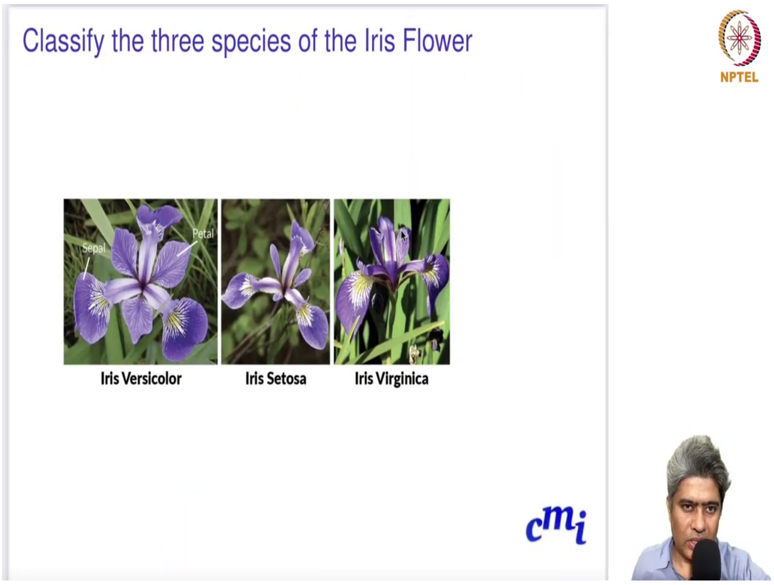
Predictive Analytics - Regression and Classification
Prof. Sourish Das
Department of Mathematics
Chennai Mathematical Institute

Lecture - 42
Multi-Class Classification with Discriminant Analysis

Welcome back to the Part C of lecture 12. In this video, we are going to start K class classification.

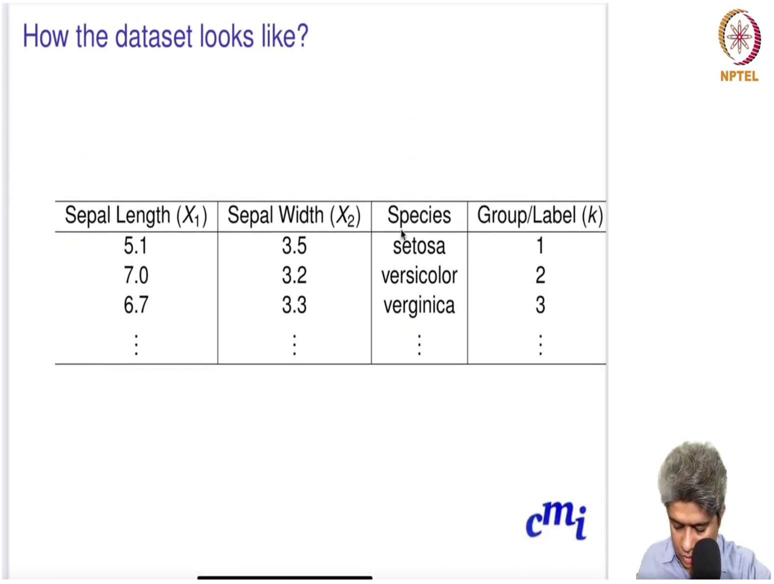
(Refer Slide Time: 00:25)



To understand the K class classification, we will; we are going to consider iris flower dataset. Its English flower iris and it has three subspecies. One is iris flower, iris versicolor, second is setosa and the third is virginica. Now, obviously, its this is the sepal and this is the petal.

This is the sepal and this is the petal now, of the flower. So, what they have done, they have; in this; in this dataset they have taken the petal width and petal length. And then similarly sepal width and sepal length, so these are the value that are being collected for different flowers.

(Refer Slide Time: 01:38)

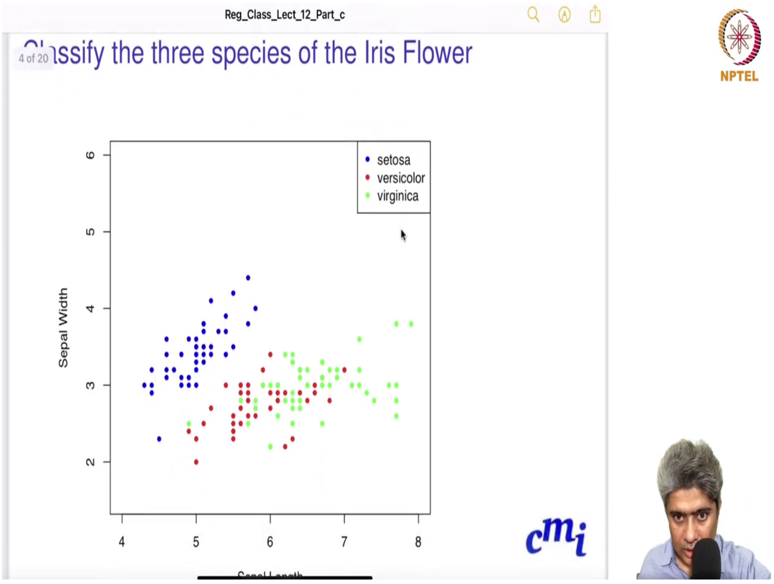


Sepal Length (X_1)	Sepal Width (X_2)	Species	Group/Label (k)
5.1	3.5	setosa	1
7.0	3.2	versicolor	2
6.7	3.3	verginica	3
⋮	⋮	⋮	⋮

And based on these flowers, we have to say whether its like you know; its just how the dataset looks like the for a particular flower, say sepal length is 5.1, sepal width is 3.5 and based on the its suppose, setosa and then we call it group 1, it belongs to group 1, we label it as a group 1.

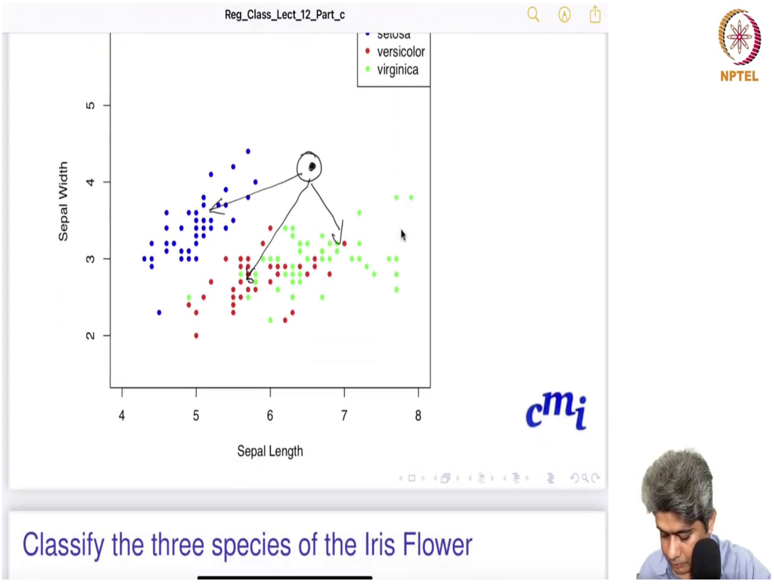
Similarly, there is another which is a sepal length of 7, sepal width of 7, 3.2, you have petal length, you have a petal width, you have a petal length, petal width and that belongs to versicolor and we call it where it belongs to group 2. So, that is how its been collected.

(Refer Slide Time: 02:28)



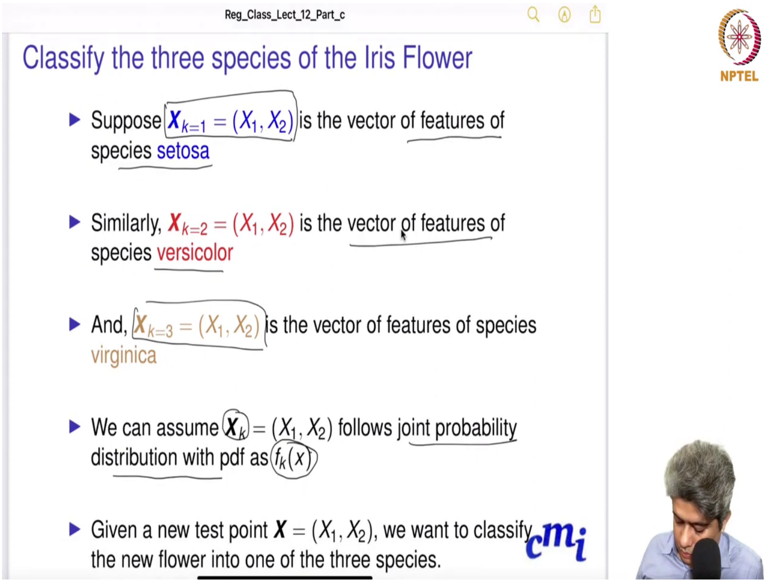
Now, that is how the data looks like. So, now, the our job is to classify these data, I mean, create a classification technique and for a new data point, say this is a new data point, ok.

(Refer Slide Time: 02:49)



I am I do not know, suppose this is a new data point, would you like to classify it as a setosa or is it going to be versicolor or is it going to be a virginica. So, for this new point, test point and a flower which got these values, sepal length and sepal width, what kind of classification, which class you would put it.

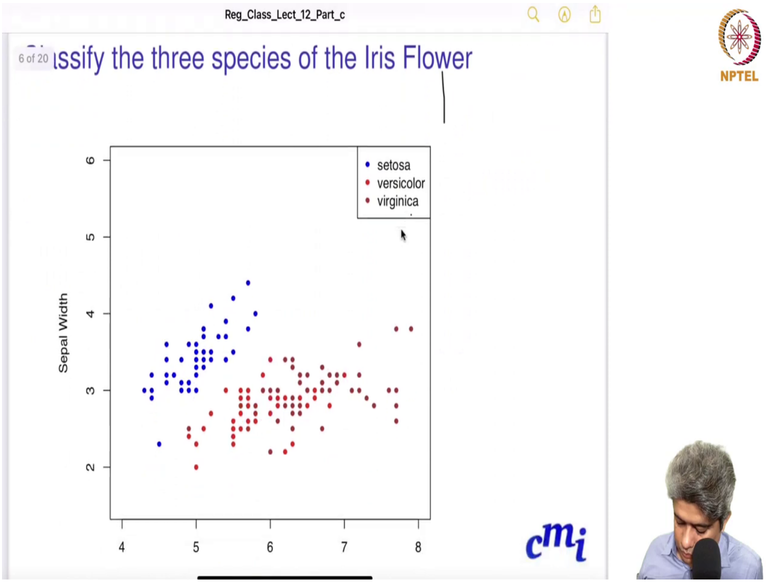
(Refer Slide Time: 03:21)



So, suppose what we can do? Now, given the that; you know given the features of the species setosa, I can create a sub set of the data. Similarly, given the feature of the subspecies versicolor, I can just create a sub set of species of the data. And similarly, I can create a sub set of the data for where all the features, all the species are virginica.

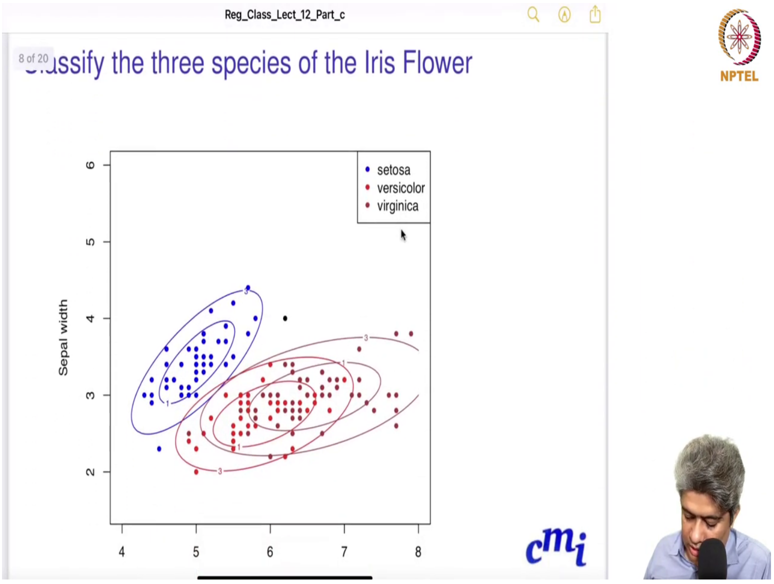
So, we can assume that X k follows joint probability distribution with some pdf probability density function f k x ok. Now, given a test point, we want to classify, then new flower into one of the three species ok.

(Refer Slide Time: 04:33)



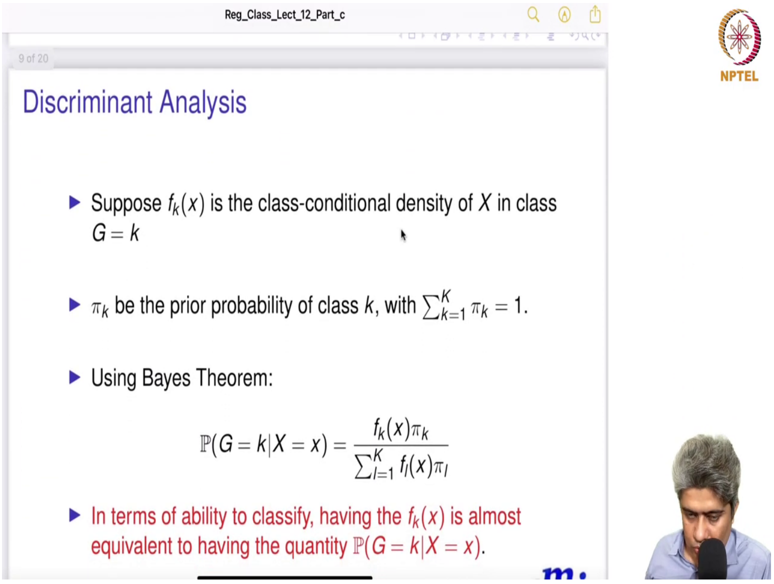
So, given a test point, which species it belongs to.

(Refer Slide Time: 04:40)



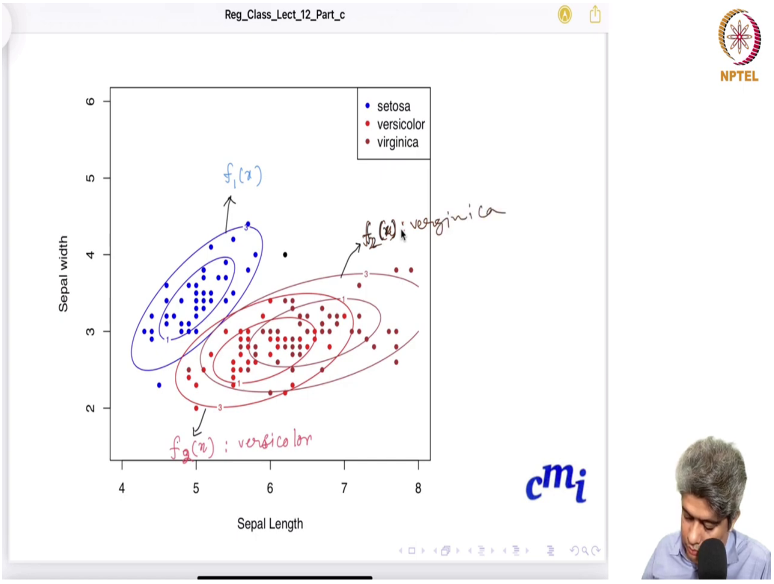
So, now, you can imagine, given the species I can try to think of these points, blue points, belongs to have their own probability distributions, which is setosa. The brown points, all the brown points which are virginica have their own probability distribution and the red points, whichever versicolor, it has its own distribution.

(Refer Slide Time: 05:17)



And so, I can think of these distribution, classify these distributions and based on these distributions, I can try to make a analysis we call it discriminant analysis. Now, suppose f k x is the class conditional density of x in class G equal to k.

(Refer Slide Time: 05:44)

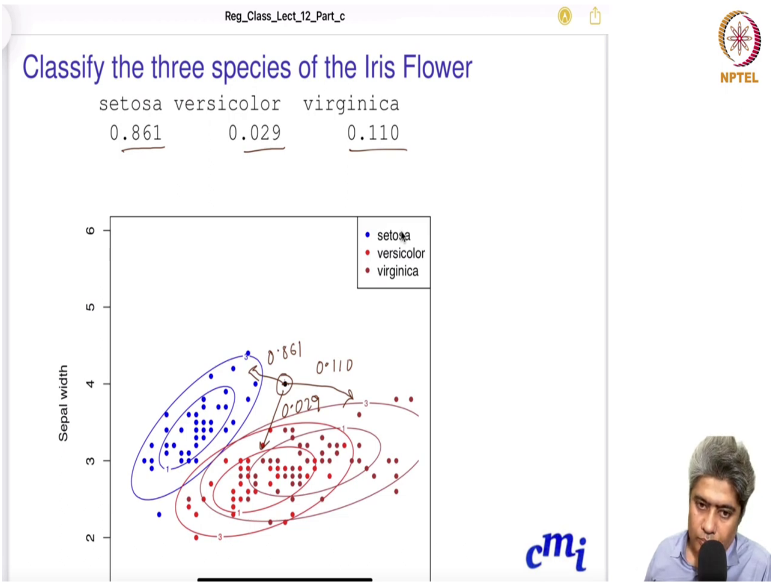


So, that means, if you go here and see that this a particular distribution, we will call it f 1 x. This is another distribution, we will call it f 2 x and this is another distribution, the red one. So, let me use a different color here. So, it is f 1 x is the distribution of all the setosa and then f 2 x is all conditional; class conditional distribution for versicolor ok, this is and this is for virginica ok, virginica ok.

Now, once you get this class conditional density, you suppose pi k be the prior probability distribution of class k and sum of the pi k is 1. Then you just use base theorem and you can compute probability of g equal to k given x. For new data point, I only know the x, I do not know which class it belongs. I am just given the data point; can I compute the probability of that the point belongs to class k.

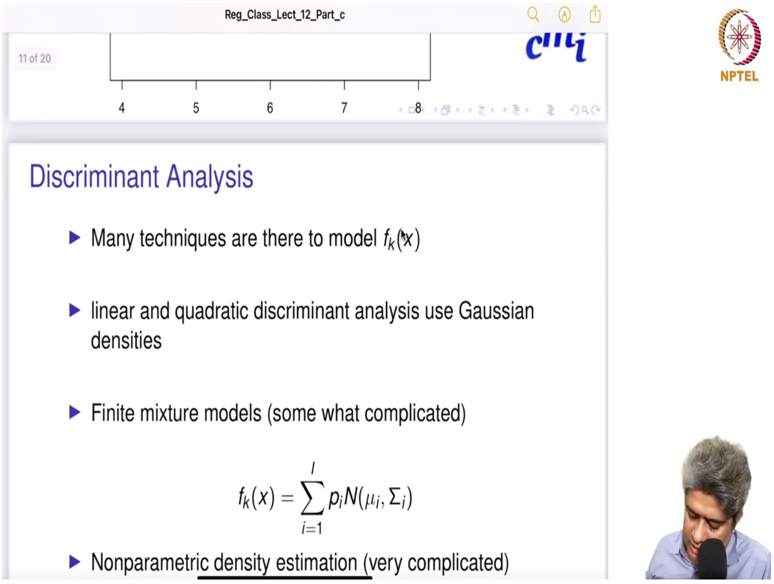
That is just apply the base theorem f k x times pi k divided by sum of the f k; f l x pi l. In terms of ability to classify having f k x is almost equivalent having quantity probability of G equal to k given x equal to x.

(Refer Slide Time: 07:33)



So, when we classify so this point, if we want to classify when we plug it in, what we found that the probability the class conditional probability or base probability is 0.861 for setosa, 0.029 for versicolor and 0.110 for virginica. So, this point belongs to virginica with probability 0.110, it belongs to versicolor with probability 0.029 and it belongs to setosa with probability 0.861. So, most likely this point is setosa.

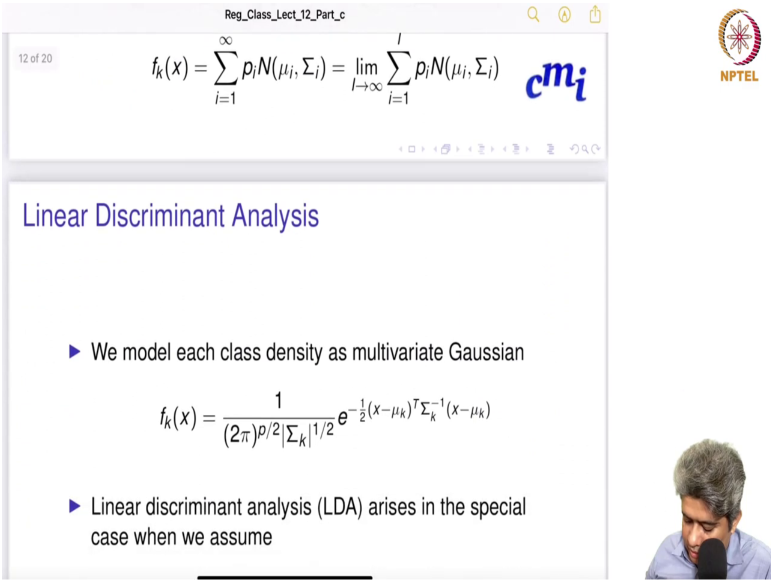
(Refer Slide Time: 08:21)



So, many techniques are there to model f k x. Now, question is how do I model f k x. There are many techniques are there to model f k x. Linear discriminant linear and quadratic discriminant analysis uses Gaussian densities. One can use finite mixture models, one can use you know nonparametric density estimation models.

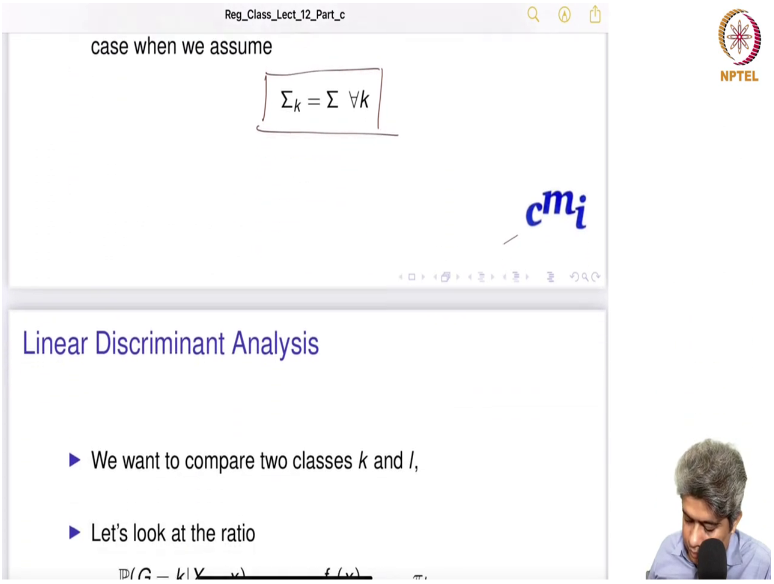
So, lot of models we can use, but in this course, we will just use assume Gaussian densities and we will only stay keep ourselves within the linear and quadratic discriminant analysis.

(Refer Slide Time: 09:09)



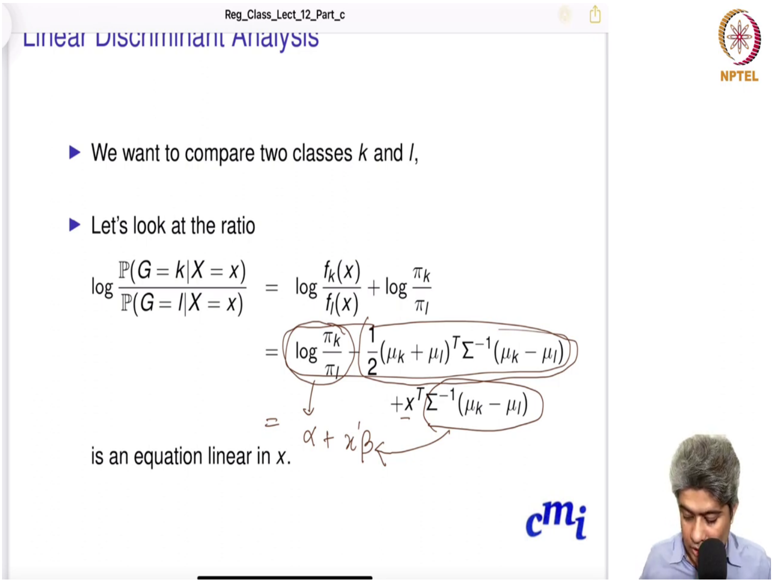
So, you can try some advanced modeling, but for this we; first we will check linear discriminant analysis.

(Refer Slide Time: 09:15)



So, if you model class density as multivariate Gaussian and then for each class, if you assume this particular thing like you know for each class covariance matrix is same, this basically homoscedasticity assumption. Then the resulting solution will be linear discriminant analysis ok.

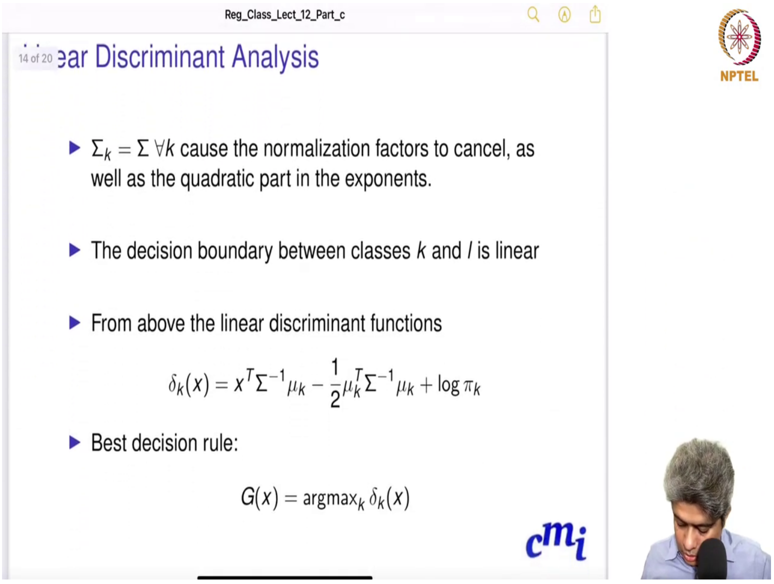
(Refer Slide Time: 09:41)



Now, so we want to suppose compare class 2 classes k and l. Let us look at the ratios, if you just look at the ratios and you can see that this pi k and pi l is already known to us and mu k mu l sigma inverse these are all known to us. So, these entire first two term is completely sort of a parameter driven. So, you can write it as sort of a this part is alpha and then x transpose sigma inverse this mu k minus mu l is sort of a x transpose beta kind of thing.

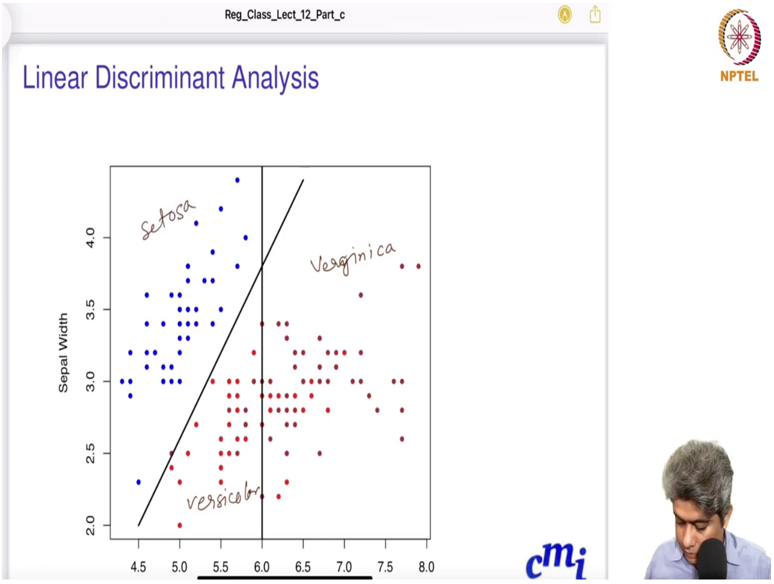
So, this is my beta so, you got; so, this ratio you are writing it as a alpha plus x transpose beta. So, this is why its this methodology is called linear discriminant analysis.

(Refer Slide Time: 10:45)



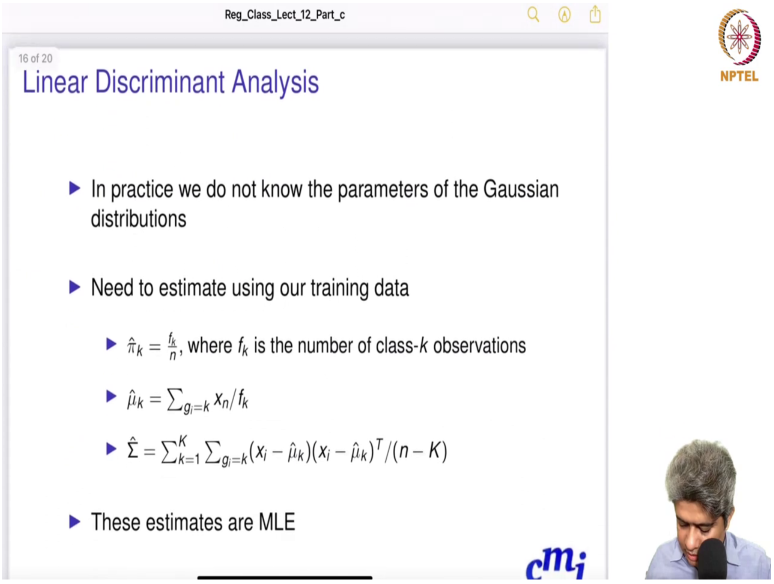
Now, if you do this linear discriminant analysis will give you a decision boundary between classes k and l and this decision boundaries are turns out to be linear, we will see about it in few slides. And from the above linear discriminant analysis you can come up with a decision boundary like this and base decision rule is just basically for each argmax of the delta k x. So, this is the best decision boundary.

(Refer Slide Time: 11:20)



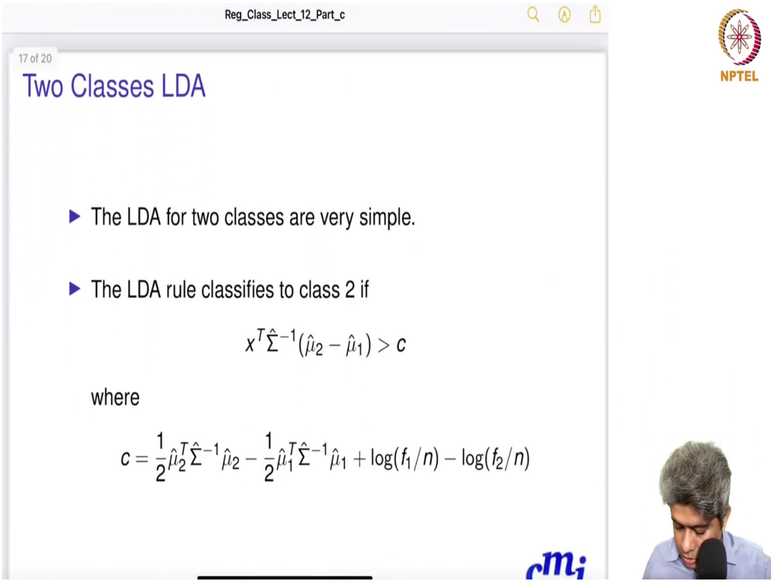
So, linear; for this particular linear discriminant analysis, we came up with a sort of this was the, so anything in this region will be effectively on the virginica linear discriminant analysis was giving us virginica, this was versicolor and this is setosa ok.

(Refer Slide Time: 11:50)

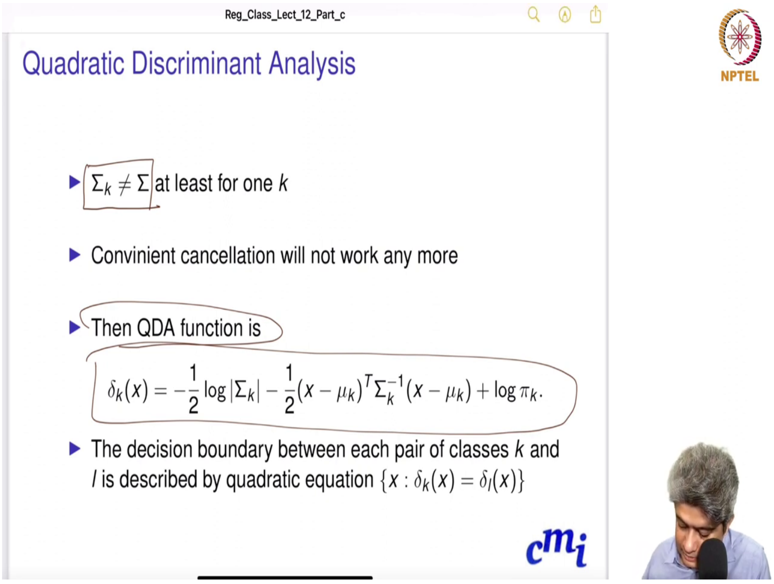


So, that is linear discriminant analysis. In practice we do not know the parameters of the Gaussian distribution. So, you need to estimate from the training data. So, how you do that? Pi k you can just take the frequency like proportion of the data points that belongs to class k mu k is simply sample mean and covariance matrix is simple sample covariance matrix. These estimates are all maximum likelihood estimates. So, you can use it.

(Refer Slide Time: 12:18)

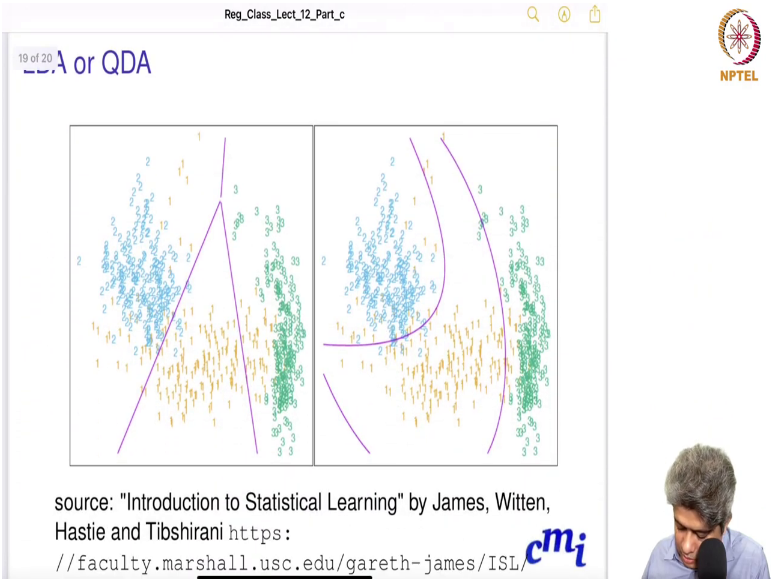


(Refer Slide Time: 12:22)

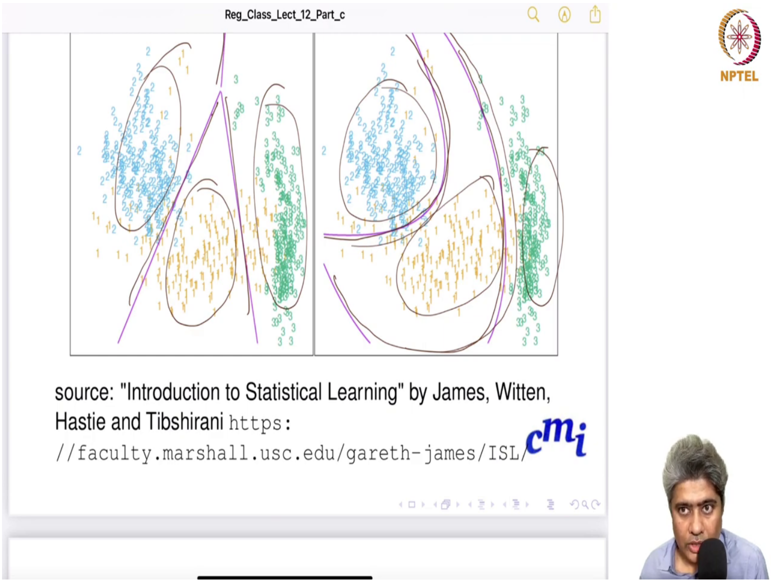


So, quadratic discriminant analysis is when if you assume sigma k are not equal to sigma and for each class you are going to compute the covariance matrix. At that time resulting solution will be QDA or quadratic discriminant analysis and decision boundary is slightly complicated, but it is not impossible.

(Refer Slide Time: 12:50)



(Refer Slide Time: 12:51)



Now, what happens you can handle it? I have taken this figure from Hastie and Tibshiranis book James, Witten, Hastie and Tibshiranis book. So, this how; this is how the linear discriminant analysis looks like, and this is how the quadratic discriminant analysis looks like.

So, they are trying to model three class problem like, the one we are doing with the Iris data set and they showed that how a discriminant analysis will behave and how linear discriminant analysis will behave. The decision boundary is like quadratic in for QDA and decision boundary is a linear in LDA. So.

Thank you very much, see you in the next video with hands on.