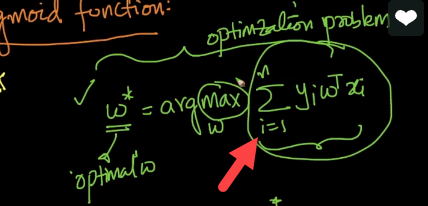
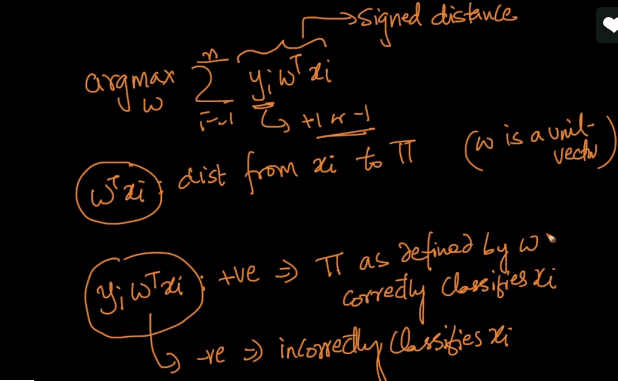
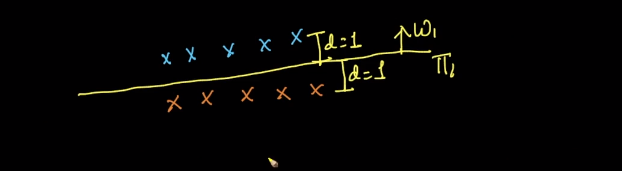
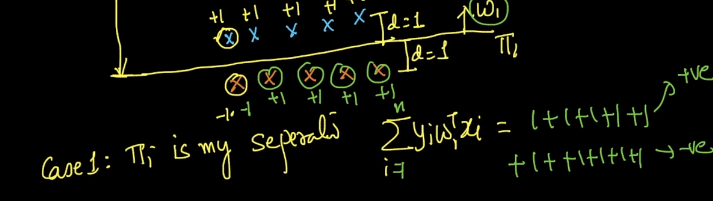
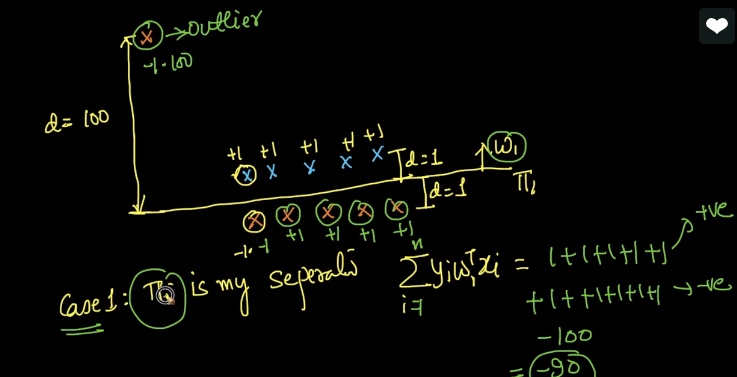
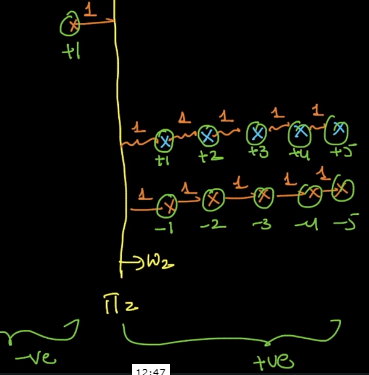
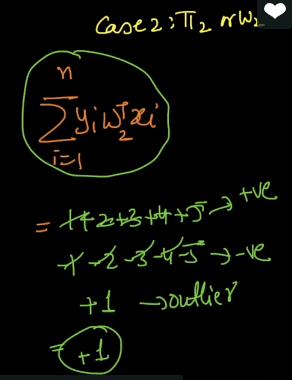
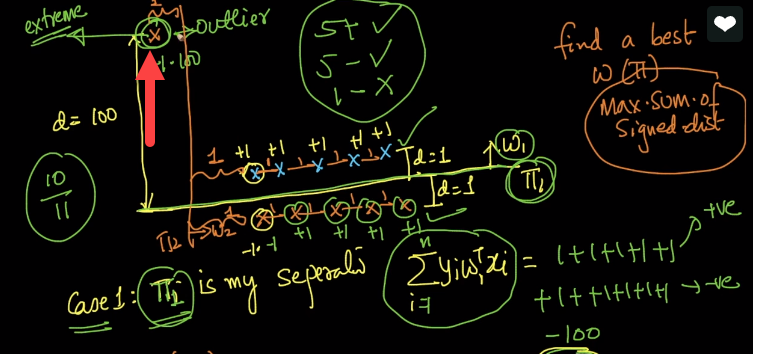
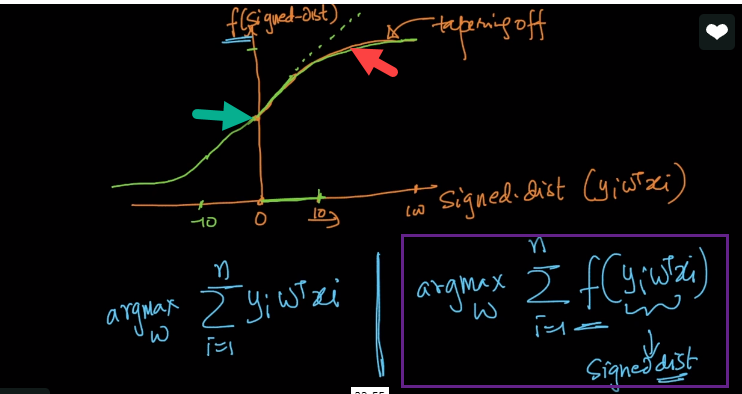
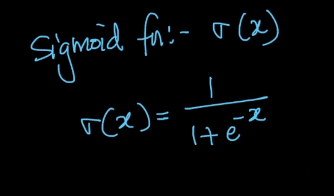
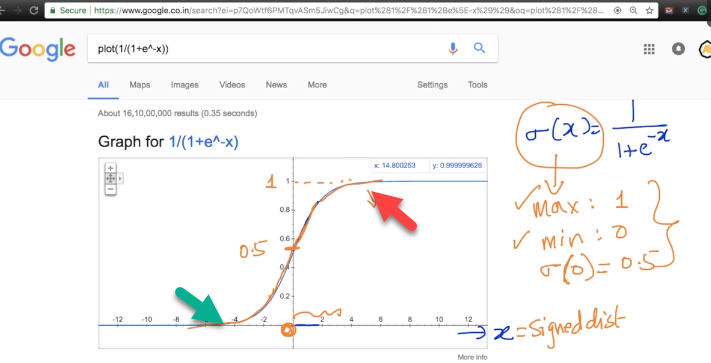
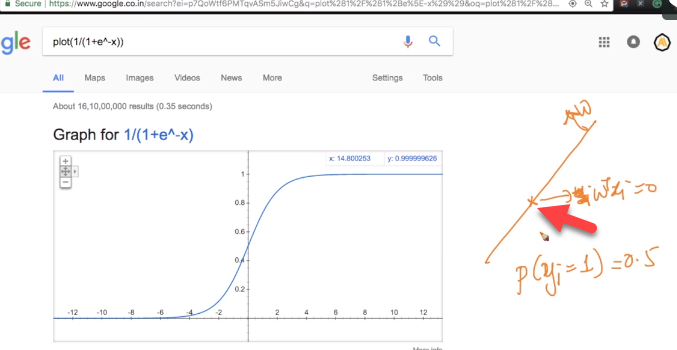
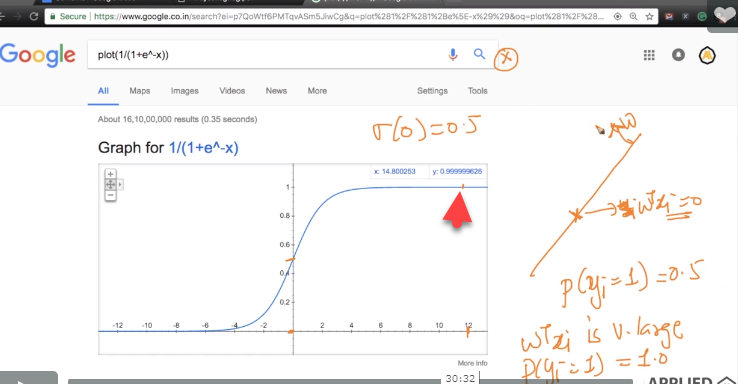
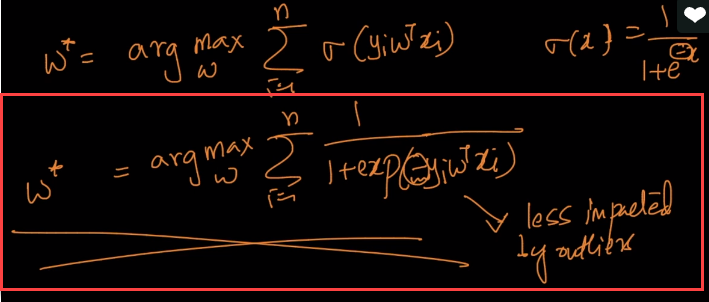
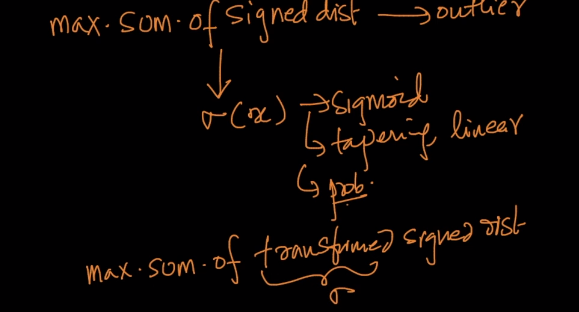
**Sigmoid function: Squashing**we already know that, it’s always best to take W(optimal value) value which gives maximum sum for ∑i=1 to n YiWTxi .The final task is to maximize correctly classified and find correct plane(π) which maximizes the signed distance.  
The highlighted part we call it as **Signed** **distance** because we are assuming WTxi : distance from xi to π while assuming W is a unit vector and Yi can take +1 or -1 as value  
If Signed distance is positive, the plane defined by “W” is correctly classified.  
If Signed distance is negative, the plane defined by “W” is incorrectly classified.  
  
**Let’s take an example where the above function doesn’t work:**  
we have a plane π1 with group of positive and negative points on either side. W1 is normal to plane and each point is having distance of unit 1 meaning d= 1  
  
For above, let’s calculate the Signed distance for both positive and negative points.  
**Case 1 :**As shown below, signed distance for positive and negative points is Positive meaning they’re classified correctly**.****For positive points,**Yi = 1 because,the given point is on positive plane.WTXi= 1, we already assuming it as unit vector  
**For negative points,**Yi = - 1 because,the given point is negative.WTXi **=** - 1,we already assuming it as unit vector which is on negative plane.  
Multiplying both here will get a positive number  
Equation works here.  
Let’s assume there is one negative point on positive side of plane with d = 100. Signed equation will be 100\*-1 which will be -100. As shown in below pic, sum will be (1+1+1+1+1+1+1+1+1-100 = -90).  
  
To summarize case#1, we draw a plane which classifies 5 +ve points ,5 -ve points correctly and 1 -ve points incorrect by which we get result as -90.   
**Case 2:**Now let’s draw a plane π2 in different way as shown in below pic**,** **For positive points,**Yi = 1 because,the given point is on positive plane.WTXi= 1, we already assuming it as unit vector  
**For negative points,**Yi = - 1 because,the given point is on negative.WTXi **=** 1,we already assuming it as unit vector and they’re on positive side of plane. Final sum will be +1 as shown in pic below.

**Summary of Case#1 and Case#2**As said earlier, the Objective here is to find W which maximizes signed distance. By looking on above, for Plane 1, we have -90 and Plane 2 we have +1. So, by default we choose +1 as it’s giving me maximized signed distance. But if go by correct classification of positive and negative points plane 1 is good than plane 2 as in plane 1, we have more correctly classified points, there by we say π2 is terrible classifier. This is happening because we have extreme point in the data set highlighted below.  
By this we can say that Signed distance are not outlier prone meaning if we have an outlier then singed distance won’t work. So how we must change this formulation such that outliers don’t impact our result, we do it by using Squashing technique. x  
**Squashing:**It’s a technique, where we can reduce the outliers in singed distance, and it’s done by tampe off the outliers once they reach above the limit defined. Mathematically, we need to write a function which will taper the outliers.  
As shown in below pic, we’re going to tamper off the after limit(marked in arrow).  
Mathematically, we need define a **function**, which will increase and after certain limit it will taper off the outliers.   
  
One such function if called **sigmoid function** and written as below.  
  
**Plot**:  
As shown in plot below, x(signed distance) increases till some extent and later will taper of, highlighted in red and green arrows.  
by looking on to plot we can say that max value of **sigmoid function is +1 and minimum is -1** because after that x is taper off. And when x =0, the value is 0.5  


There are many functions with such properties but why did we choose sigmoid function?   
it’s because if multiple reasons, one such is Probabilistic interpretation. Explained below  
if the signed distance is 0, then what is the value of Yi?   
**Point#1** It is 0.5 meaning it can be either positive or negative. Shown same in pic below   
  
**Point#2** If point is on same side of normal having very large signed distance, we get value if Yi =0.9999 which is ~1  
  
By considering both points above, we can say that, we are getting probabilistic value of given point.  
**Flow and new Formulae :**Highlighted(red box) is new formulae which has less impact to outliers rather than simple summation  


We call this function as squashing because, we can have distance in between positive infinite to Negative infinite but using this function we are squashing in such a way that we will have values in between (-1,1) by using sigma function.  
we use sigma function because it is easy to differentiate.