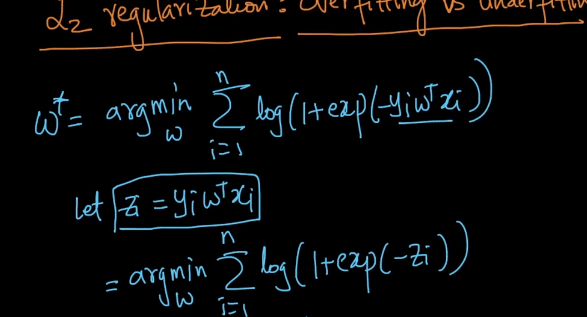
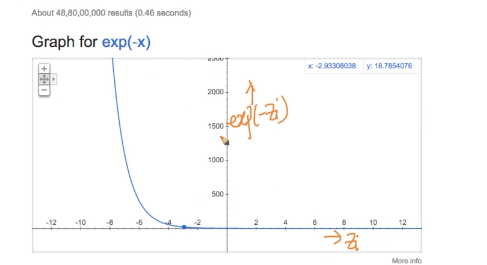
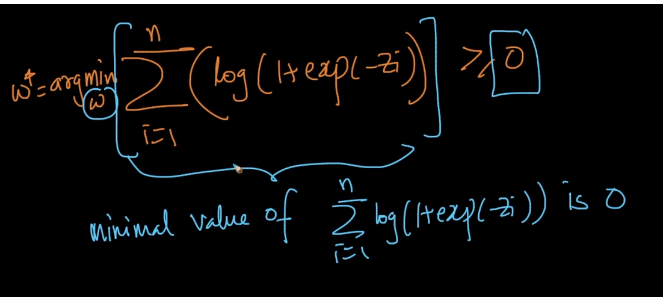
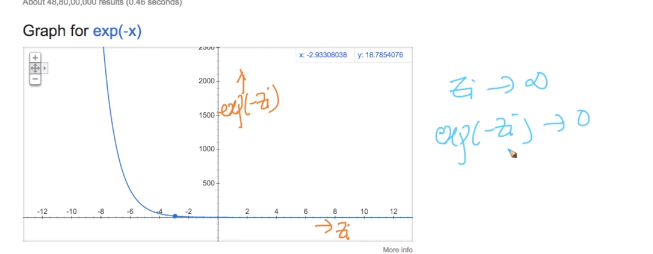
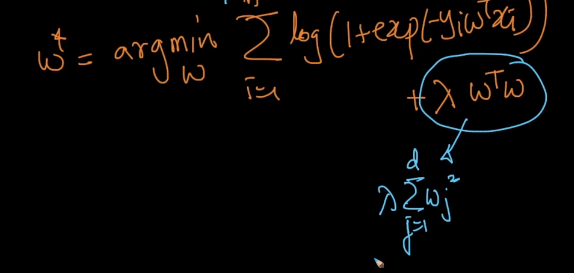
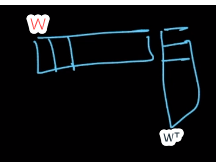
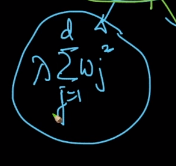
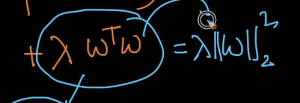
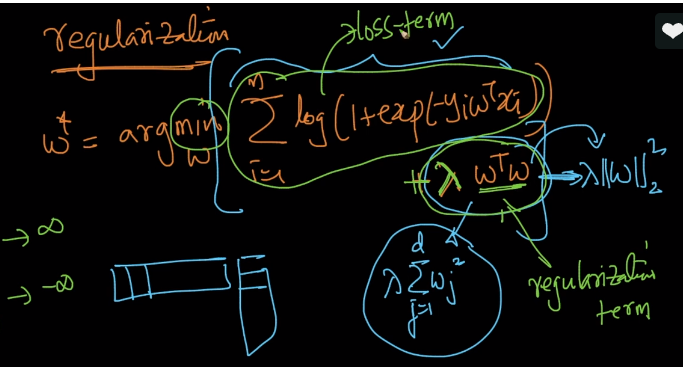
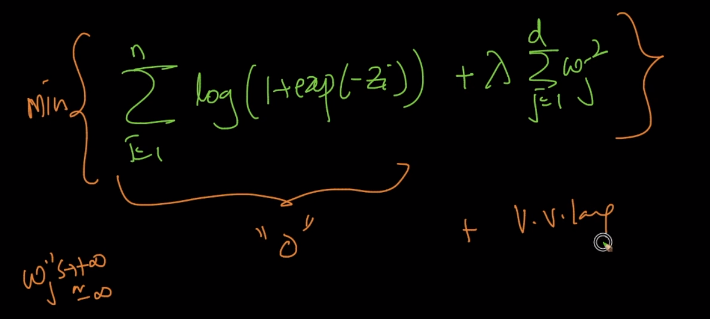
**L2 Regularization: Overfitting and Underfitting:**Below is the optimized equation, it’s always difficult to write the complete equation. So, we will write it as below   
Let’s plot for exp(-Zi),in below we can say that exp(-Z) is always positive for whatever value of Zi we take, so exp(-zi) >= 0  
  
Let’s look at whole equation, we know what log(x) is monotonical function as it’s always increases as x increases.  
So, in equation log(1+exp(-zi)), the highlighted part is always greater than or equal to 0.  
we can say that log(1+exp(-zi)) is always greater than or equal to 0 because   
on addition both log(1)=0 and we know exp(-zi) is positive number, we get some value.   
  
Let’s go back to what our original equation, we use it to find the minimum W value, so as per above, we can conclude that minimum value of W\* is 0  
  
  
**Next question, when does this 0 occur?**As Zi tends to infinite then Exp(-Zi) is also tends to 0 as per plot below meaning Zi tends to increases Exp(-Zi) is decreasing.  


In below, Xi and Yi are already given in training data set.   
  
Zi 🡪 +∞  
we need to modify W in such a way that each Zi will tend to infinite and we know what if Z🡪+∞ means Zi is positive which means our point is correctly classified. To summarize  
1) Zi = +ve ; Xi is correctly classified by W  
2) Zi 🡪∞; we have minimum value of W

If I Pick my W such that

1. All training points are correctly classified.
2. Zi 🡪∞ we have minimal value of W

Then we can say that We have best W  
  
But there is problem with this i.e., overfitting Meaning we are doing perfect job on training data but no correct results on test data  
1) As per Point a, what if some training points are in-correct or outliers.   
2) Here, Xi and Yi are fixed values and If Zi tends to infinite meaning W will tends to –∞ or +∞ based on signs of Xi and Yi meaning W’s are becoming very large, because of this we are overfitting.  
  
While constructing optimization problem, we missed one key concept that is W is normal which means WTW=1  
  
To avoid this overfitting problem, we use regularization technique   
  
**Regularization:**Equation:

**WTW – regualrization term in optimization.**  
WTW can be expanded as below   
 we can also right it as below, we call it as L2 norm of W2  
Point to be noted is, here we are finding minimum value on whole equation.  
   
So, the reason we adding the regularization term is, it won’t allow the sum to be infinite because we finding minimum value and optimization doesn’t allow it to go to infinite value any term to infinite  
For example,   
**Case#1** if the Wj’s tends to infinite (+ve or -ve) then loss term will be 0 and will be very large  
  
  
Regularization term and loss term fighting to avoid Zi 🡪 Infinite.   
Lambda is hyperparameter   
When Lambda = 0 then overfit ting will occur  
Lambda = large then influence of loss term will be decreased which means underfitting.  
Also It means, we not using training data to calculate Lambda value. Because in regularization term we don’t have any training data  
**Summarize: we use cross validation techniques to find out lambda**