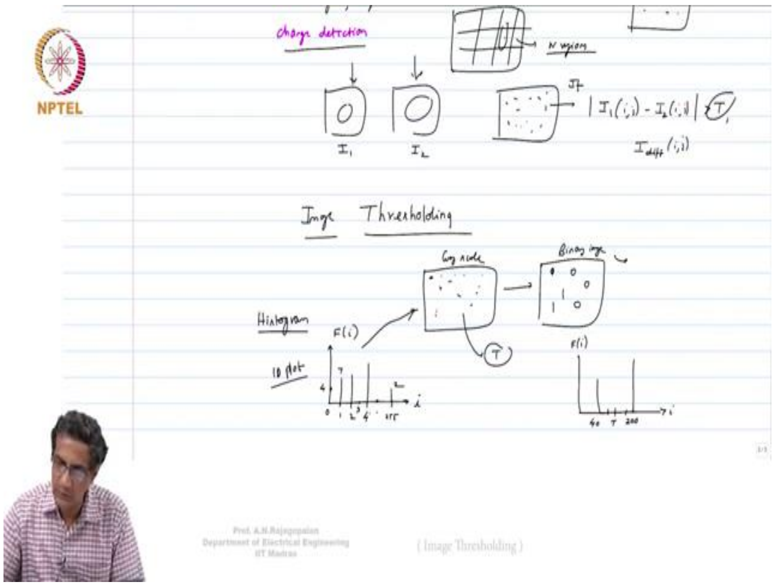
Image Signal Processing
Professor A. N. Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology Madras
Lecture 58
Image Thresholding

(Refer Slide Time: 0:16)



Now, this operation of finding a threshold is called really, is a very-very important operation called thresholding, image thresholding. Now finding a T, seems to be an important thing. For example, if I give you an image and suppose it is actually a grayscale image and suppose I asked you to convert it into let us say a binary image. Then it actually means that you need to find some kind of a threshold and then you ought to be able to examine every little intensity here, every intensity here and then you are able to say whether that intensity is greater than T.

If it is greater than T you may get 1, if it is not then you actually make it 0. So, if it is less than or equal to T, you make it 0 and that way, you kind of start entering zeros and ones and so on and then you get what is called a binary image. And there are of course many, many instances when you want to convert something that is grayscale into binary. Now, binary is one form.

You could also ask for not just binary, you could ask for what is called a multi-level image in which case you might say that it break it up into 4 regions or whatever, at 4 levels, inset of just 2 levels, you might want to break it out into 4 levels and so on. But suppose we come back to the more simpler version, which is really a binarization it is kind of a 2 class problem.

You want to be able to able to tell whether this pixel belongs to class 1 or class 2. So class 1 is like all zeros and class 2 was probably all ones. You want to kind of put them into either a basket like this or the other basket. Now, in order to be able to kind of do this, you have to be able to find T and then finding T seems to be the key. Because if you choose a T that is too high, then maybe you might actually make things, then you might do might actually end up doing things wrongly.

If you take it too small, again, no matter where you might end up doing things wrongly, therefore, there ought to be some notion of what should be an optimal value for T. So, one of the things that I actually helps you find T is really what is called an image histogram. Image histogram, what is a histogram? A histogram is something that I actually converts your 2D image into a 1D plot. So what is this 1D plot now?

So, histogram is really a 1D plot, where along the x axis you have the intensity levels, and then along the y axis you have the frequency of occurrence of these intensities. So it is like saying that x axis as values like 0, 1, 2, 3, all the way up to let us say, 255, 256 levels and all and then write your image? Suppose it has, suppose I count a, suppose in this grayscale image I count the number of times 0 occurs, maybe it occurs here, maybe it occurs there, it occurs there, I count how many times it occurs.

Suppose that occurs 4 times, then F of 0 is 4. Then I would have this, let us say F of 1 is some 7, let us say F of 2 is something else, F of this is something else, maybe red, maybe some indices just do not occur at all. It may be 3, you do not have 3 occurring at all here, maybe 4 occurs, something else occurs and maybe a 255 occurs maybe 2 times, and so on. So this plot is really called the histogram. So it is a 1D plot.

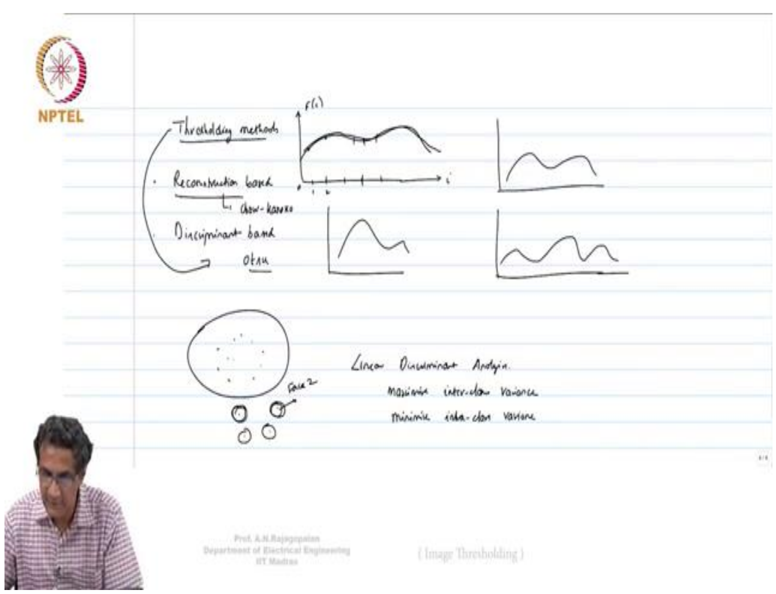
And because of the fact that we have converted information coming from a 2D and you have squeezed it down to 1D plot, then it actually means that you are going to lose information, your lost information, because you cannot go back from here to there. Even if I give you the histogram, we have not stored the stored information about where these intensities came from. All that we said was how many of them were there.

So we do not have information to kind of go back so I do not know where to put these zeros back there. So only in special cases, you might be able to go back and reconstruct, but in most cases you cannot. So the special information is gone, but that is. We are all right with that. Because that even having this 1D plot helps us a lot. Now, how does this 1D plot really help us this histogram in order to achieve a thresholding operation. What it boils down to is this. Suppose you had, suppose you are very lucky, let us say you are very lucky.

Your image simply had let us say kind of 2 values. So let me say that this is some like 40, let us say this is like 200 and this is I, after you plotted this histogram, this is what you observed. Now, your job is very easy. I mean, if you want to binarize this image, you simply say, put a kind of a threshold at anywhere in between these 2 values wherever you want.

And then you say write all the intensities in the image that have a value less than this T that you have chosen. Go to class one and all the ones that are above will go to class 2. This is very simple, this seldom happen. This kind of thing will seldom happen. So, what you will normally encounter is something like this.

(Refer Slide Time: 5:10)



You will actually encounter a histogram. Suppose I try to draw a continuous curve, you might get something like that or kind of something like this or something like that and so on. Now, you need not have just 2 more so sometimes you might also get multimodal histogram and so on, of course, all of these you have to kind of remember that i is typically a discrete value.

Therefore, even though I have drawn a continuous curve, it really means that these values exist only at 0, 1 and then 2 and so on, on 2 it has a value this, 1 it has a value this. I will just drew a continuous curve just for understanding sake. But really in an image you have only a discrete set of intensities and for which you will have the frequency of occurrence for each of those intensity values.

Now, the point is that if you have something like this and now here, you want to ask us to if you wanted to do something like something like a binarization or really a thresholding? So, binarization, let us see which is like dividing into kind of 2 classes. So, you want to know, should I put my threshold here? Should I put my threshold here? Should put my threshold? Where should I put my threshold?

So, you need some kind of a sense to pick up the T. Now, there is a host of these kind of techniques that can actually do this job what are called thresholding methods and you can kind of fundamentally classify them into 2 of these categories. 1 is called reconstruction based reconstruction based, which I will skip. Because reconstruction is something similar to a PCA that we did, in the sense that then reconstruction is like, but if you if you had this, for example, that one of the ways to do it would be to kind of would be approximate this histogram that you have with actually a sum of Gaussians.

So, you would have to find out or what should be the sigma of each Gaussian? What should be the mean of each Gaussian? What should be the weight of the Gaussian? And so on. So, basis your unknowns kind of say turned out to be that, and the whole idea is that how best can you can fit some of these Gaussian to this histogram plot.

That is why we call it reconstruction base because the error that you get you want that error to be as small as possible between the original histogram that you have with you and then a Gaussian fit, which kind of trying to do. If it is bimodal, then you have two of these Gaussian. If it is multimodal then you have, let us say if you have N modes, if you know that this you want to model it with N modes, and then you will have like N of these Gaussians.

Each one with its own mean value, and it is sigma and so on. So this is called reconstruction base, somewhat along the lines that we did for PCA and all where you want to reduce the reconstruction error. There also when we said we will pick this significant Eigen vectors and expand, then the whole idea was how do I do so that my mean SME square error is as small as possible.

So here, it is a reconstruction error. So you have this histogram, you have the Gaussians, some of the Gaussians, and they are trying to find out how, how best can this least squares fit be done. The other way to do it is really a discriminant based approach, which is what I will talk about in this course which is called a discriminant based. What this means is that, so it talks about, it is not the same as reconstruction.

So in order to give an example, by what we mean by a discriminant based approach, I will just go back to the face example that we talked about earlier. Suppose I gave you a bunch of faces. Suppose I gave you a basket full of face images. Now, this consists of let us say 4 individuals. Where face images of 4 individuals and let us say now we wanted to see how well to kind of recognize them.

One way to do it is actually go through a PCA, which will be a reconstruction based approach, we will compute Eigen vectors, or whatever Eigen faces and then we will have the feature vectors for each of these persons, for each of these people. The other way to do, so which means we choose a certain basis at which are in terms of these orthogonal eigenvectors. The other way to it is really what is called a discriminant analysis.

Linear discriminant analysis in which what you do is you kind of you pick, you kind of pick the basis that means, you pick these vectors, such that you maximize interclass variance, and minimize intraclass variance or what is called, that is between class, interclass like, between class variants, if there are 2 intra is like within class. So, what this means is that you want to choose the basis such that you get kind of classes like this.

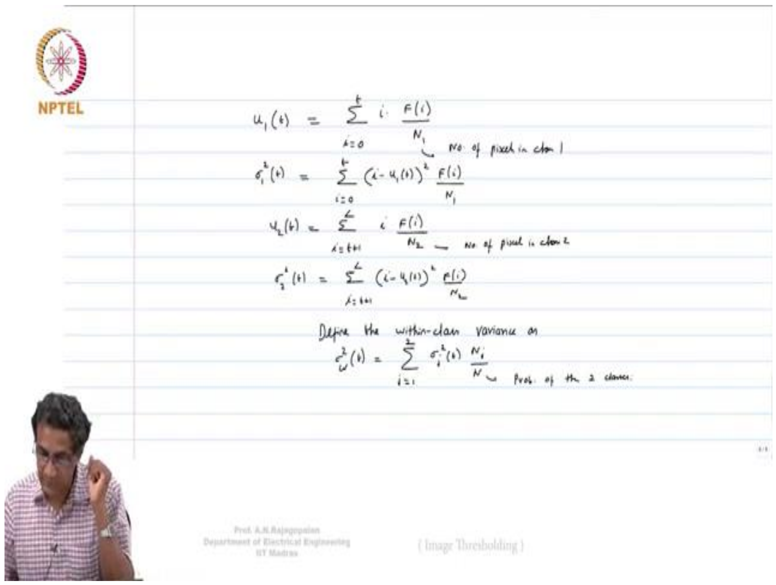
You are able to separate them, separate these classes, you do not really care about how well your reconstruction happens, but you are more concerned, or you are more focused on the fact that the separation between these classes should be high enough, but then the intraclass, that means within each class, the variance should be small, it amounts to saying that if I have a person, this is let us say person number 2, this is let us say face 2, faces of the person 2.

What it means is that even if I have images of this guy that come under different illumination, different expressions, different poses, I want to call all of them as this guy. I do not want to say that, I do not know, I mean, I do not want to say that if the expression changes and he someone else. So, within class variants, I wanted to be small, because that is when I want to ignore these sort of intrinsic variations.

I want to ignore the intrinsic variations and therefore, I want to reduce or make the within class variants as small as possible. But then the interclass, interclass is what helps me separate them, and for the interclass variance should be as large as possible. Now, the thresholding that goes by this kind of a discriminant based approach is 1 such approach that uses a discriminant analysis, a discriminant based approach is what is called Otsu.

And the reconstruction base, one of the kind of popular methods is Chow Koneke. Now Otsu, it goes after somebody's name whose name is Otsu and I want to talk about talk about how this kind of a thresholding scheme works. In more in more and more detail.

(Refer Slide Time: 12:12)



So, let us go and explain what is Otsu Thresholding? Now, the way this works is, so it is like this, I have a histogram with me and somebody gave me an image and I want to, like this is a grayscale image, but I want to kind of do a binarization and I want to convert these intensities into either 0 or 1 depending upon some criterion. Now, it could have some kind of histogram and suppose I choose some value.

So let us say I choose a small t. The point is what should be the optimal value of t and then what is my optimality? What is the optimality criterion that I am going to use in order to find my t star. So, how do I find a t star which is optimal? So, the way this Otsu goes about is like this. Now, suppose we select a threshold t, like arbitrarily.

Suppose we choose, we start with some t and then we say how to get the, get to the get to t star, which is the optimal t between let us say intensities, between 0 and L. So, here is our intensity is 0 let us say the maximum is L could be typically 255. But let us say and for the sake of generality we will say it is L. Let F of i represent the number of times the intensity i, the intensity i occurs in the image, similar to this plot that you have shown.

Let N1 be the number of pixels, pixels in let us say the number of pixels in class 1 and let N to be the number of pixels it will be the number of pixels in class 2. Now, if for a threshold T if you choose a threshold T, let us kind of look at the look at the mean let us examine, let us examine the mean and variance, let us examine the mean and variance of both the classes, of both classes, alright.

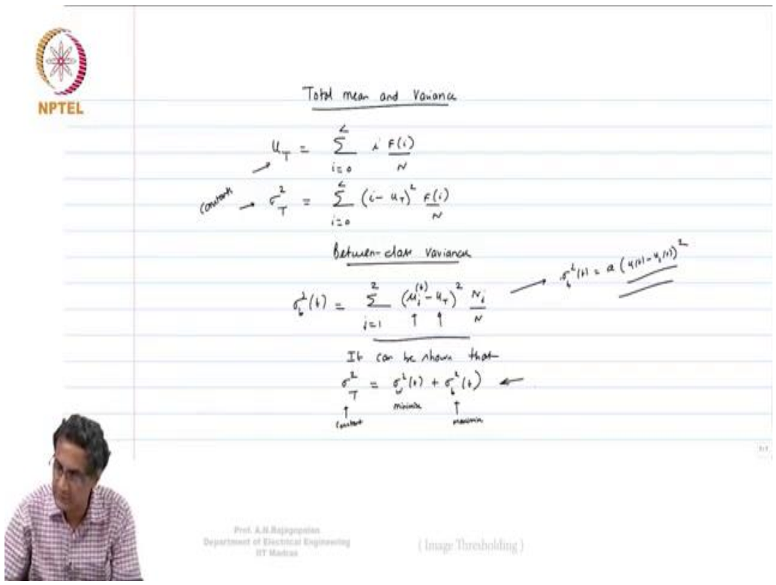
So, what does that mean? So, suppose we say that mu 1 is really the mean of the first class, this will be a function of small t. This threshold which we are choosing, so, mu 1 of t will be summation i equal to, because this just involves all those intensities that are less than or equal to t. So, we will say 0 to t then i into F of i where i is intensity, F of i is a frequency of occurrence of that number divided by N1 because class 1 has only N1. N1 is number of pixels in class 1 and therefore, the mean of this classes ifi by N1.

Now, the variance of this class let us indicate is a sigma 1 square and that is also a function of t that will be i is equal to 0 to t again, i minus mu 1 of t, now I may not repeat this mu 1 of t every time, when I write mu 1, I think we will assume that is a function of t square for F of i by N1 again. Now similarly, right for the second class we have mu 2 of t given by summation is equal to…

Now, this goes from t plus 1 because this is a class that is that is that involves intensity is greater than t. So, t plus 1 to L, i into F of i. Now, obviously divided into the number of these pixels in class 2, in class 2 and similarly, we have sigma 2 square of t that we can write a summation i is equal to t plus 1 to L, i minus mu 2 of t the whole square into F of i by N2.

Now, let us define the within class variance, the within class or what is called the intraclass variance, within class variance as sigma square w, w stands for within class, sigma square w of t is equal to summation sigma j square of t into N j by N, j equals 1 to 2 where, so where this denotes probability of the 2 classes in N j by N. That is your sigma square of t, say sigma square w of t.

(Refer Slide Time: 18:21)



Now, what about the total mean and variance of the entire image total mean and variance. If you look at the look at the total mean and variance that will not be a function of small t obviously, that is not going to be a function of small t of mu T that they submit is simply a summation. So, when you when you look at mu T you have got i is equal to all the way from 0 to L, i times F of F of i by N and I have not made this a function of t small t because it is not.

Similarly the variance is also not a function of t, So, we have not put bracket of t, is simply said is equal to i equal to 0 to L then i minus mu T whole square F of i by N. This is the overall mean what is called a total mean and variance. Now, note that right both of these are constants that mu T is a constant, sigma T is a constant these 2 sigma square T is a constant these 2 do not depend on small t.

Let us also look at the between class variance, because what we did was within class. Now, let us examine the between class variance. Between classes has indicated as sigma square b of t and So, sigma square b of t is summation j is equal to 1 to 2 mu j minus mu T the whole square into N j by N where, of course, let me well, ideally, I should be getting putting mu j of t here because mu 1 and mu 2 depend on t.

Mu t, of course, this is independent of t. This ideally I should put mu j of t but then from now on, I will drop the t. But we will assume we know for a fact that mu 1 and mu 2 that depends on small t. Now, I leave it to you as an exercise to show that like this can also be written in the form, see if we kind of look at this is looking like mu j minus mu T whole square, I know I want to leave it to you as an exercise to show that sigma square b of T can also be equivalently shown as some constant times mu 1 of T minus mu 2 of T whole square.

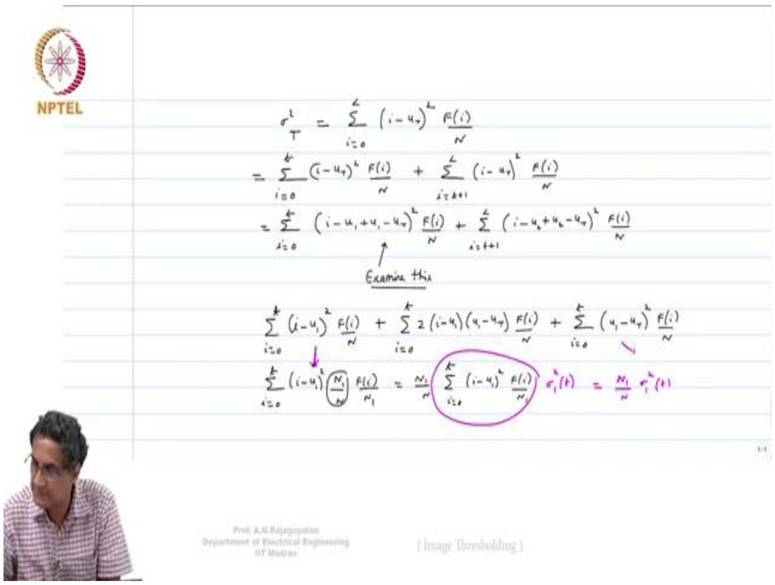
Because in a way you might be you might be able to interpret this in a more easier way than this, but then these 2 are exactly equivalent. So, when we say the between class variants, this is what we mean, either you can write it in this form or you can equally show that there is a constant you can show that, I am not going to tell you what is a you should find it out.

Now, what we want to do is we can we begin we want to show that it can be shown that sigma square T which is independent of T is in fact sigma square w of T plus sigma square b of T. Which actually has kind of a nice interpretation. What it actually means is that, because sigma square T is actually a constant and we know that this is independent of t. So, if you try to minimize within class variance, this will automatically maximize sigma square b.

Because sigma square T should remain a constant, and this is a constant, because this is a constant and minimizing the within class variance is equal to maximizing the intraclass variance or between class variance. And similarly, maximizing between class variants or if you try to maximize between class variants, then it will be equivalent to doing a minimization of the right within class variants.

So, it means that we need to do only 1 of the 2 either we should minimize this or we should maximize this. Between class either maximize between class or you should actually minimize within class variants. So, doing 1 is equal to right doing the other. Because of the fact that sigma square T is a constant. Now, we want to go ahead and show this now, how do we show that?

(Refer Slide Time: 22:46)



So, let us write down let us start with sigma square T the expression for sigma square T sigma square T is summation i is equal to 0 to L, i minus mu T the whole square F of i by N, this back. This is what we had. Now, let us split this into 2, 2 terms i is equal to 0 to T, i minus mu T the whole square, F of i by N plus summation i is equal to T plus 1 to L i minus mu T the whole square F of i by N.

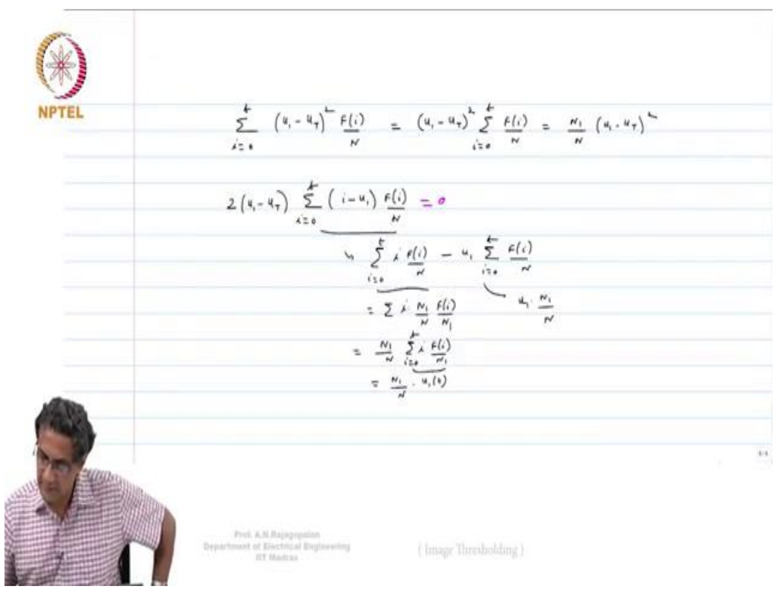
Furthermore, let us just play a small little trick we will do it as i is equal to 0 to T, I just subtract, add and subtract i minus mu 1 plus mu 1, mu 1 should be strictly a function of T, but I am not going to explicitly write that. Minus mu T the whole square into F of i by N plus summation i is equal to T plus 1 to L. Else I will add and subtract mu 2. I minus mu 2 plus mu 2.

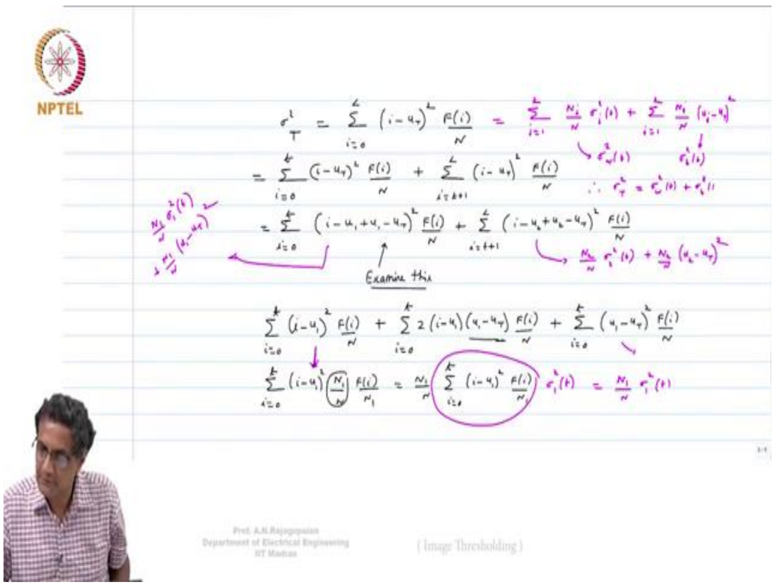
This again is a function of T, although I am not writing it explicitly minus mu T the whole square into F of i by N. Now, if you expand this, this will turn out to be summation. Now, we will be instead of looking at both the terms we will just look at 1. Examine this, examine the first term, examine this. So, if you examine this what do you get you can expand this summation i minus mu 1 the whole square into F of i by N plus summation 2 times i minus mu 1 into mu 1 minus mu T into F of i by N plus summation.

All the way, of course, I equal to 0 to T here, I is equal to 0 T here, i is equal to 0 to T here obviously, mu 1 minus mu T the whole square into F of i by m. Now, this term right we can rearrange it a summation i is equal to 0 to a T i minus mu 1 whole square into N1 by N into F of i by N1. Because we just we just multiplied and then divided by N1, right. So, F of i by now, if you pull out N1 by N this will be equal to N1 by N summation i minus mu 1 whole square into F of i by N1, i equal to 0 to T.

Now, this clearly, you can actually identify this with, this is sigma 1 square of T correct. If you can again go back and check the expression is what we had for sigma 1 square of T. Therefore, this whole thing reduces to N1 by N sigma 1 square of T. So, the first term, so the first term this reduces to this. Now, let us get a look at the last term the last term is mu 1 minus mu T the whole square.

(Refer Slide Time: 26:05)





So, summation i is equal to 0 to T mu 1 minus mu T the whole square into F of i by N. Because of the fact that mu 1 and mu T are both independent of I therefore, this can be written as mu 1 minus mu T the whole square into summation for F of i by N, i equals 0 to T. This we know is simply i equal to 0 to T if you count all the all of F of i that is simply equal to N1. Therefore, this is N1 by N into mu 1 minus mu T the whole square and now examine this middle term now, which is still remaining.

Which is i minus, so 2 times i minus mu 1. So, this term and so, out here what we see is mu 1 minus mu T is independent of i therefore, right that can be pulled out. So, we will get 2 times mu 1 minus mu T into summation i equals 0 to T mu then what do you have here? i minus mu 1, sorry i minus mu 1, I minus mu 1 into F of i by N. Now, this can be split as summation i into F of i by N, i going from 0 to T minus mu 1 which is of course independent of i into summation i equal to 0 to T, F of i by N.

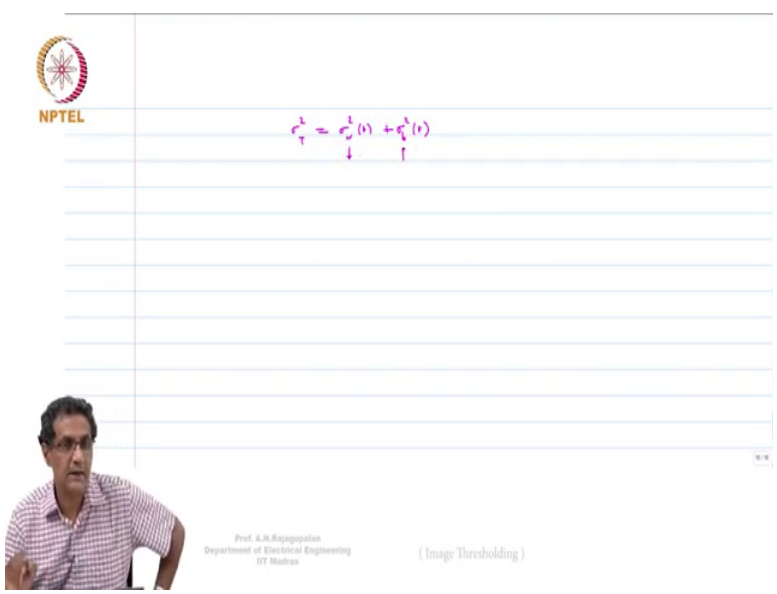
Now, if you see this is i into F of i by N and this we can write as summation i into N1 by N into F of i by N1 which is equal to N1 by N into summation i into F F of i by N1, this we know is mu 1 of T. Therefore, this will be equal to N1 by N into mu 1 of T. This is i equal to 0 to T of course. Look at this what have we got here, this summation is simply mu 1 into summation i equal to 0 to the F of i is N1 and 1 by N.

Therefore, if these 2 are exactly equal and they subtract each other therefore, this term will go to 0. Therefore, what this means is that the first term, if you look at the first term here, this term, this the first term is this 1. So, this term in fact reduces to N1 by N sigma 1 square of T which is what you had here and the second term which is N1 by N into mu 1 minus, plus and 1 by N into mu 1 minus mu T the whole square.

Similarly, the second term right when you do it from i equal to T plus 1 to L I leave it you as an exercise please show that this is equal to N2 by N sigma 2 square of T plus N2 by N into mu 2 minus mu T the whole square. And therefore, what happens is So, the sigma square T right if you add up so, here right this whole expansion will then be summation j equal to 1 to 2 sigma N j by N2 sigma square j of T plus summation j equal to 1 to N j by N, mu j minus mu T the whole square.

So, these terms come from here and this we know is nothing but our sigma square b of t, sigma square w of t and this we know is sigma square b of c and therefore, it follows that therefore, sigma square T is equal to sigma square w of T plus sigma square b of t.

(Refer Slide Time: 30:57)



Now, the implications are as follows. Given the fact that sigma square T now is equal to sigma square w of T plus sigma square b of t and so if you find a T such that it maximizes between class or minimizes within class we are done.