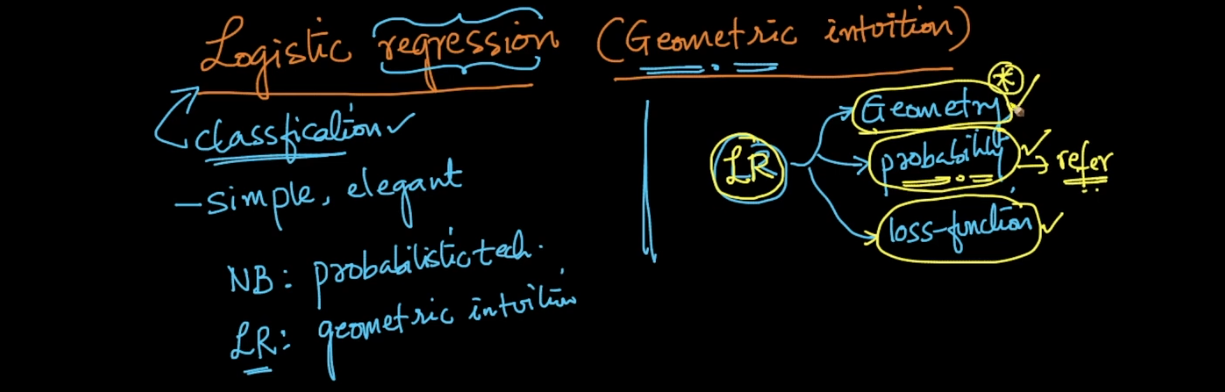
Logistic Regression :

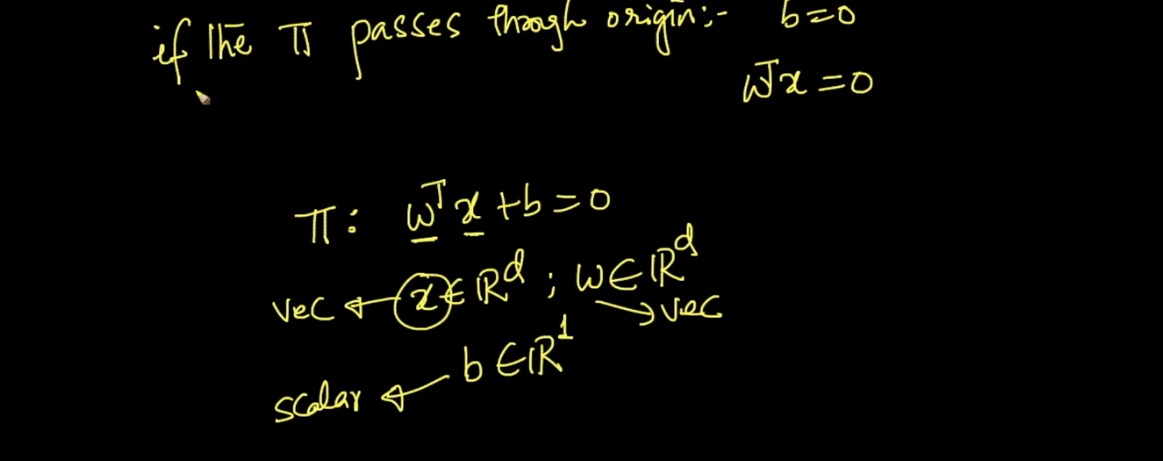


So in logistic regression our job is to chose appropriate plane means w .



W t x + b if line passing through then b will be 0 .

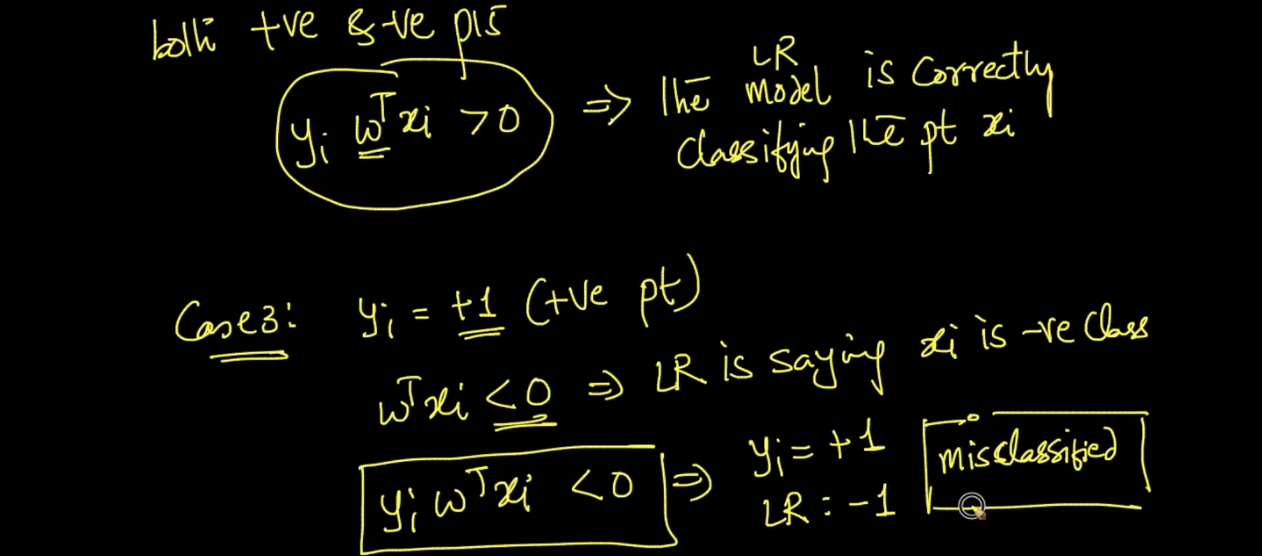
So main task is to find perfect plane which can easily separate data .

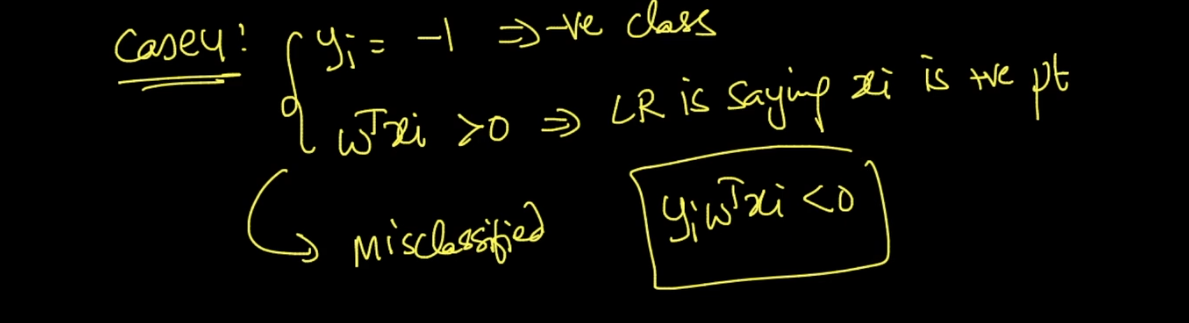




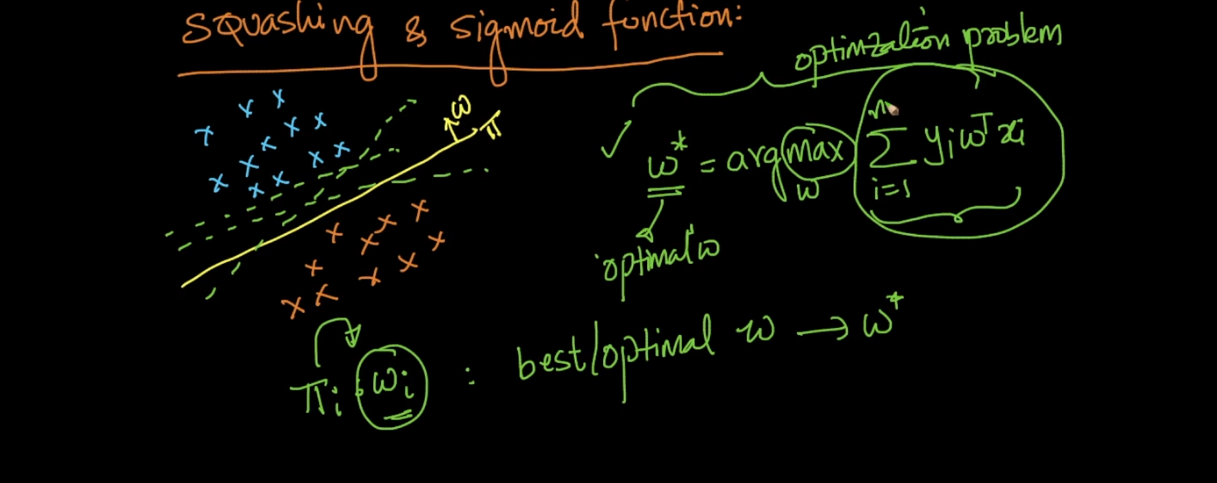
See above image lets take new 2 points x1 and x2 .now x1 is in same direction of w hence it will + and x2 opposite .







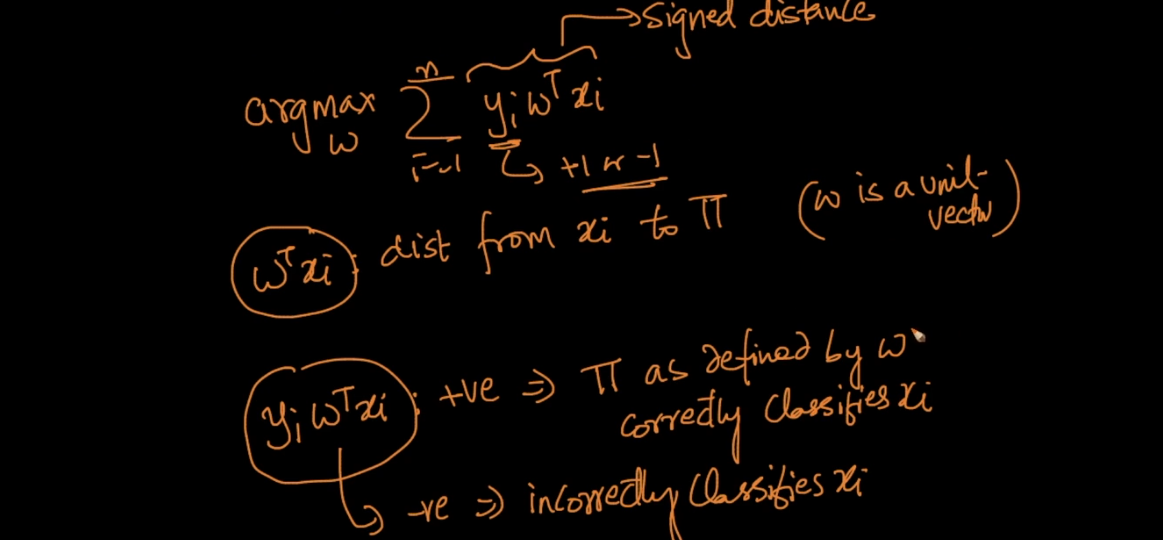
So finally we need best w so we can separate max points .



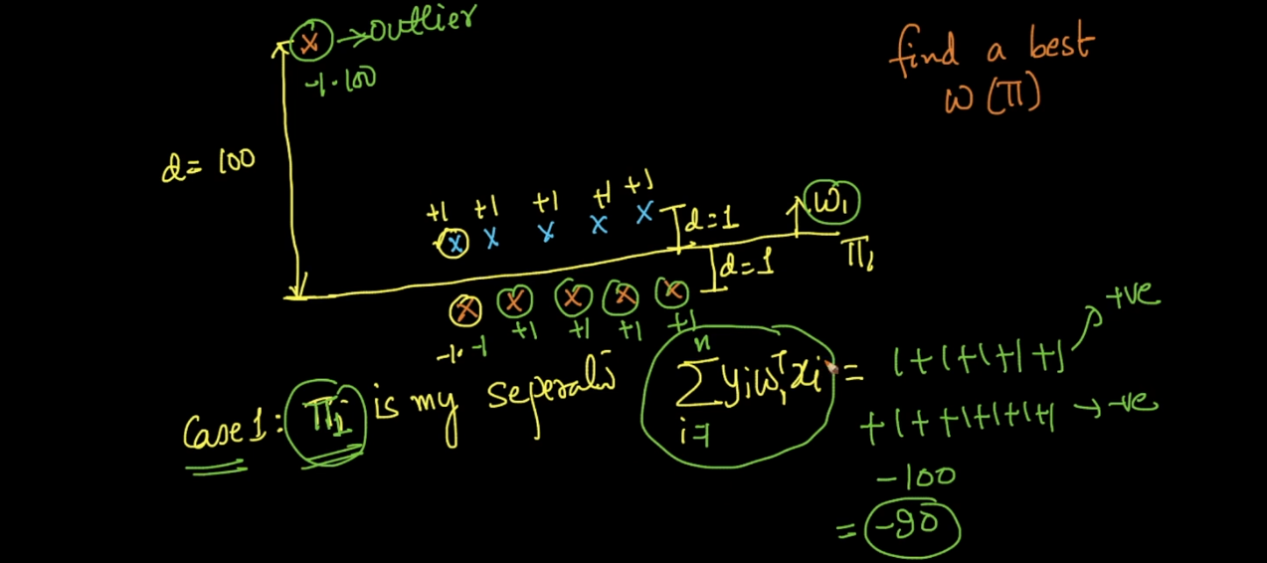
So see above image we found function of w we want to find max signed sum so we will get plane n good direction that’s what we study .

But there is a big problem of this concept with out l .it will perform very very bad . lets see below images .

Y will be +1 or -1



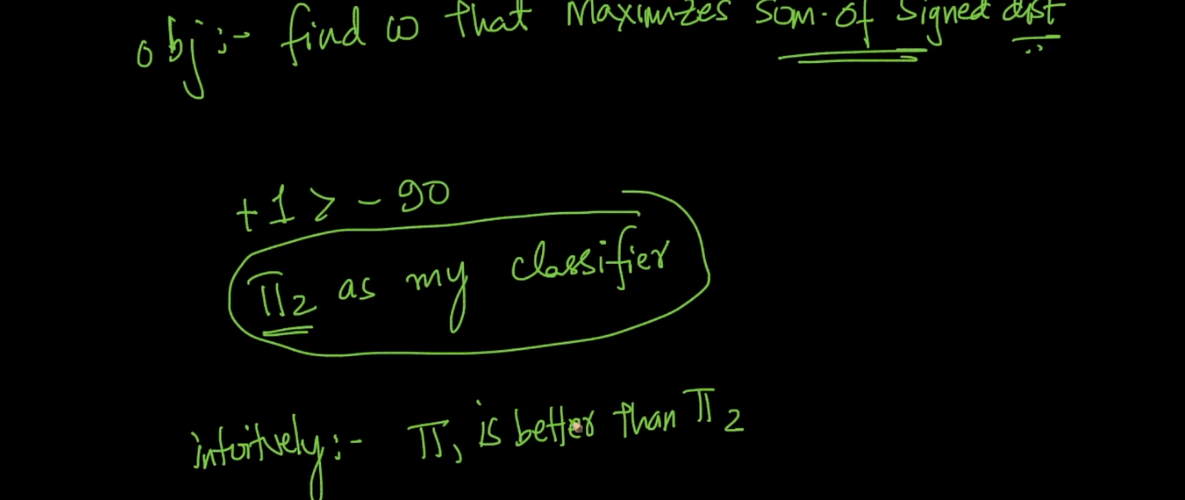
See case 1 :



Here we get very very small signed sum just because of only 1 out l.

Since we get very bad value we will move or plane so next we get :

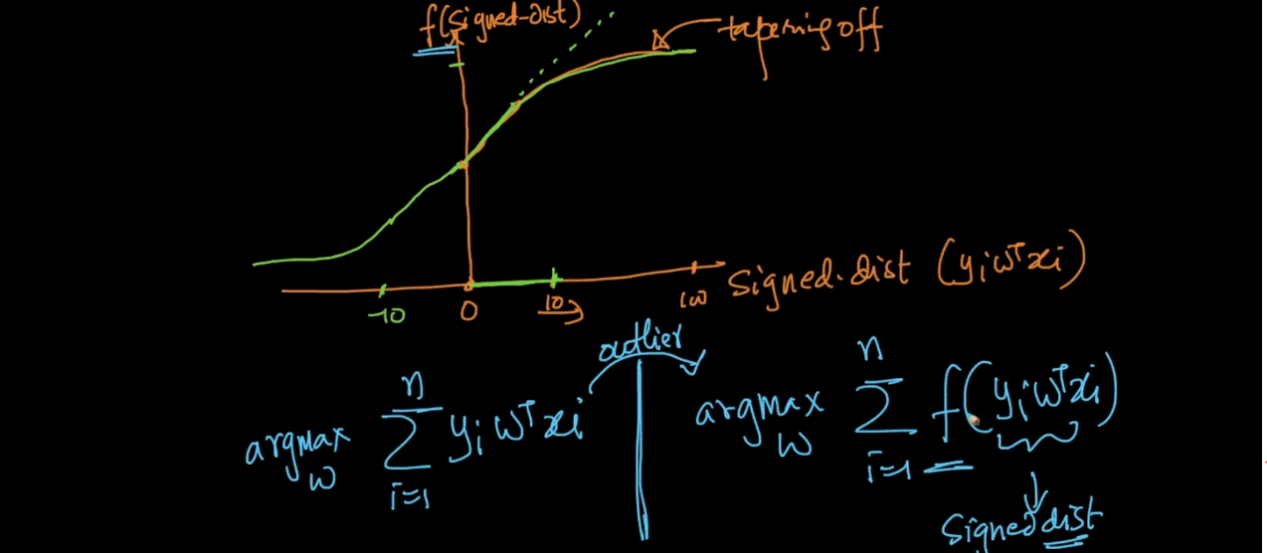




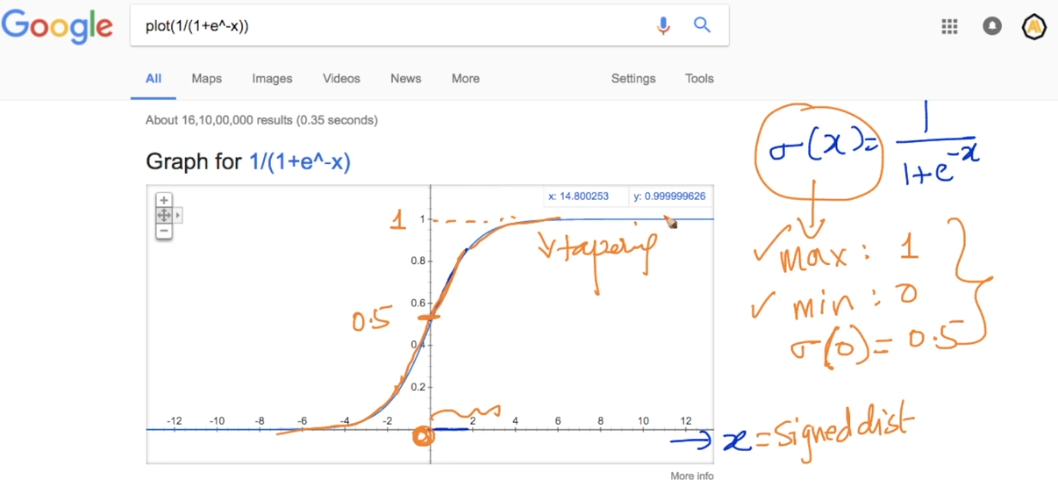
Practically pi 1 is best but we get wrong plane just because of 1 out l .

So we need any function which will handle this type of out l. fo

That function is called Sigmoid function .



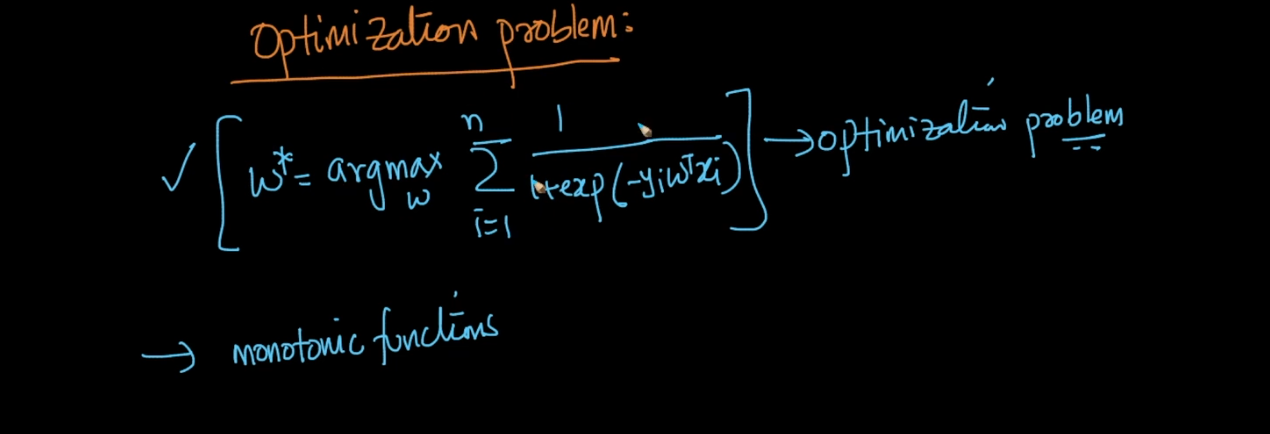
This function will squish out l. values from very large to small how ?



This function has great probabilistic interpret .

Lets say we are predicting y =1 for some x value

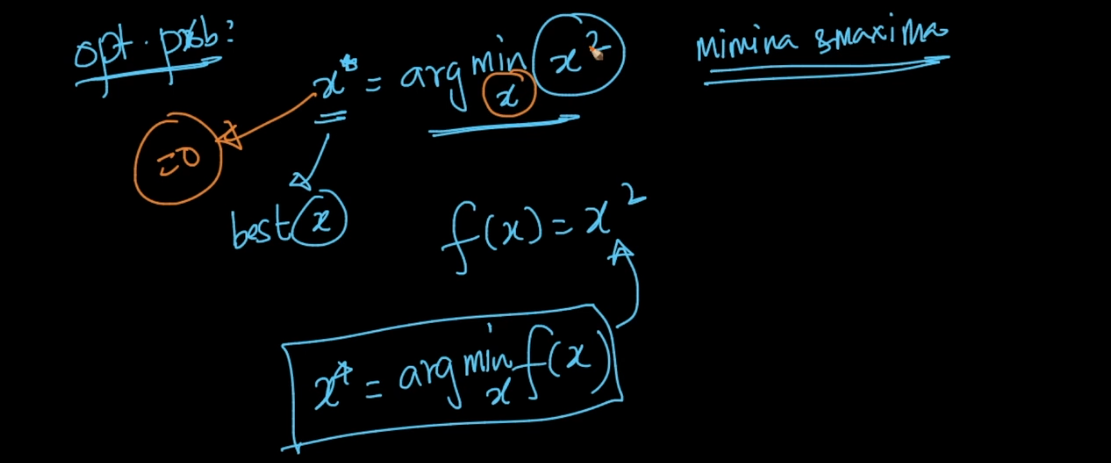
And lets say x is very large like out l . 12 in above example then this function return 99% prob value for belong to class 1 .

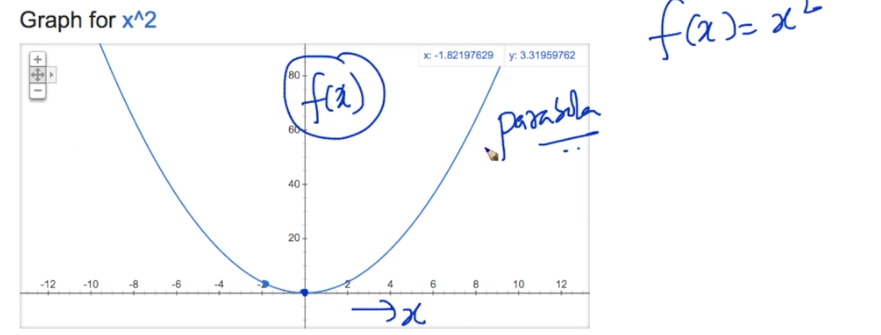


Log x is a monotonic function means as x inc log(x) also increase .

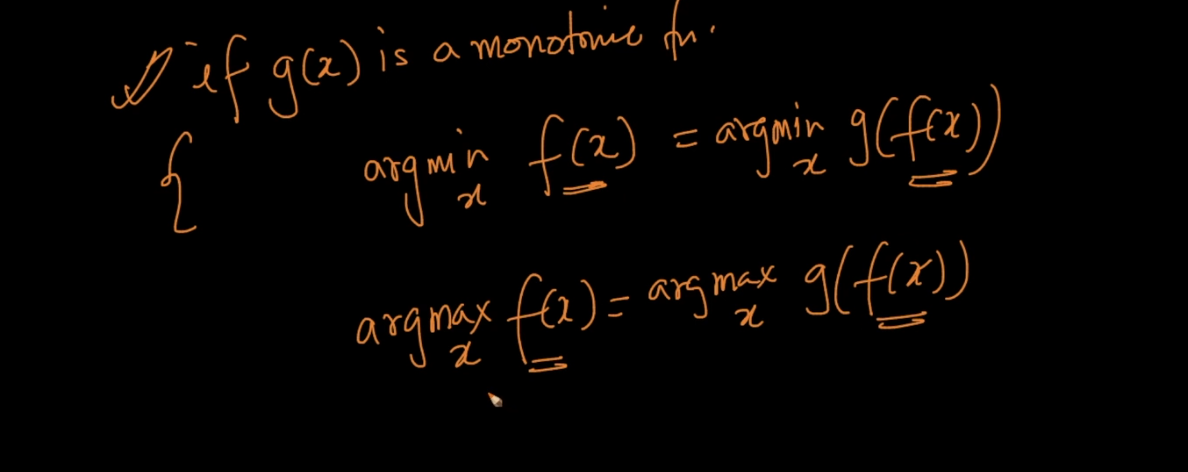


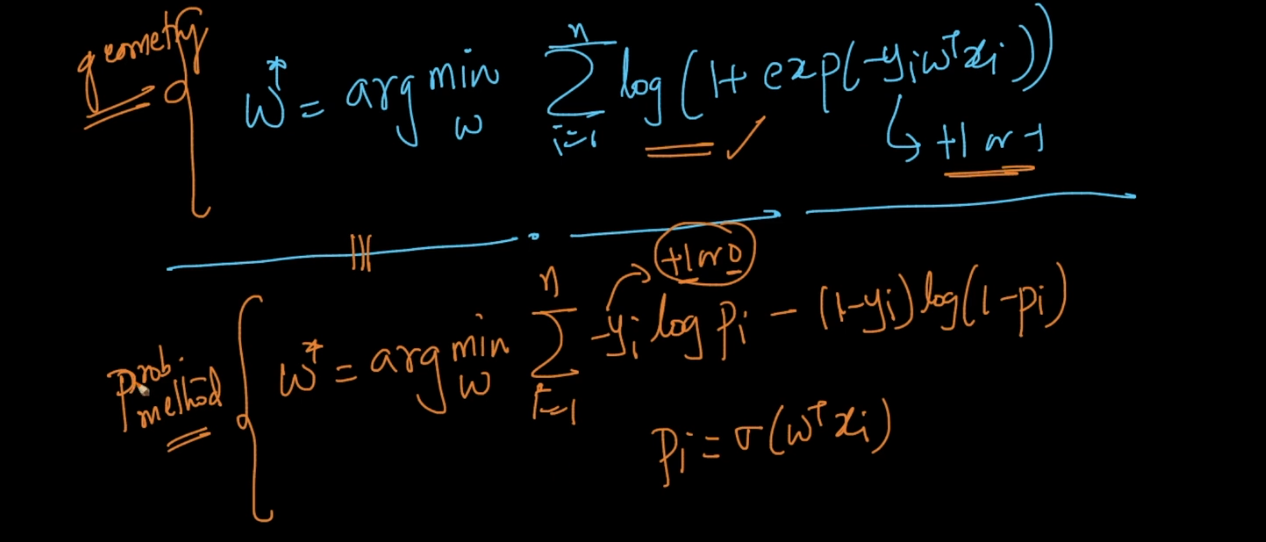
See below image that is opt problem it means find best x which minimize x2 .

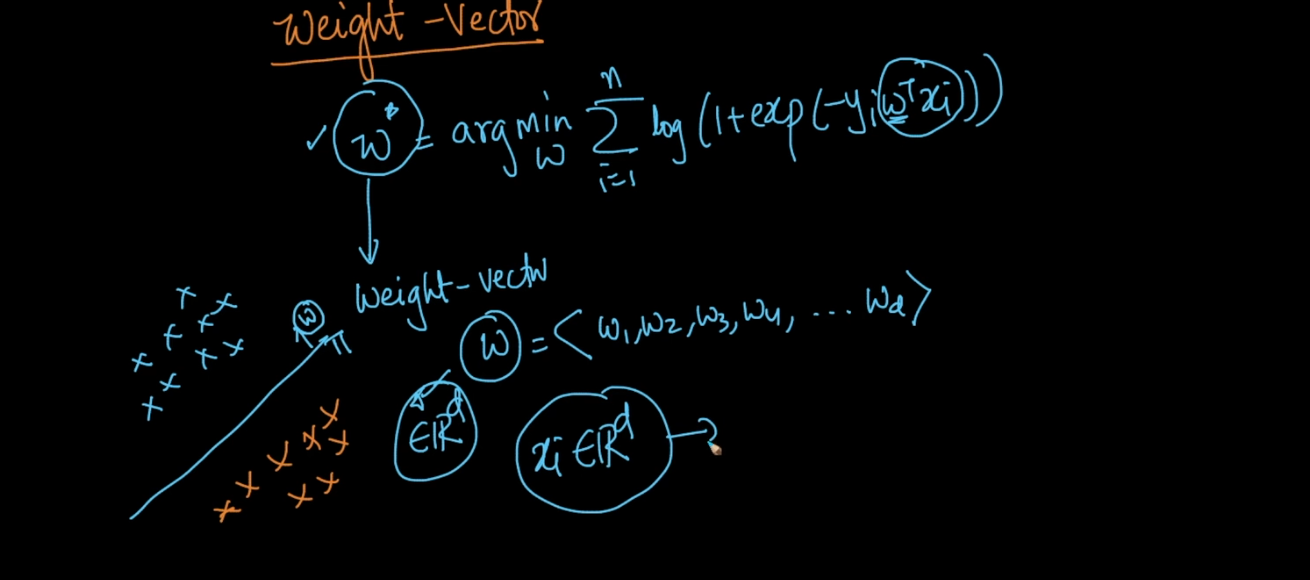


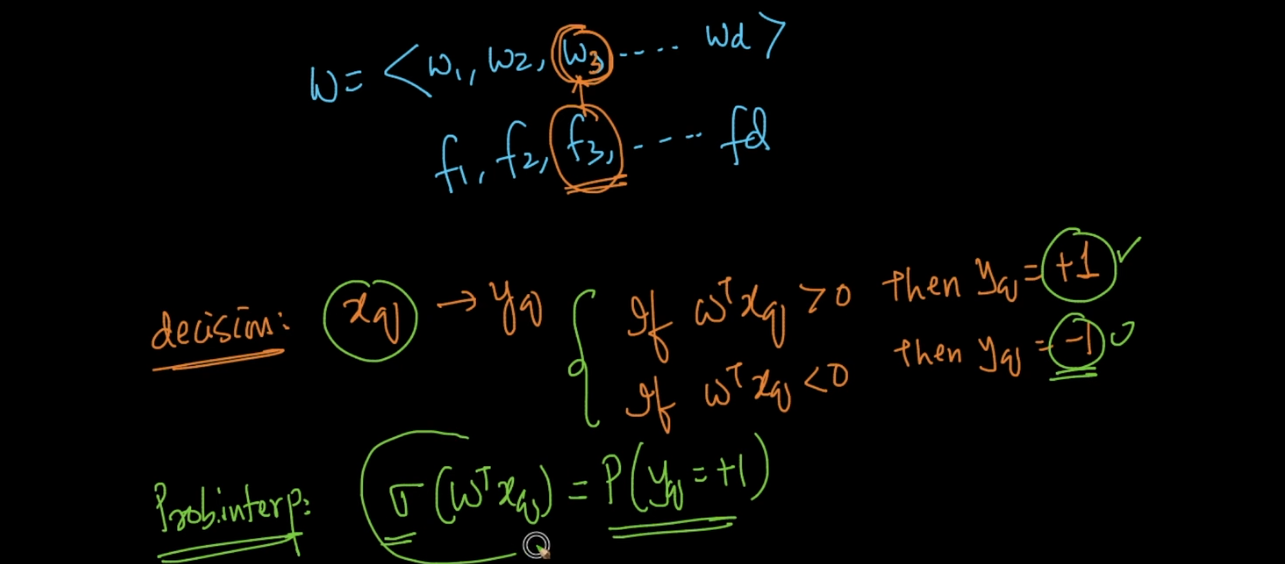


See x = 0 which minimize f(x) .

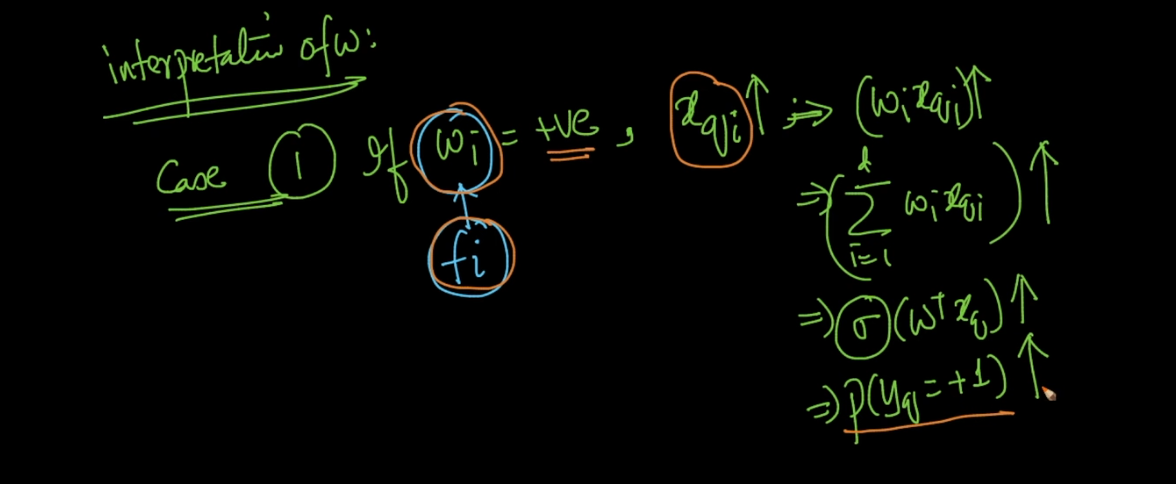


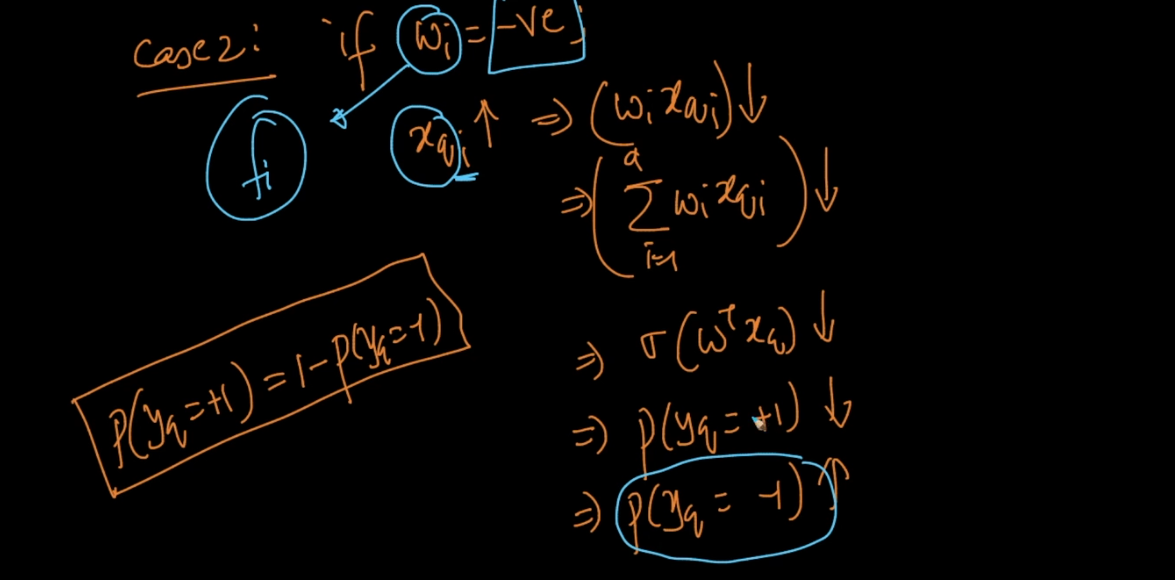


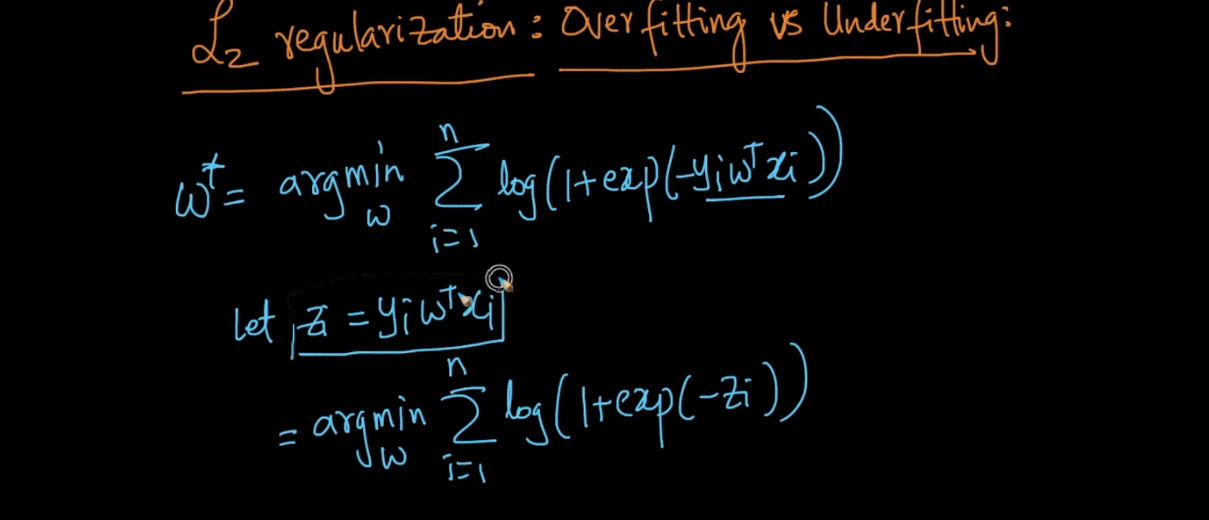




For each feature we have weight . we can find prob of new point will be + using sigmoid function .



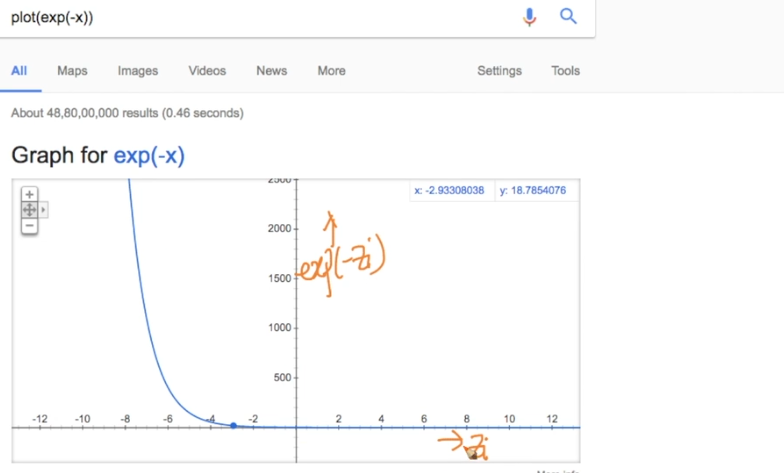




Lets take first exp part of equation .

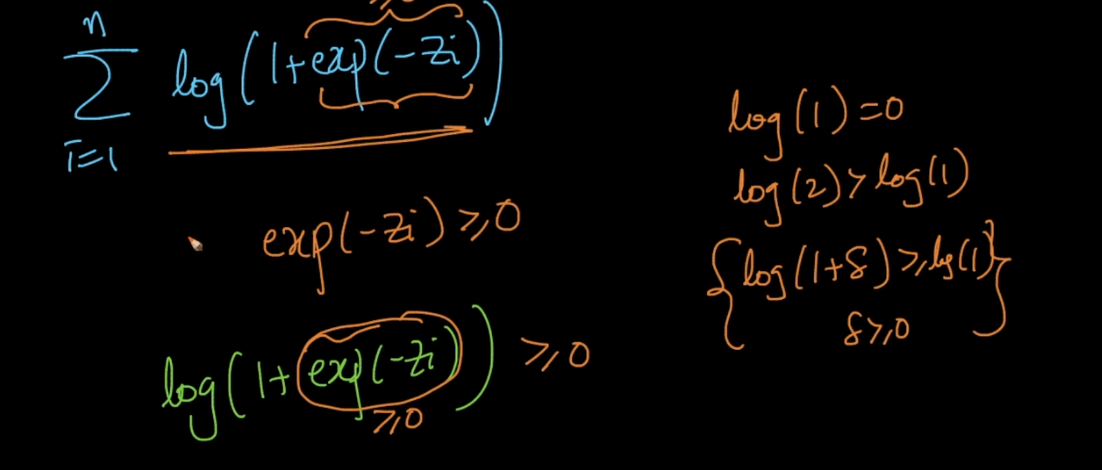
Exp(-Z\_I)

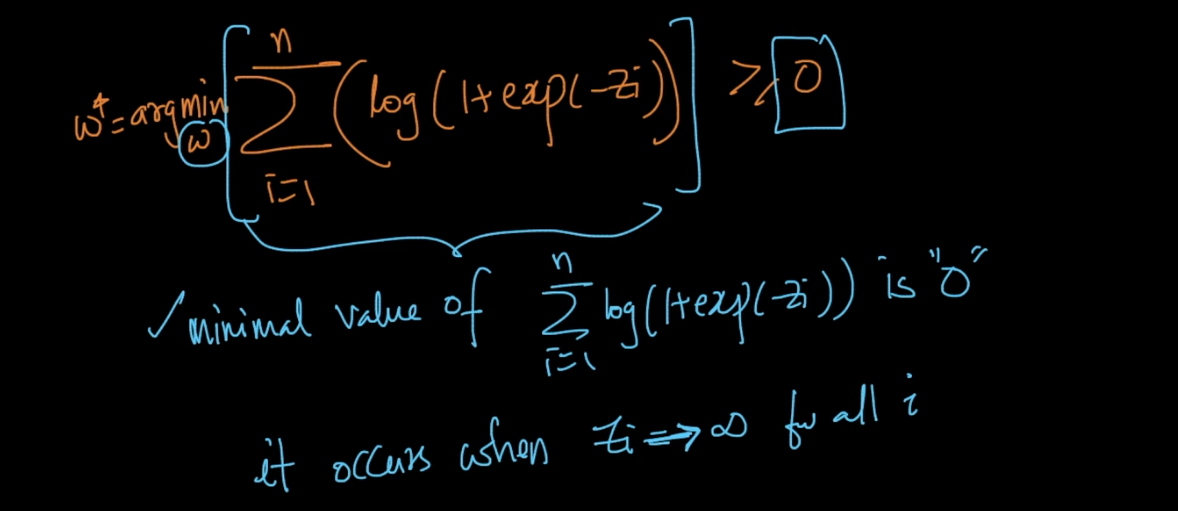
If we plot exp(-x) on Google we get this value will always >= 0 .



Now we know log is monotonic increasing function so log(2) > log(1) similarly ,

Log(1+e(-z)) is also always > =0 .



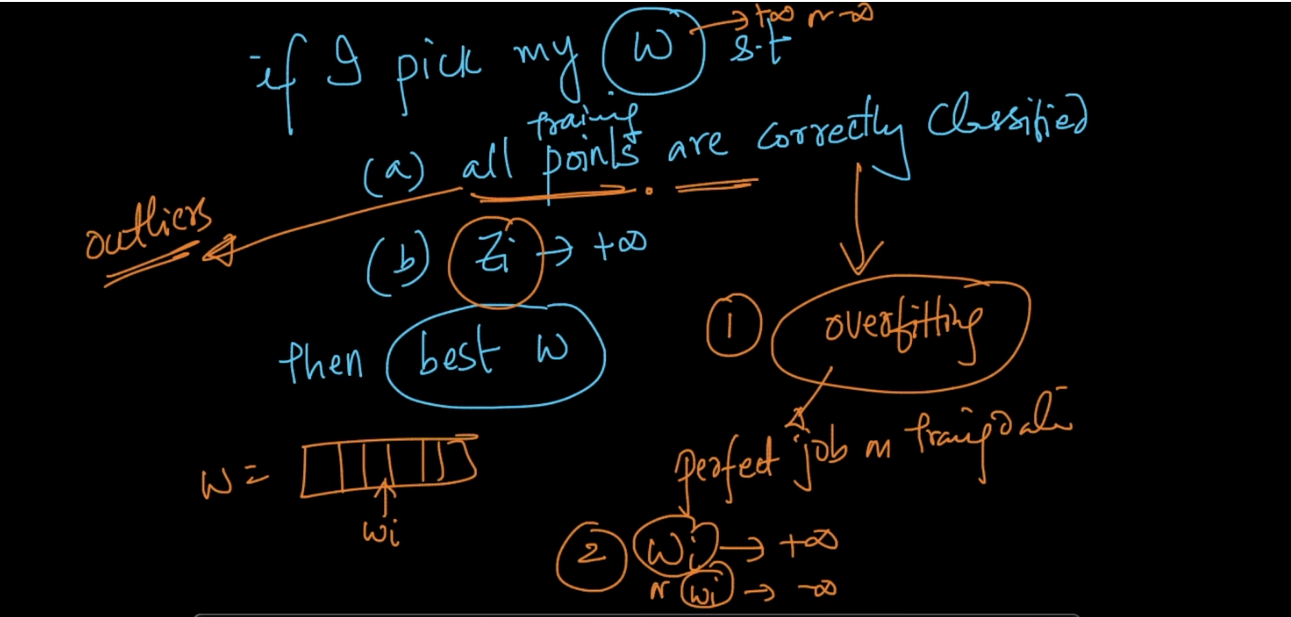


We want to find min value for w and min here is 0 .

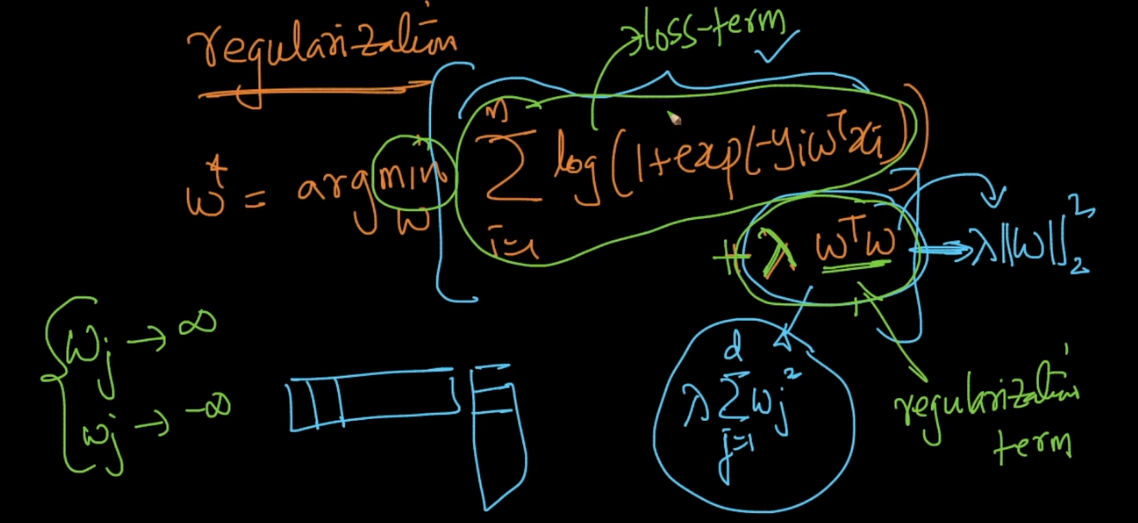
We achieve these when sum of all z\_I = inf .

So here x and y are fixed from train data we cant do anything with that only remaining is w so we modify w in such a way so we get sum of z = inf .

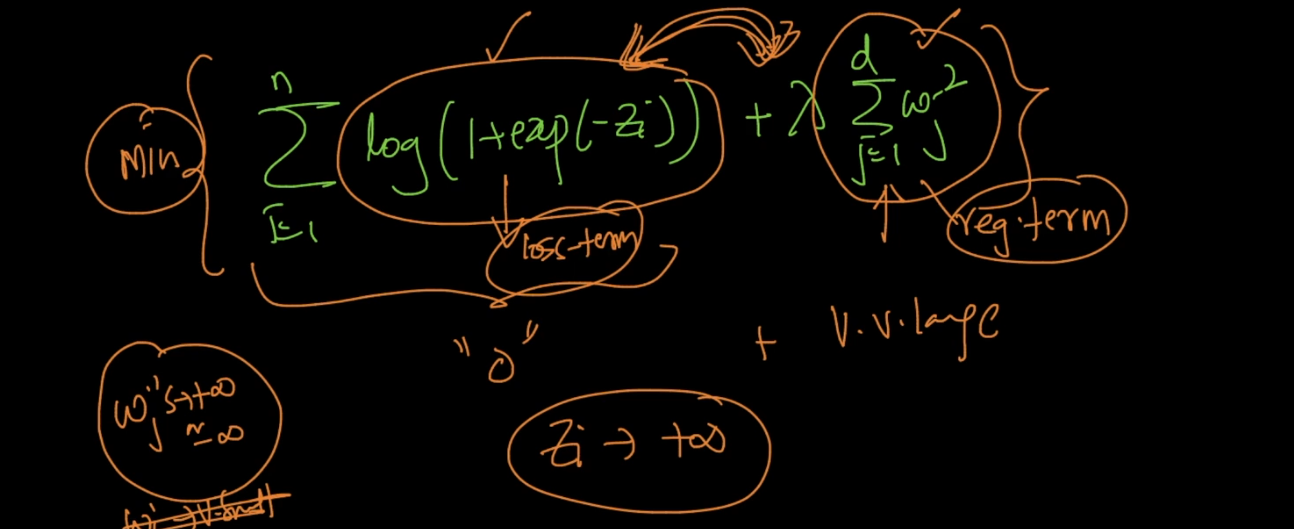
Lets assume we get perfect w but with that we get over fit issue also . because we are trying to fit perfectly out l. also so this will work good on train data but on test data it will fail this is overt fit issue .



For that purpose we use regularization which is done by adding :



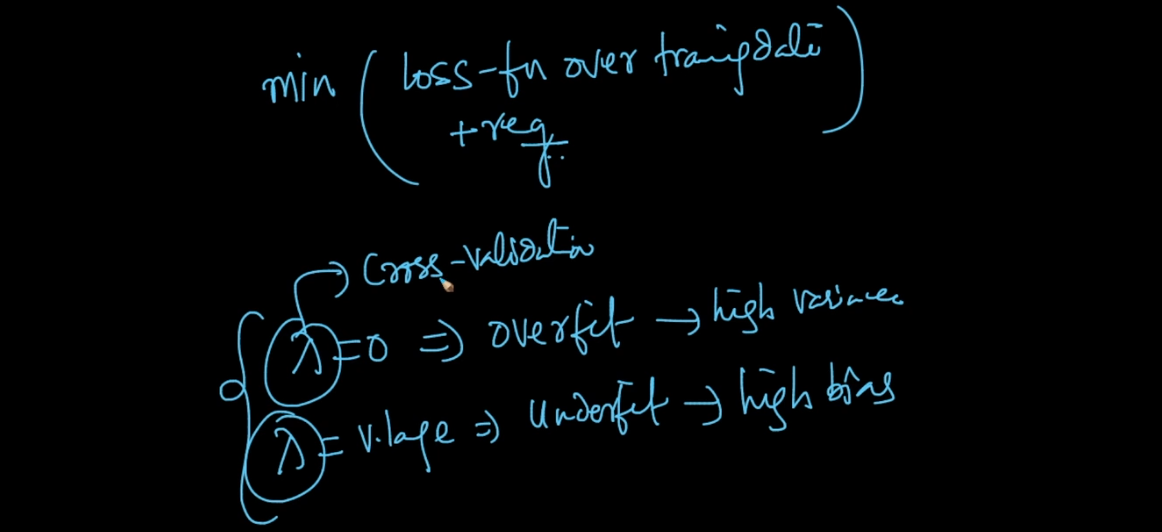
Because of that new term it will never allow to become + inf to w\_I . it will always oppose .



So if we use

Lamda = 0 then we get over fit problem

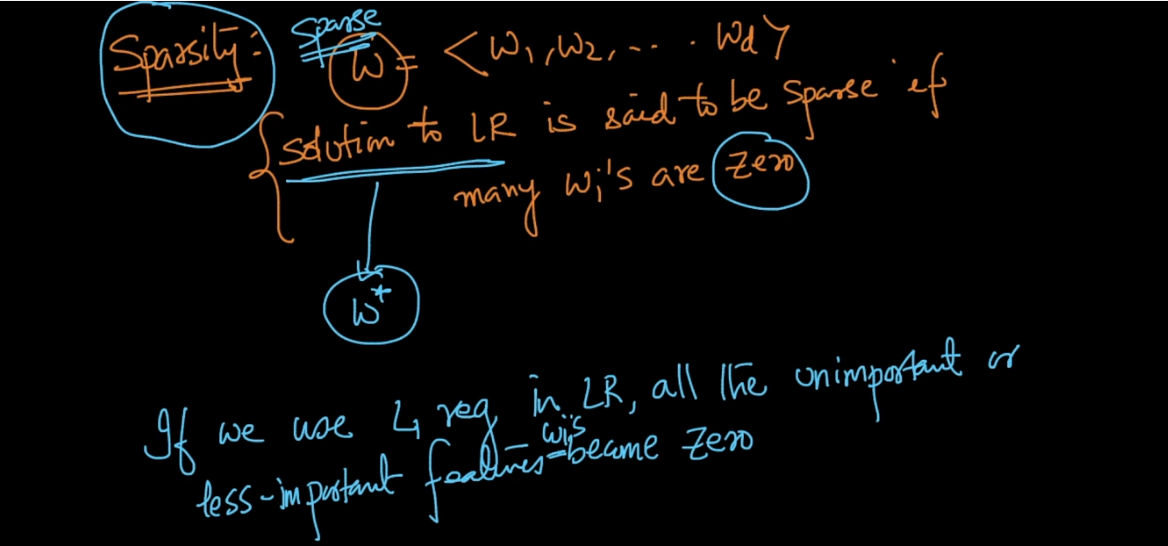
Lambda = very large then we will get our loss term very low and reg term very high so we will not use train data in that case and we get under fit problem ,.

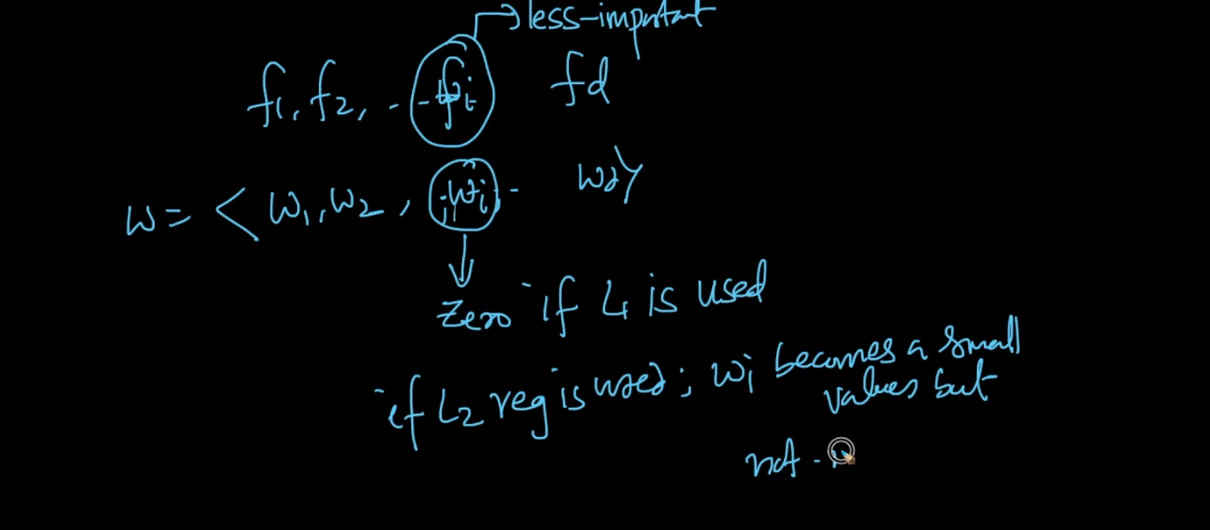


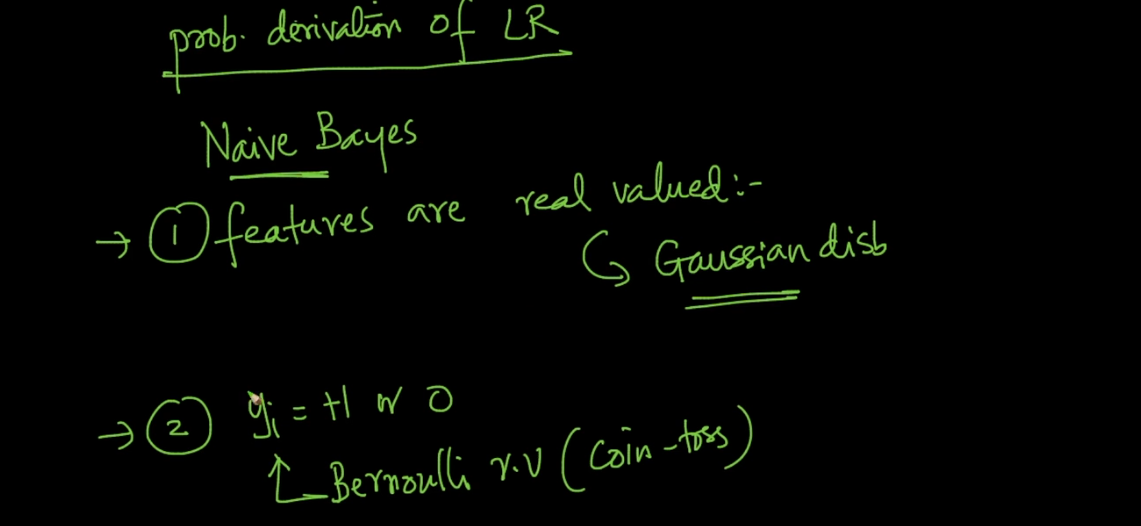
Another very popular alternative for L2 is L1 reg .

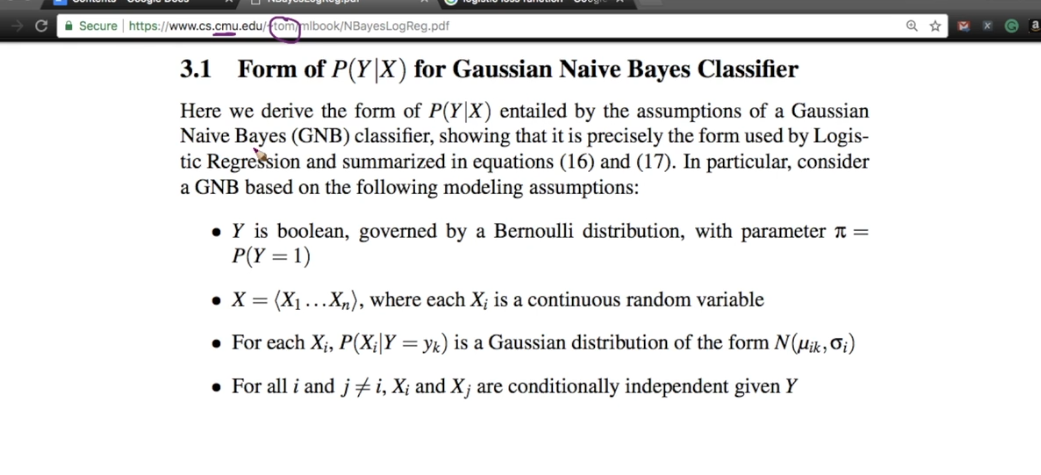
Because L1 have Sparse feature . We can say Regularization Sparse when it has maximum 0 values .

So if we use L1 then all unimportant features becomes zero .

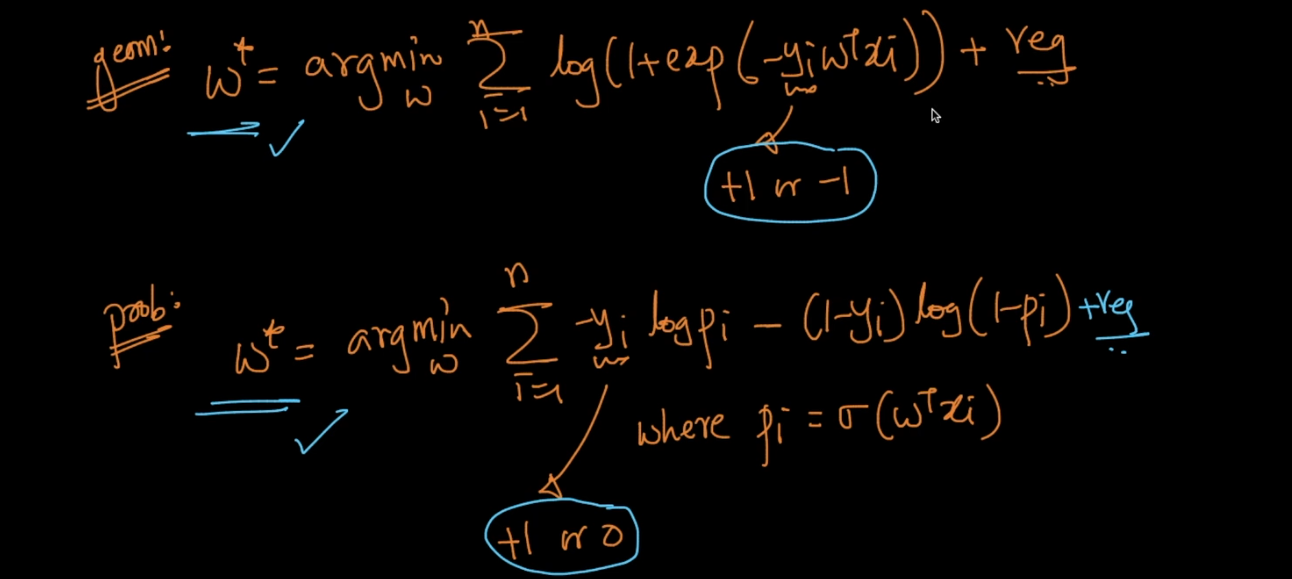








Below image both function work as same .



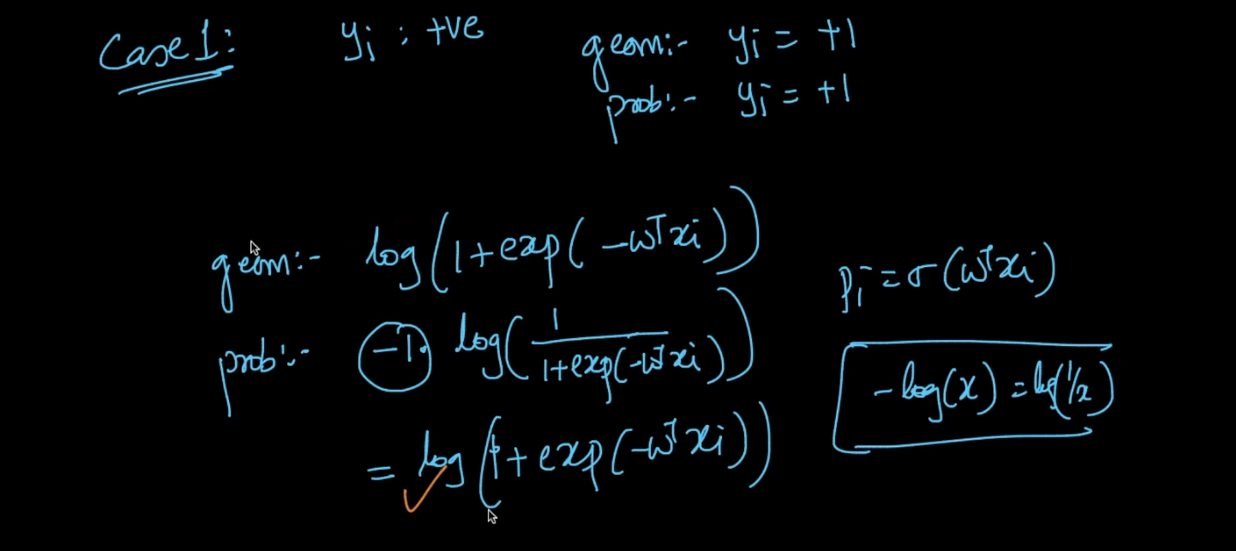
Proof :

Lets take case 1 :

Y = +

Means y\_I(Geo) = +1

Y\_I(prob) = +1



After putting all values we get same result for both .

