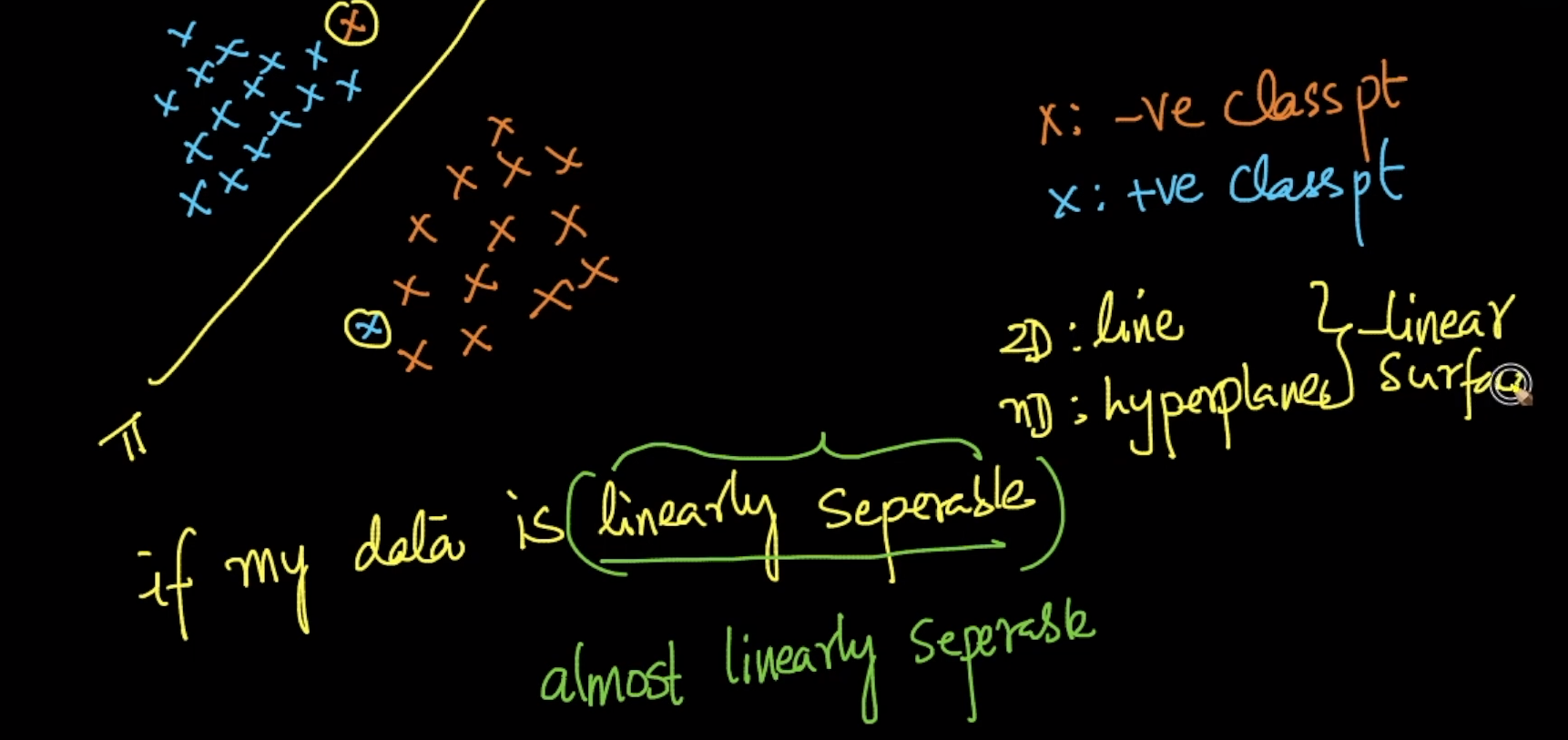
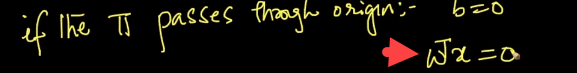
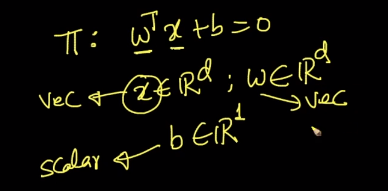
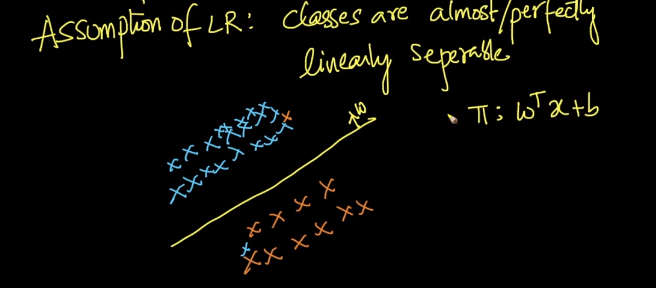
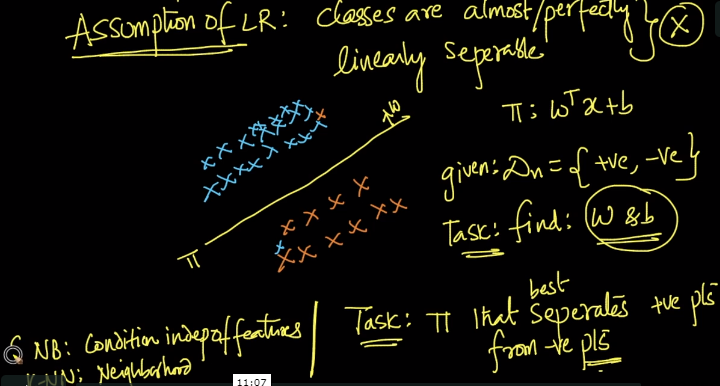
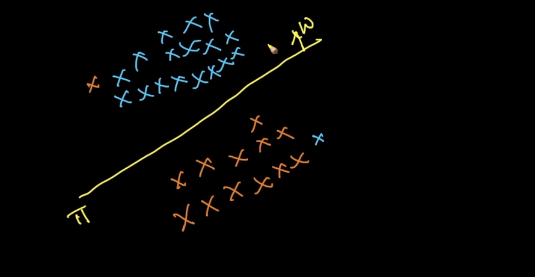
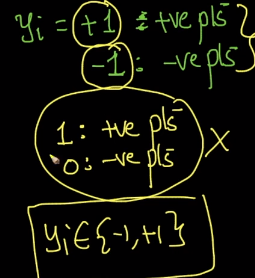
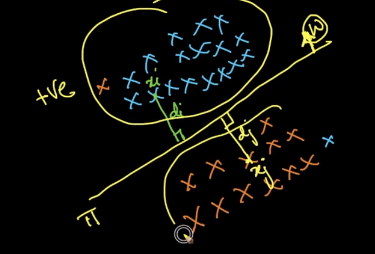
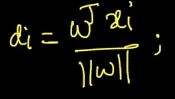
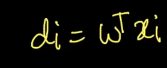
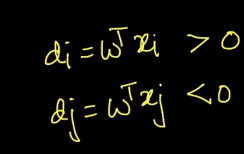
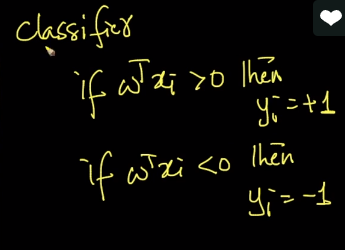
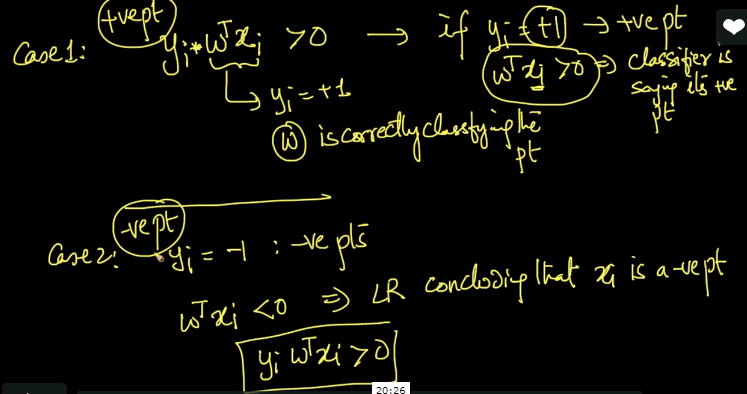
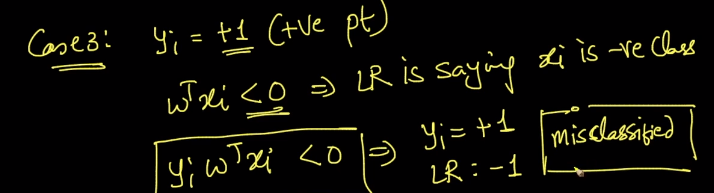
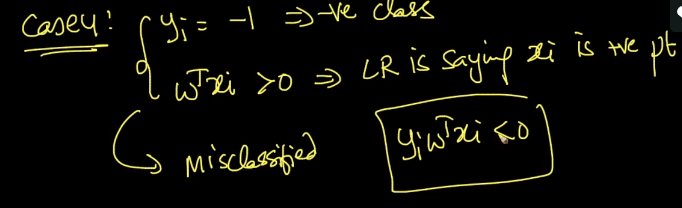
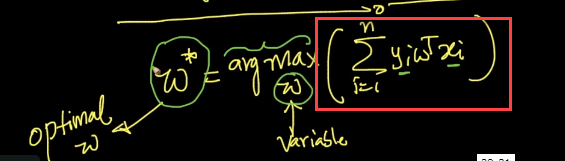
**Geometric intuition of Logistic Regression:**

It’s a classification technique.  
we can be intuited by different techniques like geometry, probabilistic, Loss function.   
Geometry is intuited as follows.  
Linearly separable : If we can classify positive and negative class by a linear surface  
Almost linearly separable : Almost all the points are separated, only few points are not separated.  
2D- Known as line  
ND- Known as hyperplane  
  
Line: Linear surfaces  
Circle, parabola, elipses : Quadratic surfaces  
For any given line, there will be normal which is like perpendicular line.   
Plane can be represented as **π: (w,b)** where w is the normal of the plane and b is the intercept of the line.  
Equation of line in plane **y=mx+c** in 2-D. and equation of line or plane with high dimension will be as shown in pic  
  
If the plane(π) passes through origin, then b=0 and equation of plane will become   


In general equation of plane, we can use it only if following conditions are true  
**x** is a data point and in D-dimensional space , **W** is also a vector in d-dimensional space and **b** is scaler belongs to 1-d dimensional space   


**Assumption : separation in LR makes is Classes are almost/ Perfectly separable. Example below pic with equation of plane when w is the normal**

we already have +ve and -ve points and we need to find **W** and **b.**Meaning, Task in logistic regression is find the W and b which corresponds to the plane such that this plane best separates positive points from negative points by assuming both the classes are separable.  
In any algorithms will something by default **for** **example**:  
In Naive Bayes, we assume the features are conditional independent.  
In K-NN, we assume neighborhood of a point is same as a point.  
  
**Mathematical explanation of LR:**  
Let’s assume we have linearly separated plane, we need to fund out **π** and **W**  
Earlier we used to assume 1: +ve points and 0 :- ve points but in LR, we assume as follows,  
**Yi = +1 : +ve points** **-1 : -ve points**Yi belongs to {-1,+1} where Yi is class label  
Let’s assume a point **Xi, Xj** in a plane and distance of that point from plane is **di,dj** where assuming **W** is normal to plane **π** passing through origin, so B =0   
Formulae to calculate the di isand let’s assume **||W||** is unit vector meaning **||W||** = 1, we can discard the denominator.  
the equation changes for point **Xi** **Xi**is on the same side of **W** and **Xj** is on other side of the plane. Based on this we can say thatwe can define the classifier as well   
  
we need to remember that, Decision surface in logistic regression is Linear plane  
  
Based on above assumptions, we will define 4 cases. Task here is we are deciding the plane W to get the correct classification of points  
**Case#1:** All **Positive Points side**  
Here Xi and Yi are points from the data set D Where Yi = +1 here  
if Yi = +1 point and WtXi > 0 classifier saying it’s +ve point.  
on multiplying both, we getting result positive point i.e., **Yi\*WTxi > 0,** then my plane “W” correctly classifying the point.  
because here plane **“W”** saying that, given point is positive and in data set D it’s already +1 which is positive.

**Case#2:** All **Negative Points Side**  
Here Xi and Yi are points from the data set D Where Yi = -1 here  
if Yi = -1 point and **WtXi** < 0 classifier saying it’s -ve point.  
on multiplying both, we getting result positive point i.e., **Yi\*WTxi > 0,** then my plane “W” correctly classifying the point.  
because here plane **“W”** saying that, given point is positive and in data set D it’s already +1 which is positive.  
 **In case #1(Positive) and case#2(Negative) Points  
YiWTxi > 0 🡪 Logistic model is correctly classifying the Point xiCase#3:**Now, Let’s say Class label Yi = 1 (+ve pt) and Classifier WTxi < 0 LR saying xi is -ve class -1  
when we multiply positive class with negative class, final equation will be Yi\*WTxi < 0 meaning LR concluding the point as -ve Point which is wrong or misclassified point.  
  
**Case#4:**Now, Let’s say Class label Yi = -1 (-ve pt) and Classifier WTxi > 0 LR saying xi is +ve class meaning 1  
when we multiply positive class with negative class, final equation will be Yi\*WTxi < 0 meaning LR concluding the point as +ve Point which is wrong or misclassified point.  
  
  
At the end, we want our classifier to be very good meaning minimum number of misclassification or correctly classified points. In other words, we want Yi\*WTxi > 0  
So we can use below formulae to define the maximum W which can be written as W\*  
  
argmax = whichever W variable gives maximizes the sum(Highlighted), call that variable as W\* and that W will be normal to plane where Yi and Xi as data points of dataset which are constant  
The Highlighted value will be positive if we correctly classified and negative if they’re incorrectly classified.