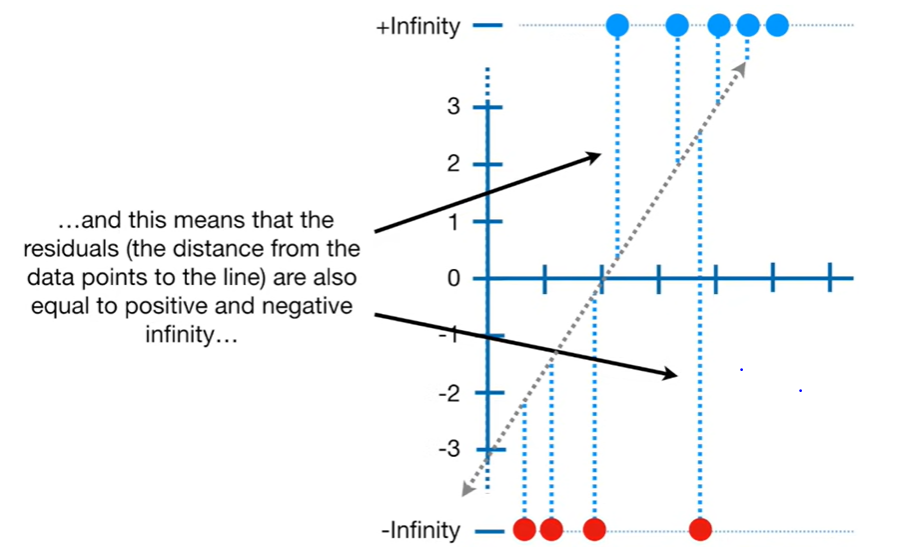
Logistic Regression in Machine Learning

**MAXIMUM LIKELIHOOD IN LOGISTIC REGRESSION**

To find the best fit line (and corresponding S curve ) for Logistic regression, we use the concept of Maximum Likelihood.

Note:

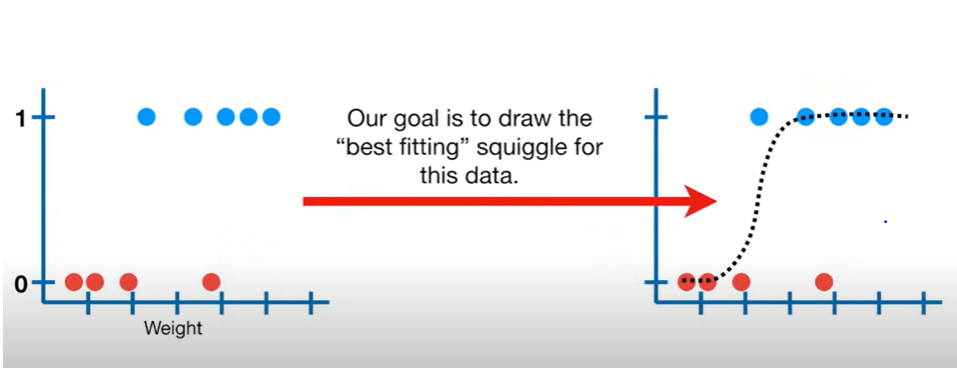
1. We cant use least squares technique to get best fit line because residual (y – distance from red and blue pts . to the fit line (see fig)) is infinity . Therefore we cant get residuals and therefore not minimize them to get best fit line, using least squares method. Hence we use the Max likelihood method.



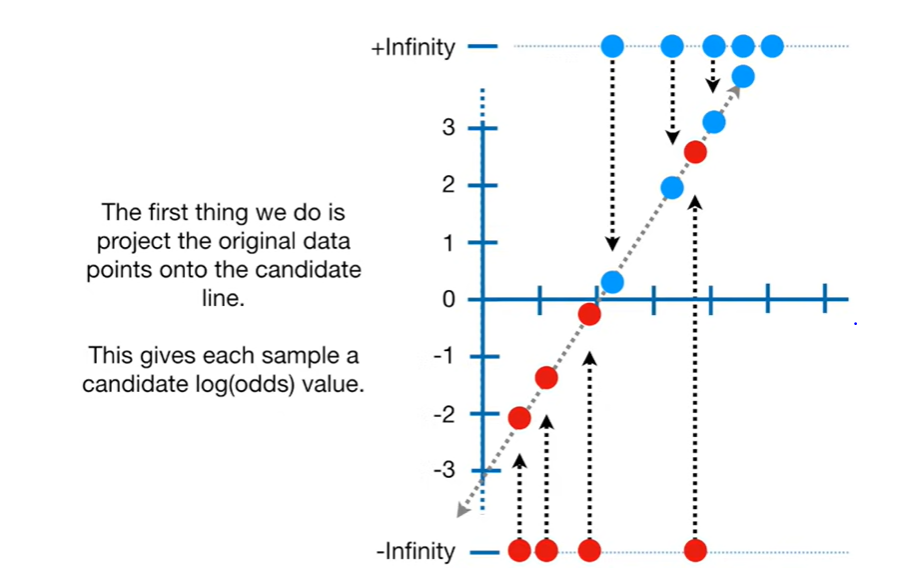
1. **We DO NOT go from best S-curve to best Linear curve.**

Instead we do this:

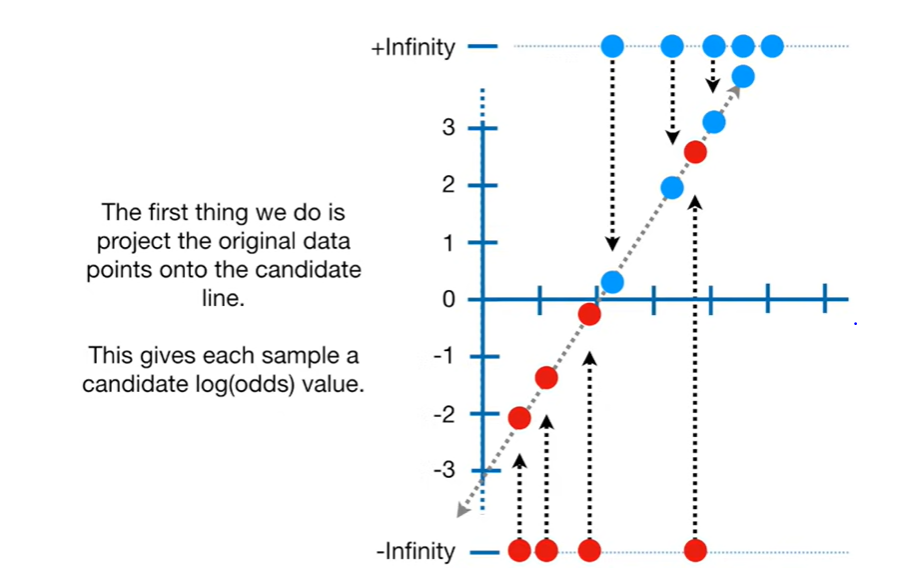
1. Goal is:



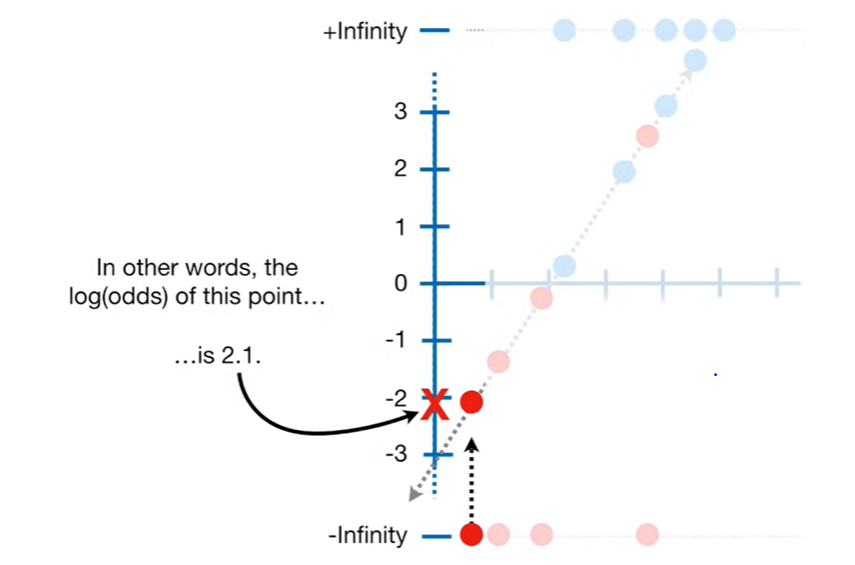
1. Now we use Maximum Likelihood to find the best fit line**. We start by placing these pts at +/- infinity (blue at +infinity and red at – infinity ), whereas there x -coordinates remain same .**
   1. Then, choose a random line as our candidate.(see dotted line below fig)



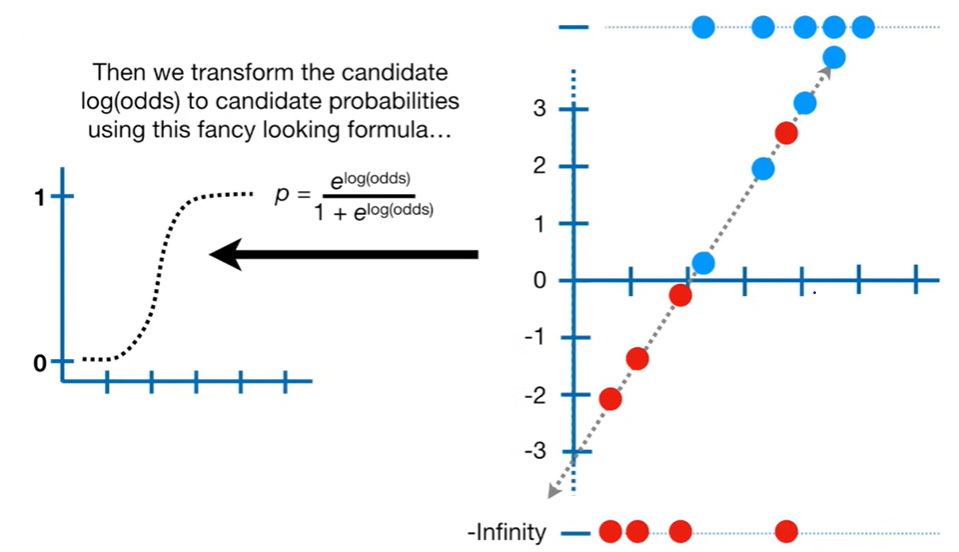
* 1. project data pts to line.
  2. This gives us log(odds) value corresponding to each sample .



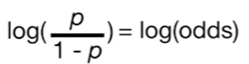
Example: for this red pt at the -infinity (see fig below) , the corresponding log(odds) value is -2.1.



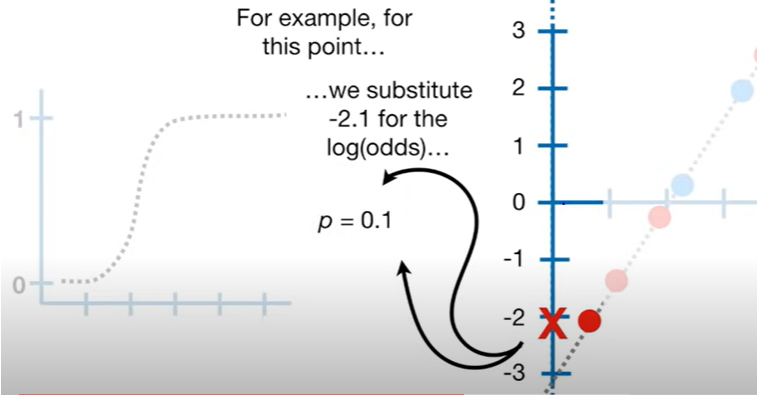
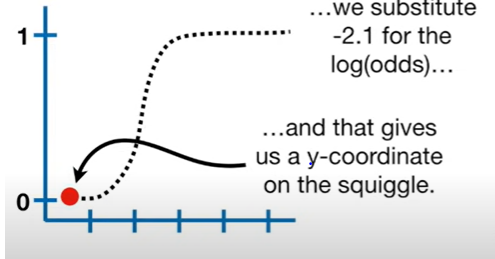
* 1. Slly we get log(odds) value for all data pts (red and blue both).
  2. Then we transform these log(odds) values to probabilities(=y value or g(z) for S curve) value using the formula shown in fib below.



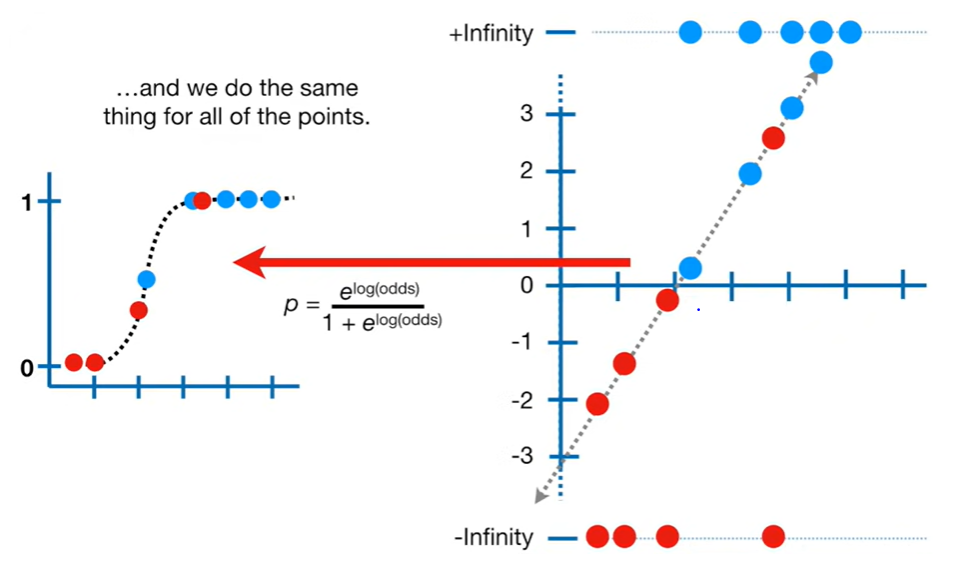
This formula is nothing but a re-ordering of the formula

 where RHS is known to us and we want to get the p = probability value = y (previously used notation) = g(z).

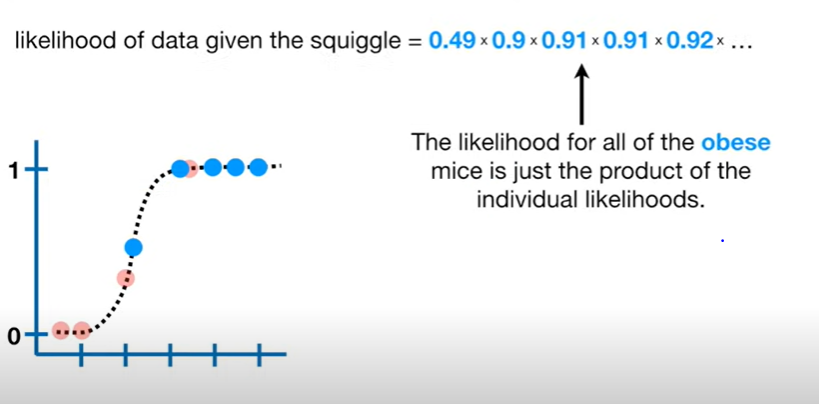
* 1. Example log(odds) = -2.1 , on solving we get p = 0.1(= y coordinate value for S curve)

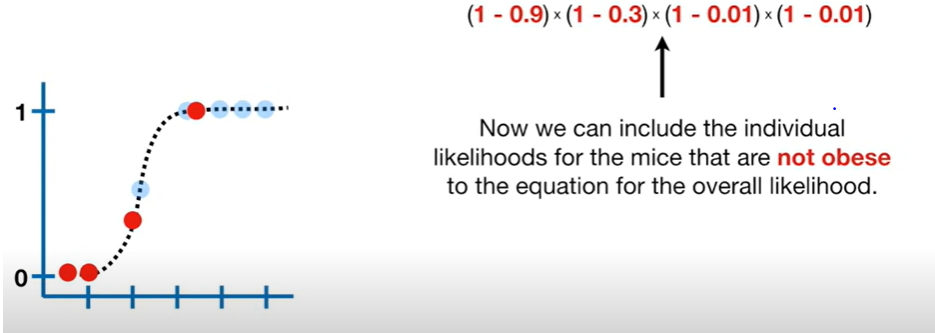
* 1. We do this for all the log(odds) value and plot them on the probability curve. We get the pts and then we



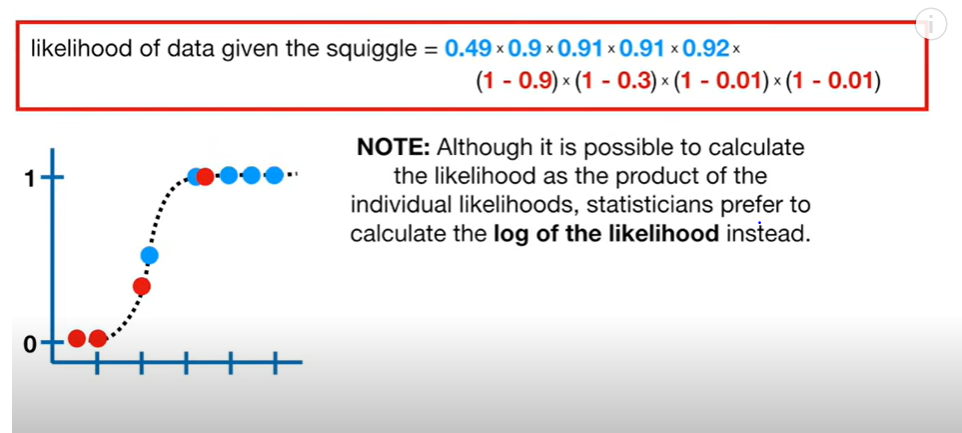
* 1. Calculate the Likelihood for all the mice by simply multiplying their individual probabilities of being obese and Not obese.

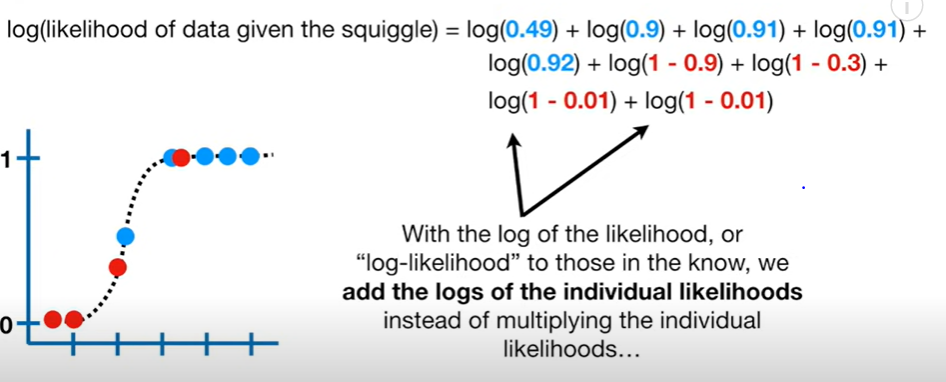
Example: for mice that are obese 

Example: for mice that are not obese. (They still have a prob. of being not obese which is simply = 1- prob(of obese)



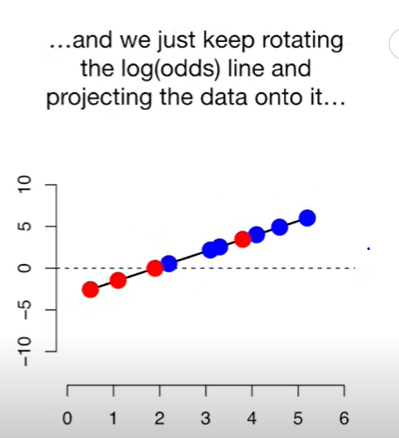
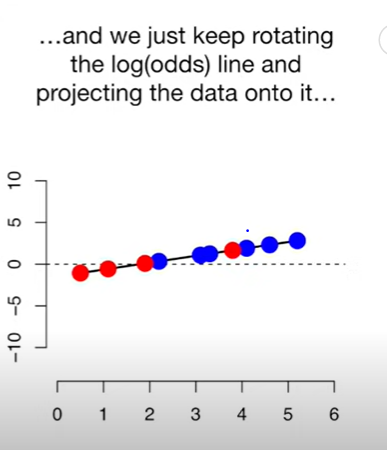
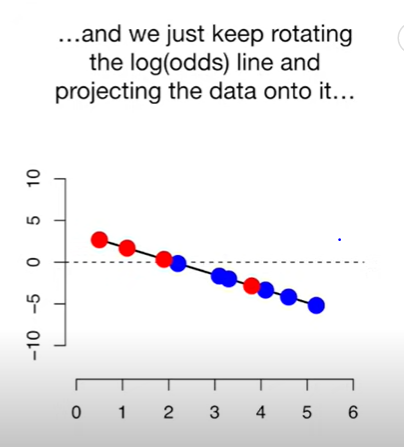
* 1. Thus, their overall Likelihood of the mice is :

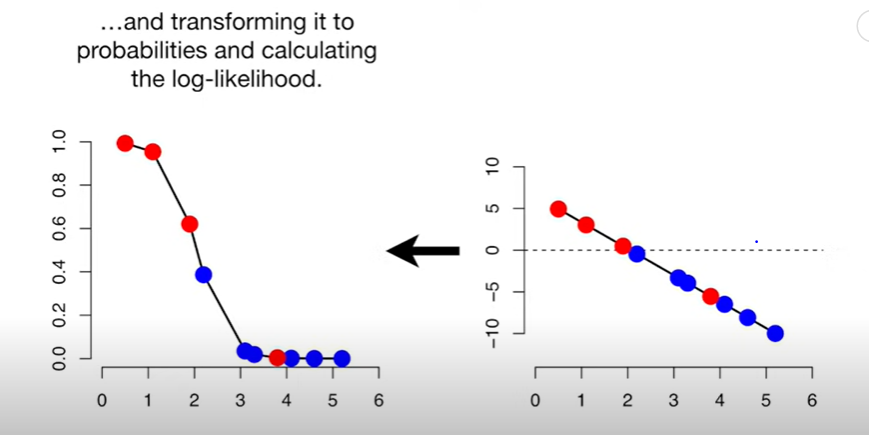




So , the Log(likelihood of data ) = -3.77 (using above data for example).

* 1. This means that Log(Likelihood ) of the original candidate line (linear line assumed at the beginning) is -3.71 .
  2. Now , we rotate the line (to get a new candidate line) and do the same calculations again to get the Log(likelihood) value for each Line.



* 1. **The optimum or BEST FIT line will result in the Highest value of Log(Likelihood ) or highest value of Likelihood.** This is done by the algorithm.

**R-Squared and its p value for Logistic Regression**

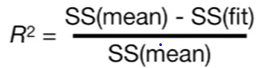
We know that the above line and s-curve obtained using Maximum Likelihood is Best Fit, **But , How do we know if they are Useful .i.e. we should also look at the relation b/w the weight and obesity** and see if they are at all related or Not. This is done using R-square and p value test (just as in Linear Regression ) .

However, **we wont be using Residual**, R-squared, to get these because here, the residuals have very large value (i.e. Infinity ) .

**NOTE**: There is no general consensus among Mathematicians on how to get R-squared value for Logistic Regression. We will be discussing one of the commomnly used method , although methods can vary .

Note:The method discussed below is below is very similar to what we learnt in Linear Regression to get R-squared value.

Recall the formula :

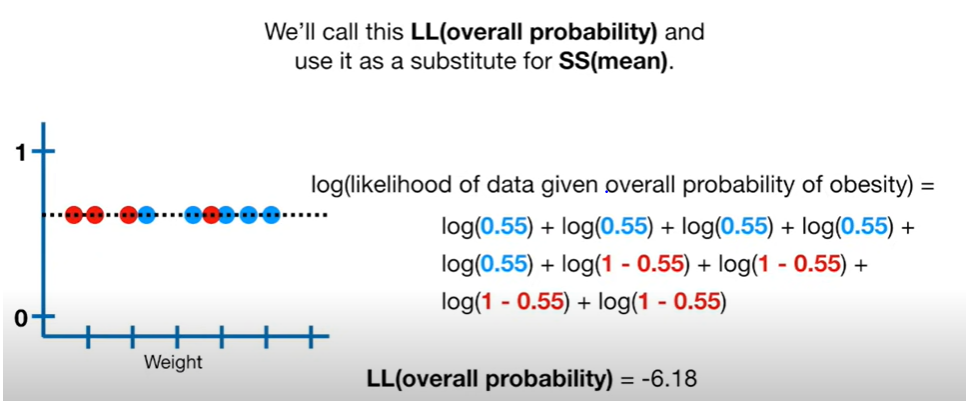


Method for Logistic :

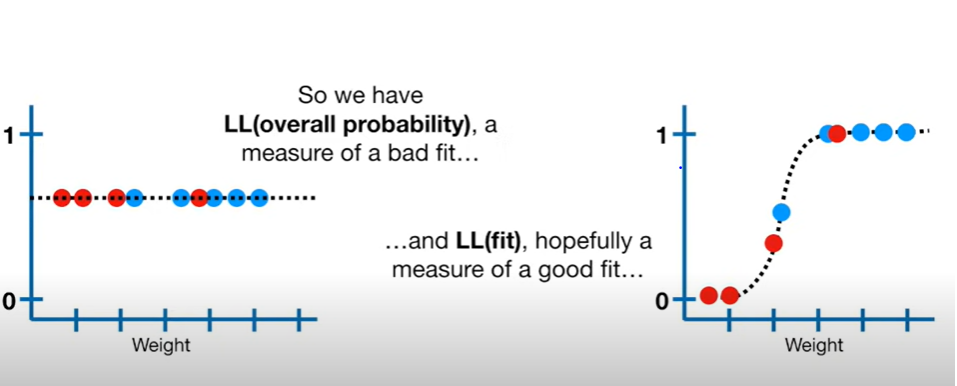
* 1. We replace **the SS(fit) value in above formula with LL(fit )** , which is nothing but Log(Likelihood) value of the (best) fit that we have earlier calculated. This LL(fit) is a measure of SS(fit). Say this LL(fit) = -3.77
  2. We calculate “the Overall Probability of Obesity “

i.e. Prob (odds of Obese) = p = (No of Obese mice/Total mice) = say , (5/9) = 0.56.

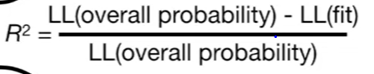
Plot it on the S curve.(see below fig), we get a horizontal Line at y = 0.56 .



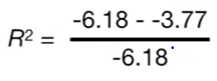
* 1. Now get the Log(Likelihood ) value (see above fig). This is used as **a measure of SS(mean) = LL(overall Probability )**
  2. Now we have the two values needed i.e LL(fit) and LL(overall prob) .



* 1. Now our New R-squared is given by



Plug in values we get R2  = 0.39



**NOTE:**

1. **Similar to Logistic regression, R2 values values go from 0 for poor models, to, 1 for good models .**

Example:

R2 = 1 implies 100% of variation in obesity is explained by the weight of the mice.

R2 = 0 implies No variation in obesity is explained by the weight of the mice.

1. A bad model will have a LL(fit ) value (obviously for best fit only, because R2 is being calculated for best fit only) relatively large negative value , whereas , a good model will have relatively lesser negative value or 0 value for LL(fit) (again for the best fit model only)

Ex: If LL(fit) = -3.7 and using different set of data we get LL(fit) = -4.5 , then obviously the 1st model is the better one. So, use the 1st data set to train your model.

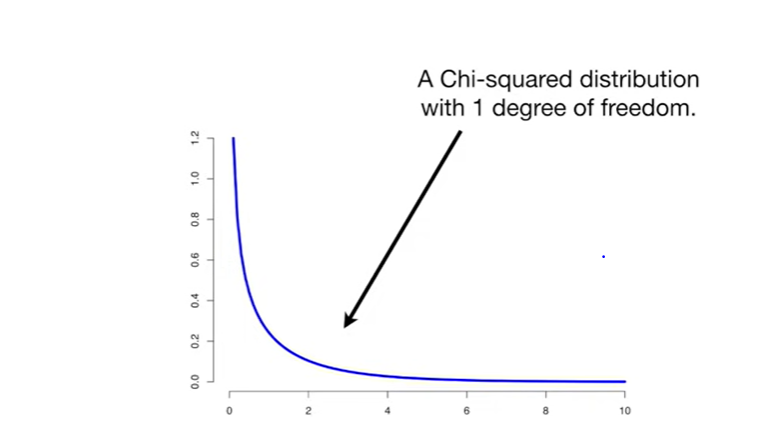
**P- value test for the R-Squared :**

Here, p-value is calculated using the **Chi-squared curve** (unlike the normal distribution curve used in Linear regression).

The d.o.f of the chi- squared curve is determined using

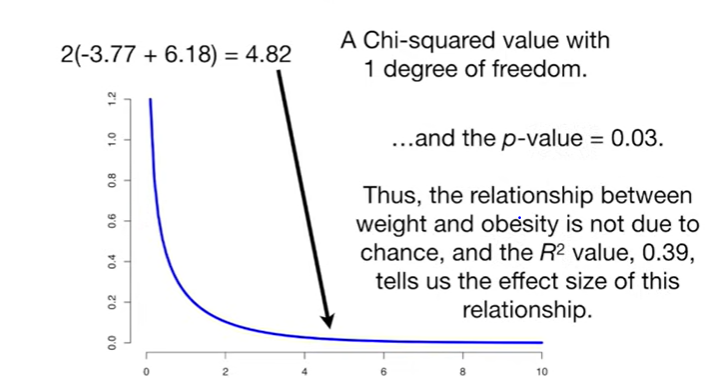
= No of Parameters (for LL(fit)) – No of Parameters(for LL(overall Prob.)

In previous case , d.o.f = 2-1 =1 . So, refer to the chi-squared chart with dof = 1.



The Value on the X-axis is given by : = 2(LL(fit) – LL(overall probability)) .

Considering our Prev. example’s values , x = 2(-3.77 - -6.18) = 2(-3.77 +6.18) = 4.82 and Corresponding **p-value is 0.03 .**



Recall:

**Typically, if p-value is less than 0.05 (p-value<0.05) ---> then, we Reject the null Hypothesis (Ho) and accept the Alternate Hypothesis (H1).**

## What is a null hypothesis?

For most tests, the null hypothesis is that there is no relationship between your variables of interest or that there is no difference among groups .i.e here Null hypothesis Ho : is that there is No relation between Weight and Obesity .

Thus, since our obtained R-squared value is 0.03 (<0.05) so, we reject Ho and conclude that our data set does have a relationship b/w weight and obesity and it is not due to coincidence. Hence, it is a good model .