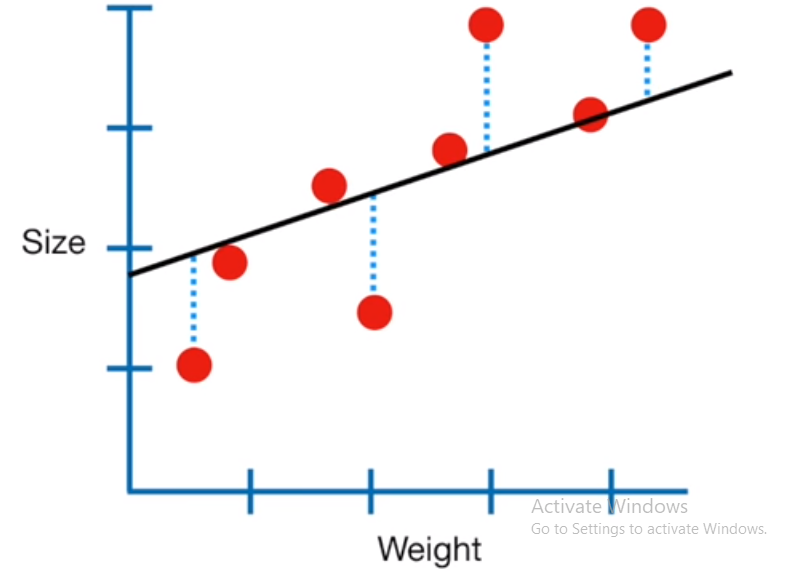
**Logistic Rgression: Fitting a line with Maximum Likelihood**

* Quick review with the linear regression-



* Line with the smallest sum of squared residuals (the lease squares) is the line chosen to fit.

Q- same least square method is suitable in case of classificatiopn problem (Yes/No)? No but why????

***Lets talk about the Logistic Regression……***

***(Prerequisit- Introduction of Odds and Odds Ratio (OR))***

Let’s begin with probability. Probabilities range between 0 and 1. Let’s say that the probability of success is .8, thus

**p = .8**

Then the probability of failure is

**q = 1 – p = .2**

Odds are defined as the ratio of the probability of success and the probability of failure. Odds are determined from probabilities and range between 0 and infinity. The odds of success are

**odds(success) = p/(1-p) or p/q = .8/.2 = 4,**

that is, the odds of success are 4 to 1. The odds of failure would be

**odds(failure) = q/p = .2/.8 = .25.**

This looks a little strange but it is really saying that the odds of failure are 1 to 4. The odds of success and the odds of failure are just reciprocals of one another, i.e., 1/4 = .25 and 1/.25 = 4.

**Another example**

This example is adapted from Pedhazur (1997). Suppose that seven out of 10 males are admitted to an engineering school while three of 10 females are admitted. The probabilities for admitting a male are,

**p = 7/10 = .7**

**q = 1 – .7 = .3**

If you are male, the probability of being admitted is 0.7 and the probability of not being admitted is 0.3.

Here are the same probabilities for females,

**p = 3/10 = .3**

**q = 1 – .3 = .7**

If you are female it is just the opposite, the probability of being admitted is 0.3 and the probability of not being admitted is 0.7.

Now we can use the probabilities to compute the odds of admission for both males and females,

**odds(male) = .7/.3 = 2.33333**

**odds(female) = .3/.7 = .42857**

Next, we compute the odds ratio for admission,

**OR = 2.3333/.42857 = 5.44**

*Thus, for a male, the odds of being admitted are 5.44 times as large as the odds for a female being admitted.*

Range (oddss ratio)= 0🡨🡪 + infinity

In view of classification-

**Objective function (vauge):** try to maximize odds ratio for some set of Inputs (assume to belong in one class) and minimise some other set of inputs (assume to belong in another class).

In this view, we require a function that give direct ralationship between inputs (features) and odds ratio.

f(x) 🡨🡪 Odds Ratio

View as independent and dependen variable.

Funadmetally, we are trying to draw a line that will separate the observation in groups.

Line (β0+β1∗x1+...+βn∗xn ) 🡨🡪 Odds Ratio

Odds Ratio 🡨 🡪 Line (β0+β1∗x1+...+βn∗xn )

But, how to define that relatioship mathematically,

Log (Odds Ratio) 🡪 β0+β1∗x1+...+βn∗xn --------eq. (A)

Suppose, z=β0+β1∗x1+...+βn∗xn

Log (Odds Ratio) = z

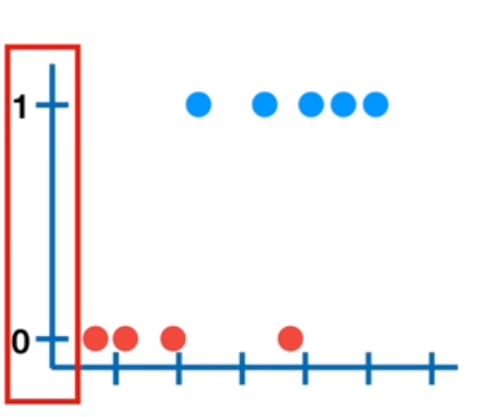
It is called logit and the range is negative infinity to positive infinity. The logit transformation allows for a linear relationship between the response variable and the coefficients.

response variable: logit

Independent variable: coefficients(line)

**Objective function (refined):**

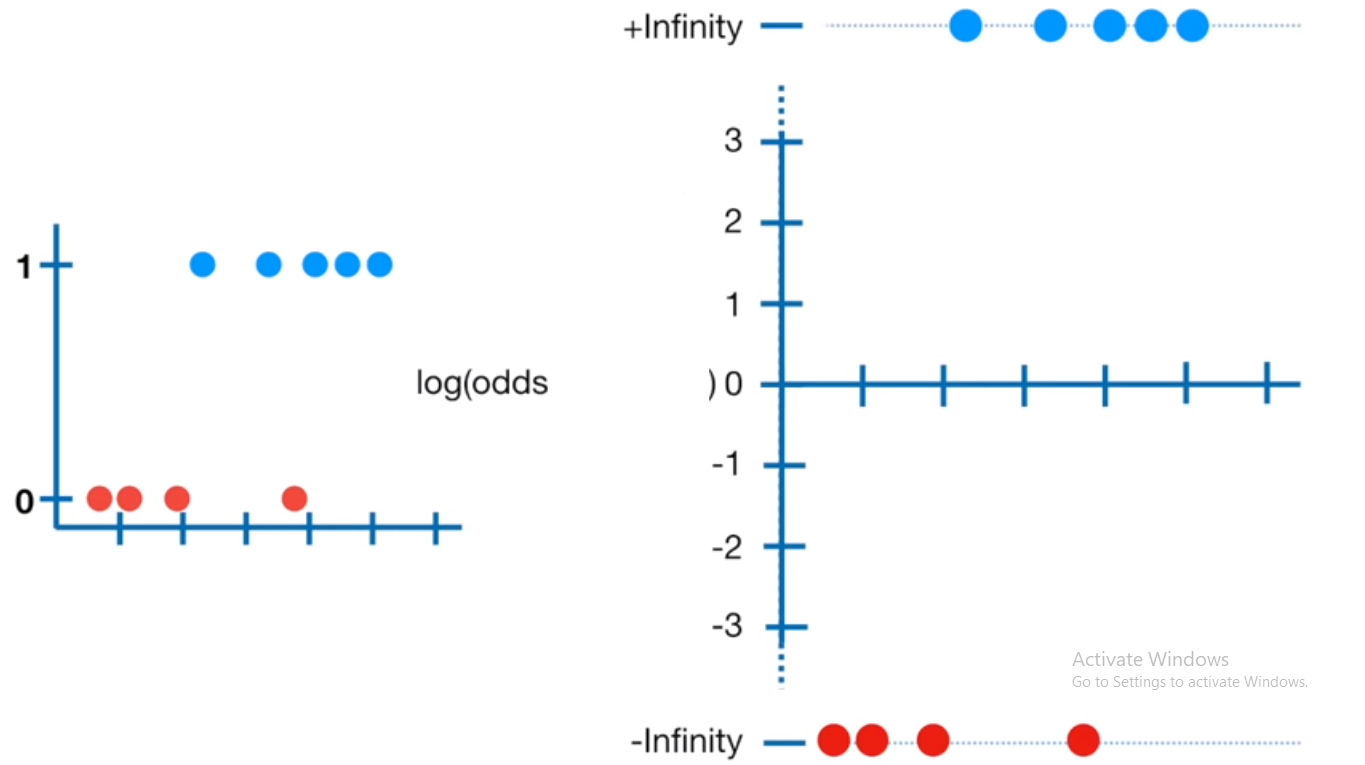
In term of probability, it should be enough high for some set of observations and vice versa.



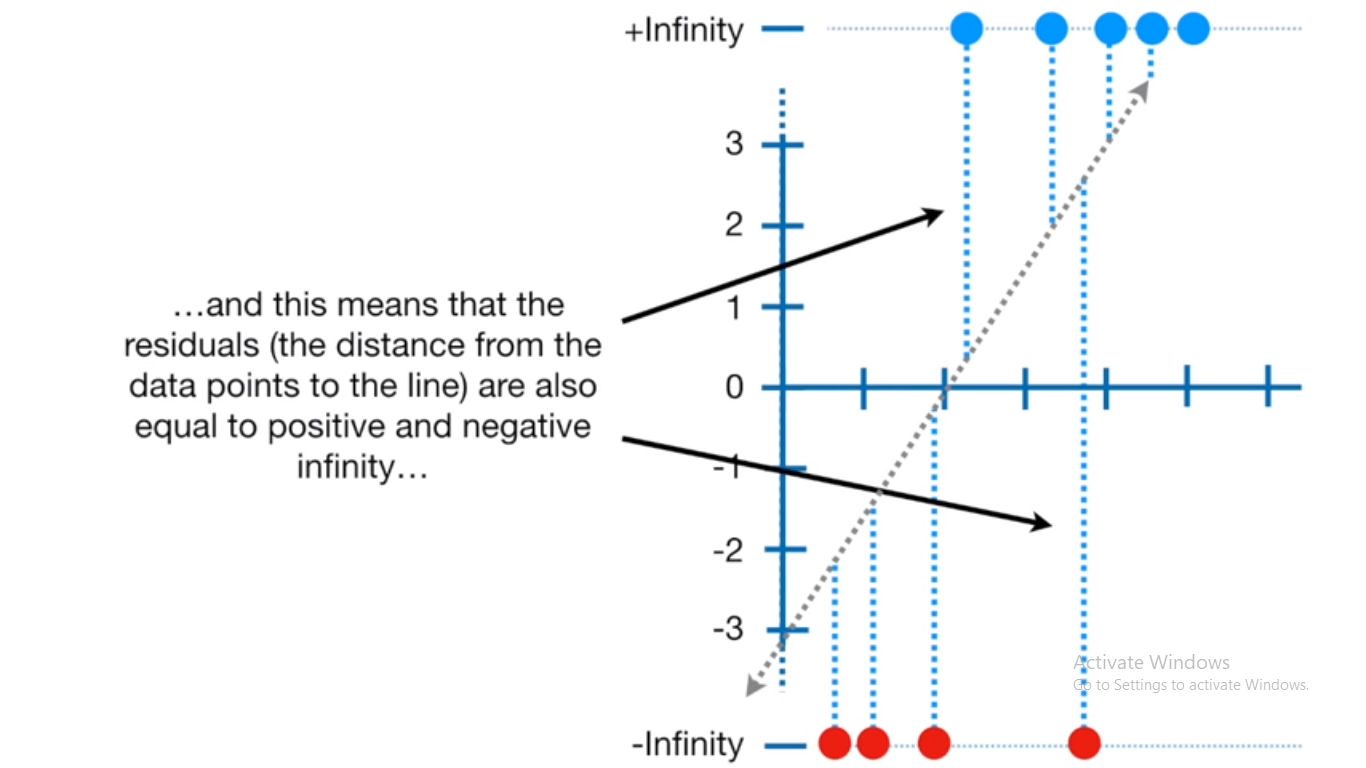
But , logit (Log (Odds Ratio)) range from-

–infinity 🡨🡪 + infinity

Response variable range= –infinity 🡨🡪 + infinity



Task: try to fit a line using *least square* method against response variobale which range from –infinity to +infinity.



Conclusion: we can not use least squres method to fit the best fitting line. Instead we use maximum lekelihood.

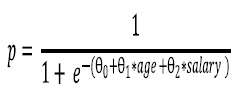
Meanwhile, we need arrangement that given output in term of pobabaility.

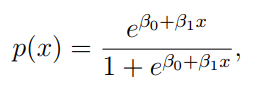
 ln[p/(1−p)] =β0+β1∗x1+...+βn∗xn

ln[p/(1−p)] = z

p=1/1+exp^−z,

where z=β0+β1∗x1+...+βn∗xn



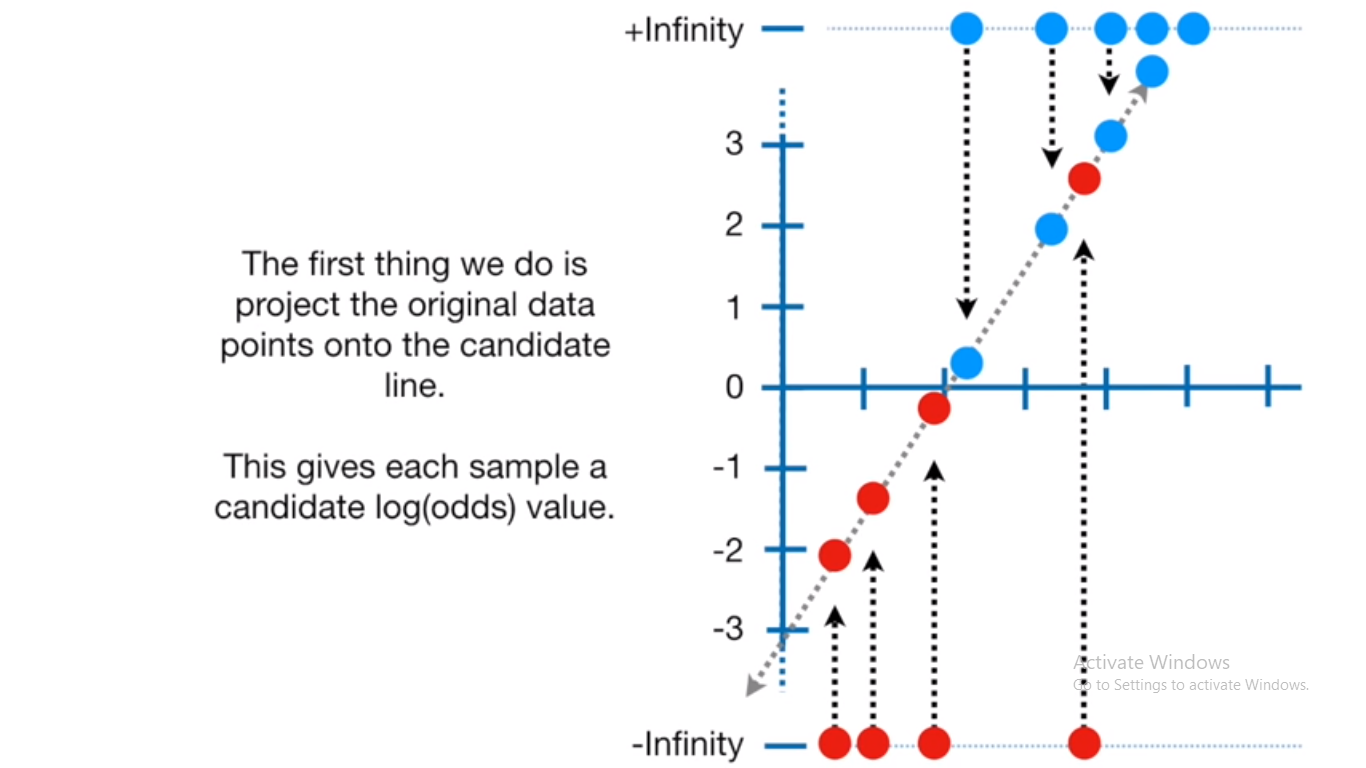


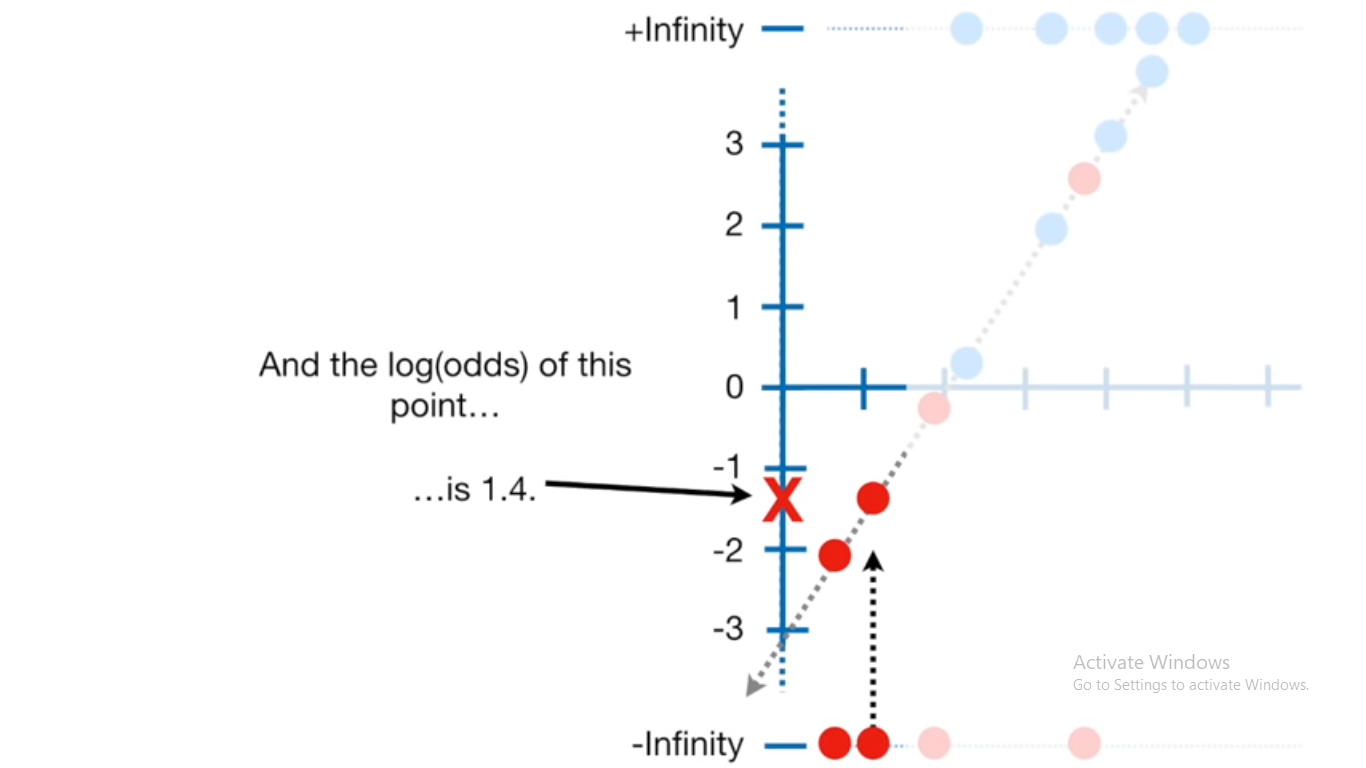
**Working of likelihood function-**

We had equation:

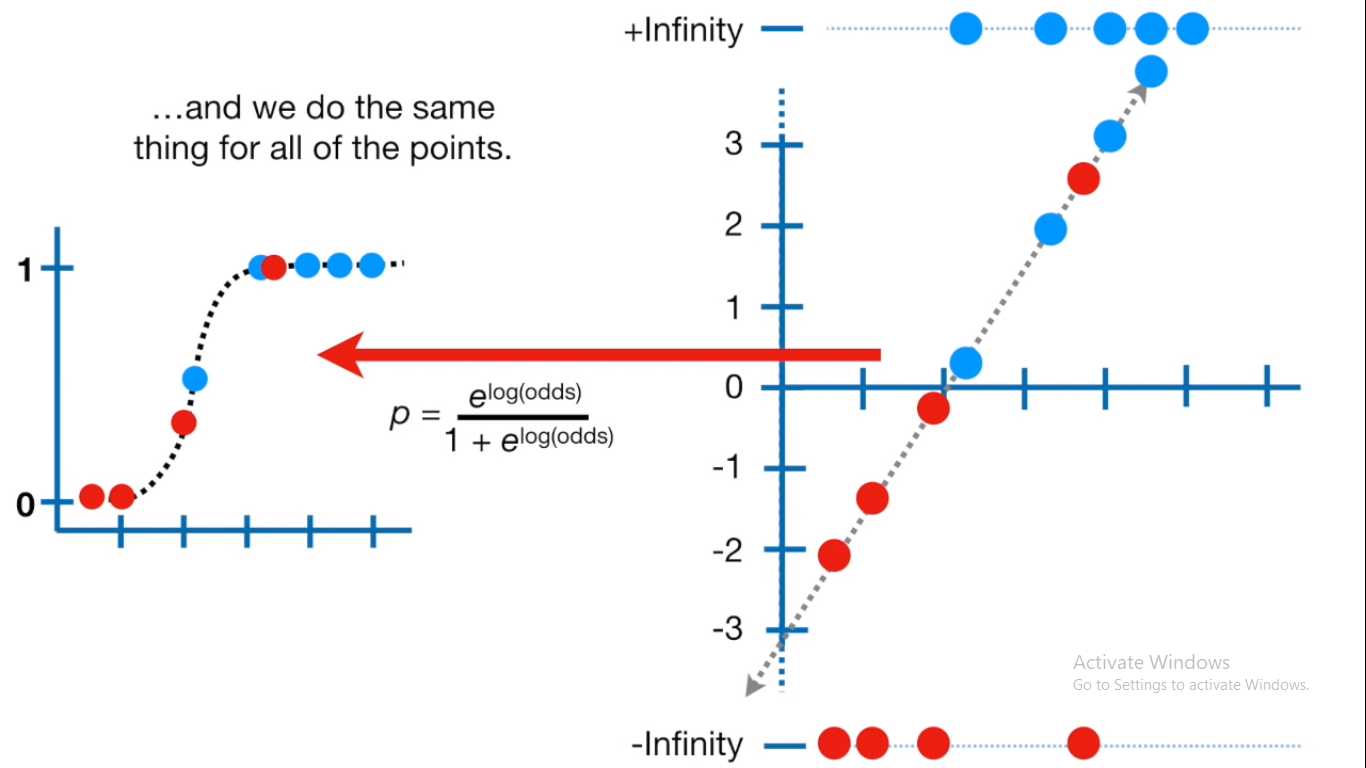
log (odds ratio)= Z

odds ration = ez = e β0+β1∗x1+...+βn∗xn

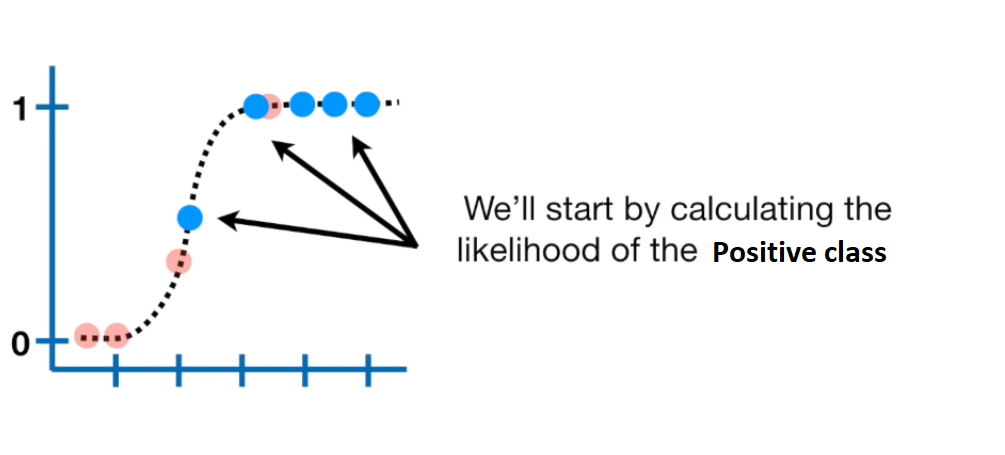


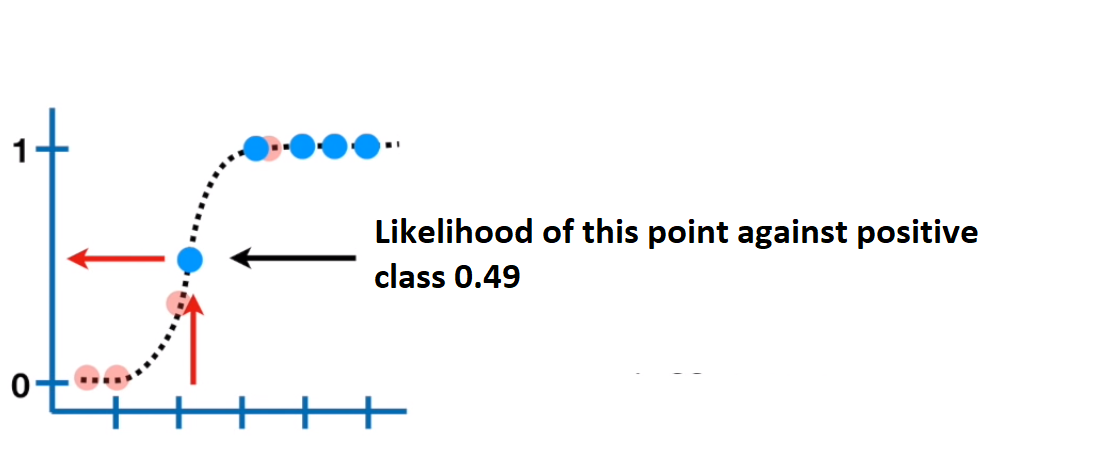


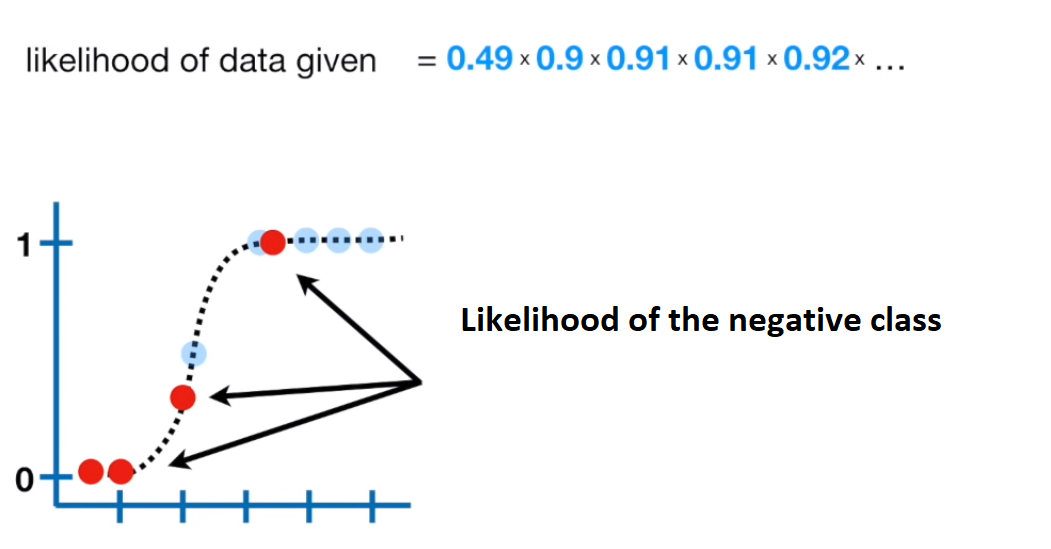
P= ez/ (1+ ez)

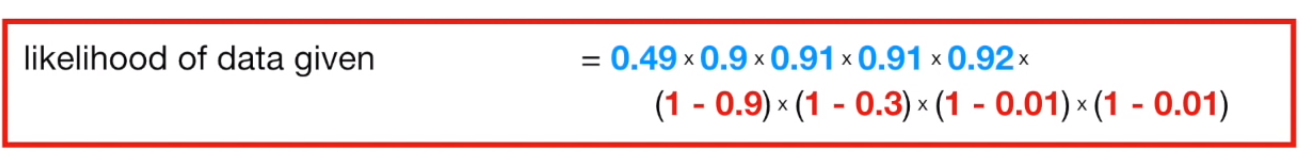


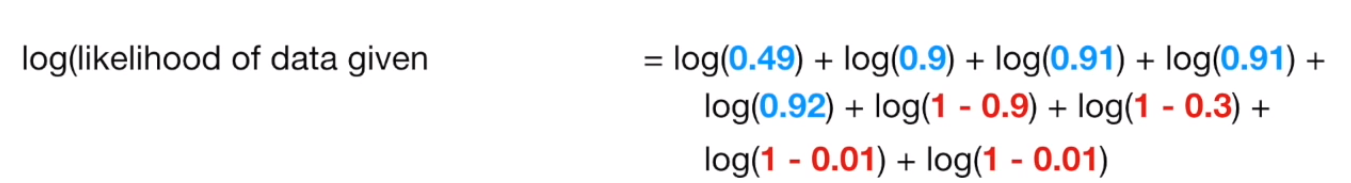
Repeat the same for every observations

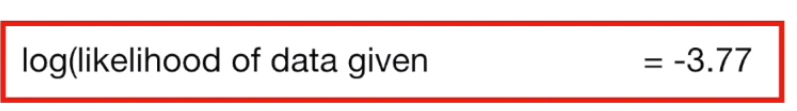


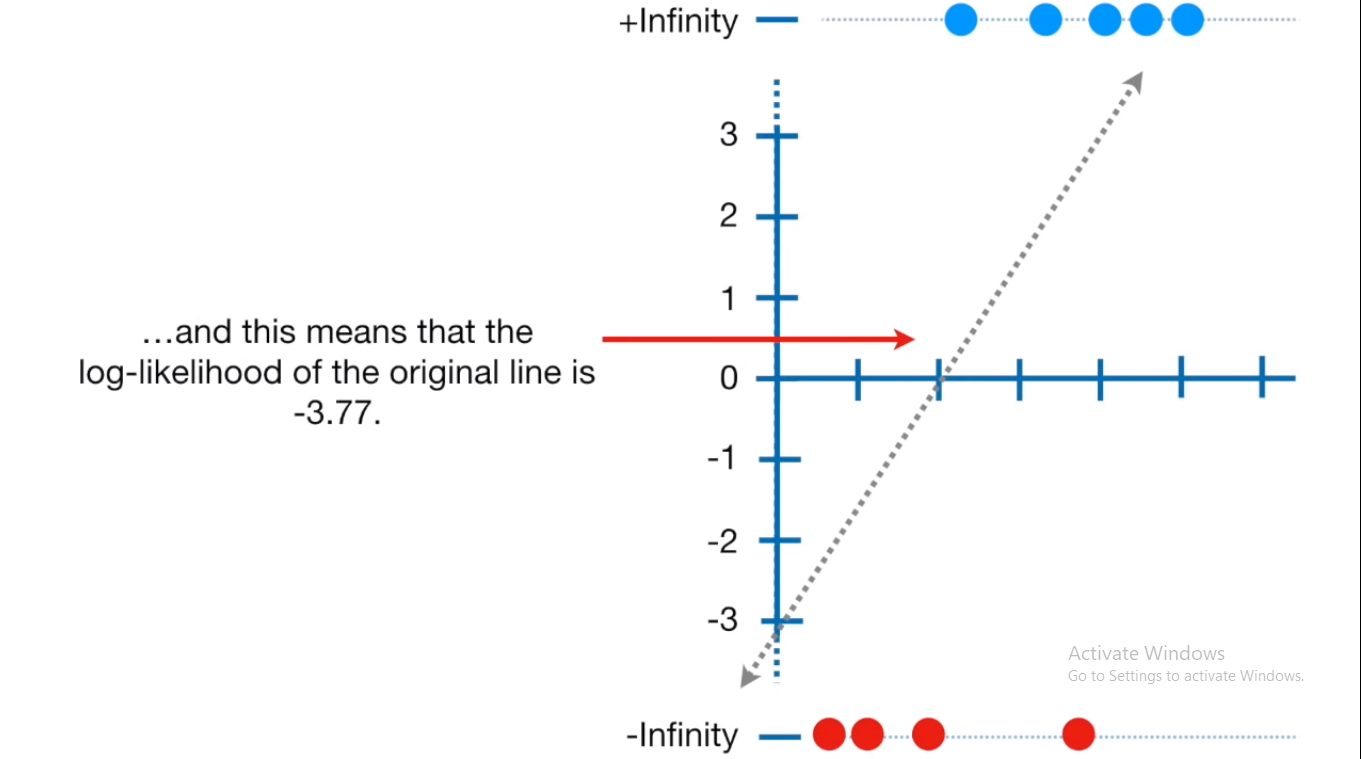




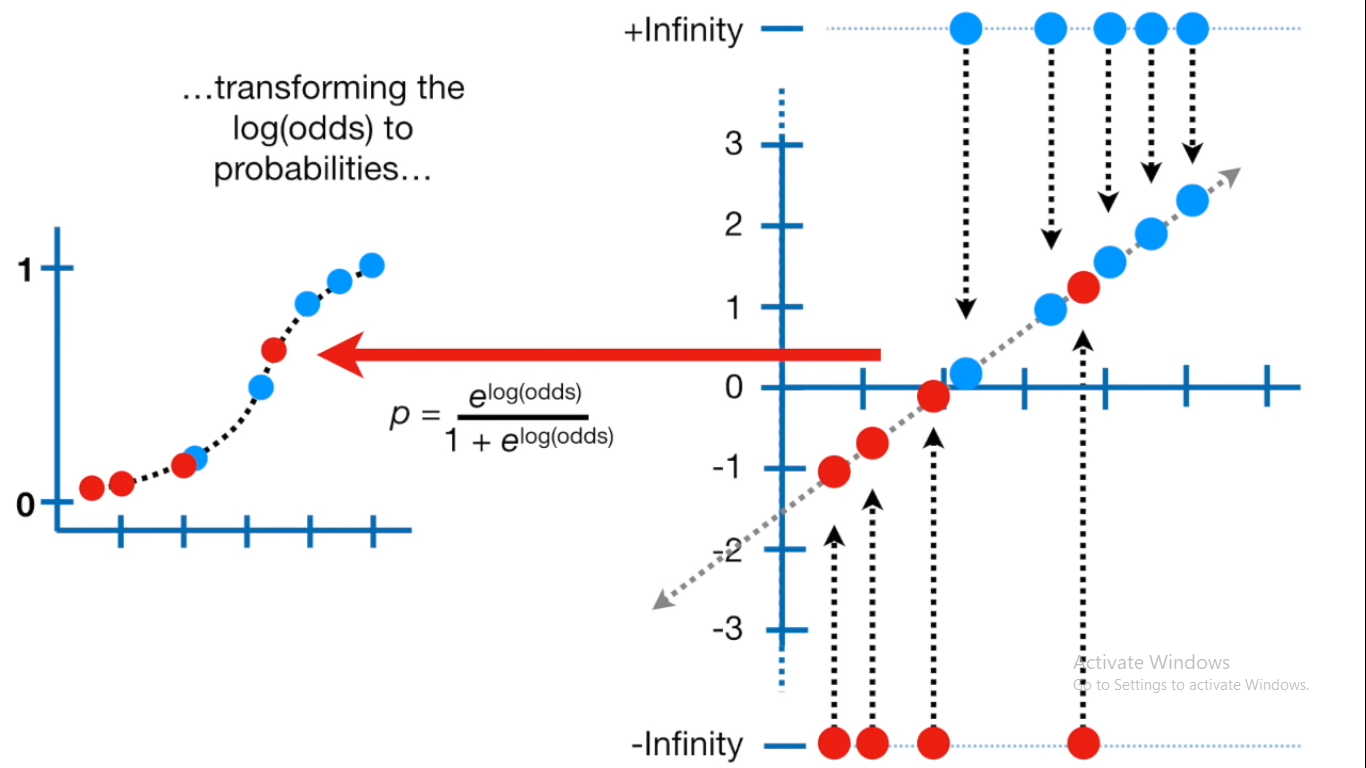








Rotate the line and again calculate the likelihood and try to maximize it.



Likelihood Ratio (LR) Test

log LR-Test

* Used to compare two nested regression model
* When one model (model-1) is nested within another mode (model-2)
* Model-2 is unrestricted model with p2 parameters
* Model-1 is restricted model with p1 parameters
* Null Hypothesis: restricted model (model-1) is statistically better than unrestricted model (model-2)
* Number of restriction: Q=p2-p1

Formula:

2(loglikelihood\_unrestricted\_model - loglikelihood\_restricted\_model)

-2(loglikelihood\_restricted\_model -loglikelihood\_unrestricted\_model )

* LR Test follow chi-square distribution with Q dof.

Study: Compare two regression model with LR-Test:

x\_reg1 = lm (y~x)

logLik(x\_reg1) Restricted model (Model-1)

suppose: 5430.173 (df=3)

x\_reg2=lm (y ~ x1+x2+x3)

logLik(x\_reg2) unrestricted model (Model-2)

suppose: 5432.885 (df=5)

Put in formula

2\*(model2-model1)

2\*(5432.885-5430.173)

= 5.424 ----------- A

LR-Test critical value

qchisq(% of confidence, Q (q2-q1))

qchisq(0.95, 2)

= 5.9914 ---------B

Interpretation: since LR test (computed) < LR test (critical)

5.424<5.99

So, we cannot reject hypothesis, and so restricted model (model-1) is better.

Alternate

The likelihood ratio test can be performed in R using the lrtest() function from the lmtest package or using the anova() function in base.

library(lmtest)

lrtest(mod\_fit\_one, mod\_fit\_two)

i.e. sequence does not matter

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | #df | logLik | DF | Chisqu (LR-Test) | Pr>chisq |
| Model-1 | 3 | 5430.2 |  |  |  |
| Model-2 | 5 | 5432.9 | 2 | 5.424 | 0.6642 |

P-value >0.05, we cannot reject null hypothesis, model-2 is not significant

anova(mod\_fit\_one, mod\_fit\_two, test ="Chisq")

A logistic regression is said to provide a better fit to the data if it demonstrates an improvement over a model with fewer predictors (restricted). This is performed using the likelihood ratio test, which compares the likelihood of the data under the full model (unrestricted) against the likelihood of the data under a model with fewer predictors (restricted). Removing predictor variables from a model will almost always make the model fit less well (i.e. a model will have a lower log likelihood), but it is necessary to test whether the observed difference in model fit is statistically significant. Given that H0 holds that the reduced model is true, a p-value for the overall model fit statistic that is less than 0.05 would compel us to reject the null hypothesis. It would provide evidence against the reduced model in favor of the current model.