Indian Institute of Technology Madras
Presents

NPTEL
NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING

Pattern Recognition

Module 02

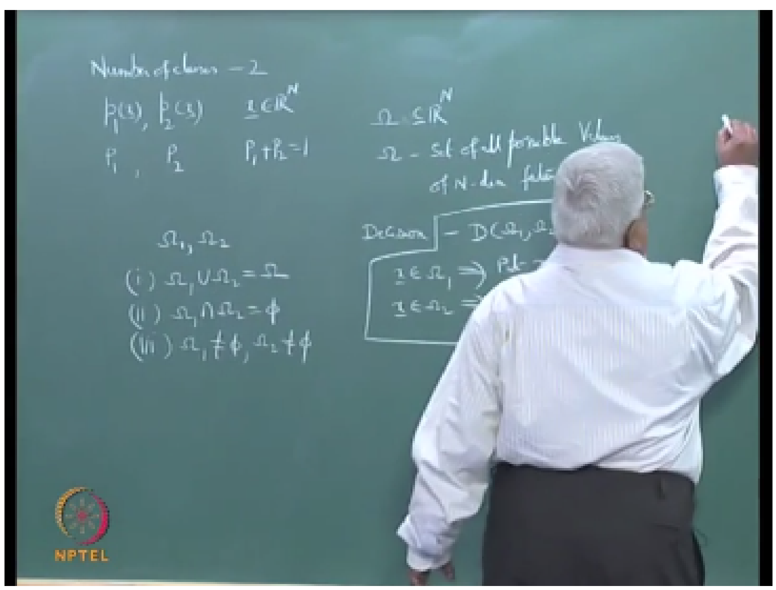
Lecture 01

Types of Errors

Prof. C.A.Murthy
Machine intelligence unit.
Indian statistical institute Kolkata

I was mentioning about the different types of errors okay, and let me give you the corresponding mathematical formulation of that let me give you the corresponding mathematical formulation so how does one give them mathematical formulation it goes like this say you have let us assume that the number of classes.

(Refer Slide Time: 00:32)



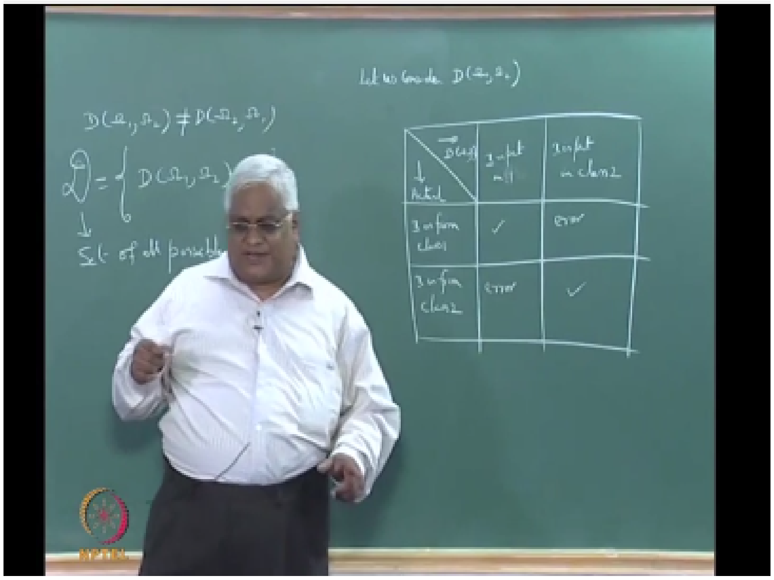
Is 2 number of classes is 2 and the conditional probability density functions for the class one is let me call it p1 and for the class 2 it is p2 so for a point X it is p1 X and for a point X here is p2 x I am assuming that X is belonging to RN and the prior probabilities this is capital p1 and this is capital p2 and p1+s p2 = these are prior probabilities okay now let us take Ω to be let us take Ω, to be the set of all possible values of the features you have N number of features the set of all such possible values that this can take.

So this that that set can be either RN R it may be a subset of RN either it is RL or it may be a subset of RN ok so for the sake of convenience I am just writing this thing to be subset of RN that means the places where X is not belonging to Ω, then p1 is zero and as well as p2 is zero is it clear to you the places where X is not belonging to Ω, that means if X belongs to Ω, complement then p2 is zero and P 1 is zero and I can as well write X is belonging to RN no problem ok Ω, is the set of all possible values.

Of the feature vectors Ω, is set of all possible values of the set of all possible values of n dimensional feature vector that is Ω, so Ω, can be RN or Ω, is a subset of RN now what is the meaning of making a decision the meaning of making a decision is we have to divide Ω, into we have to make a partition let the partition of Ω, let me just write it as Ω 1 and Ω 2let me just write the partition of Ω as Ω 1 and Ω 2 that means Ω 1 Union Ω = Ω 2 Ω 1 intersection Ω 2 is equal to null set + 3 Ω 1 is not is equal to5 and Ω 2 is also not is equal to 5.

These 2 are non empty sets the intersection is null such and the union is partition Union is the whole space and what we will do is that now we will make a decision the decision let me just represent it by the decision let me just represent it by D Ω 1 Ω 2 the decision I am just representing it by D Ω 1 Ω 2what is the meaning of this the meaning is X belonging to Ω 1 implies we will put X in class 1 we will make a decision that it is going to class 1 X belonging to Ω 2 implies we will put X in class 2.

That is the decision is this clear to you we have made a partition and then for a partition we have made a decision like this now note that this whole thing is decision that means D Ω 1 Ω 2 means actually this the decision D Ω 1 Ω 2means actually this note that D Ω 1 Ω 2.
(Refer Slide Time: 06:04)



Is not e is not equal to D Ω 2 Ω 1 why because here if X is in Ω 1 will put X in class 1 but here if X is in the first set we will put it in class 1 the first set means class 1second set means class 2 so naturally Ω 1 Ω 2 is not is equal to D Ω 2 Ω 1 first set is always for class. Second set is always for class 2 so the Ω and Ω 2 is not is equal to D Ω 2 Ω 1 okay, now let us look at this set D the set D is all such D Ω 1 Ω 2all such D Ω 1 Ω 2 that means we are looking at all possible decisions not even a single decision.

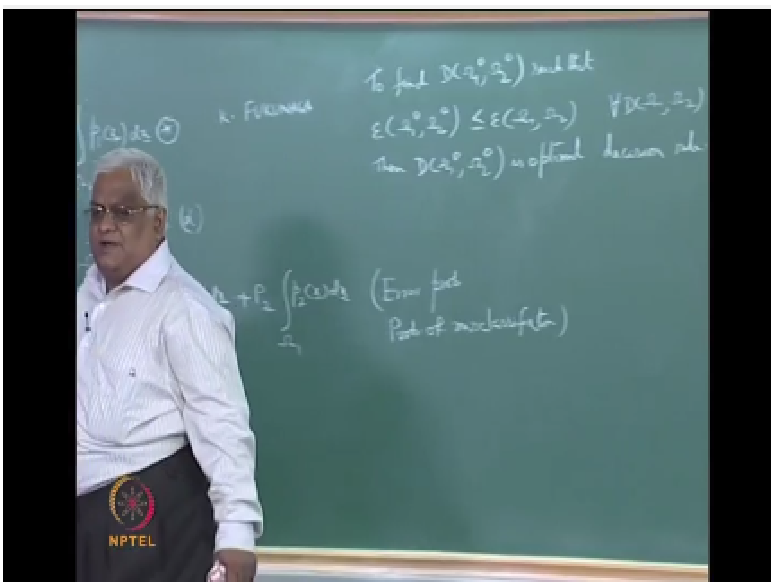
We are not considering every decision we are considering every decision we are considering when I say that this is D is the this is the set of all possible decisions is the set of all possible decisions is it clear to you this is the set of all possible decision every decision is here okay, now for a decision let us consider a single decision let us consider a decision let us consider one decision D Ω 1 Ω 2 this is one decision D Ω 1 Ω 2 now what can happen we this one this is the decision D Ω 1Ω 2 that is here X is put in Ω 1here X is put in negative x is put in class 1 X is put in class 2.

This is the decision decide this is actual that means here X is from class 1X is from class 2let me repeat if you take a point X it may be actually from class 1 or it maybe actually from class 2 now your decision it may make that X may be put in class 1 or X may be put in class 2 it is clear now what happens is if X is actually from class 1 and if you also put it in class 1 then there is no harm similarly if X is from class 2 and if you put it also in class 2 then there is no harm.

But if X is from class 1 and you will put it in class to disease and okay this is an error similarly X is from class two and you will put it in class one then this is also an enemy it is clear so for every decision by $\Omega 1 \Omega 2$ you have this now if you have three classes for two classes you have two such places where you have error then if it is three classes at how many places you will have errors it is clear it is for two classes it is 2 square - 2, 2, 4, 3 classes it is 3 square - 3 6 4 ,4 classes it is 4 square - 4, 4 into 3 12 right and if it is K number of classes.

The places where you can have this error is K into K -1ok at K number of classes the places where you can have error is K into K- 1 so then what is the mathematical formulation of this till now it is fine but how does we calculate this error well so these are the possible places now let us look at the mathematical formulation this is when X is from class1 we will put X in class 2.

(Refer Slide Time: 12:45)



So how does it happen X is from class 1and we will put X in class – we will put X in class - means X must belong to $\Omega 2$ X is from class 1 but we will put X in class 2 that means X is class to me only when X is belonging to Ω 2.We will put it in class 2 that is integral over p1 XDX or Ω 2 okay this corresponds to this if I write it as start this corresponds to this star and if I write this as α this is X is from class tube but we will put it in class 1 this is your α and this naturally depends on what this Ω 1 and Ω 2 are for a partition are for a decision rule.

D Ω 1 $\Omega 2$ this is one error expression for one error this is expression for one error okay now but what is the probability that X is from class 1 the probability that X is from class 1 is P 1 and the probability that X is from class 2 is P2 and so the total error is if I call it as Ɛ and this is dependent on Ω 1 and Ω 2 this is actually P 1this is the total error probability this is error probability this is also known as probability of misclassification this is also known as probability of misclassification.

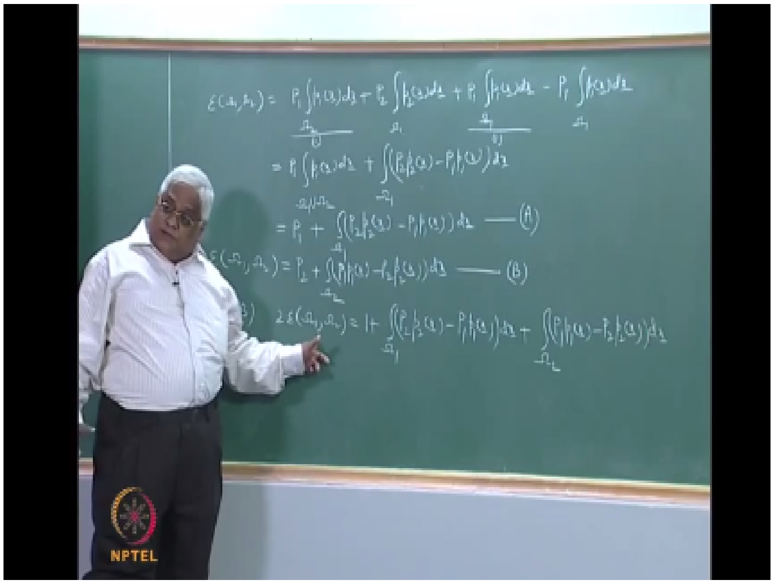
I would like to make a small remark when I wrote this expression I am not bothered about cost of misclassification I am only bothered about probability of misclassification that is if the point from class one if it is put in class two what is the cost of misclassification that is the cost of misclassification can be enormous whereas probability may be small our probability may be hi misclassification may be I mean the cost may be small that is basically I was telling you that for different error probabilities are for different errors whether you will give equal weight or same weight it depends on the problem under consideration here.

I am assuming same weights I am not bothered about the cost of misclassification our I am assuming that the cost of misclassification is same I need to multiply this thing by the corresponding cost or misclassification I need to multiply this by the corresponding cost of misclassification which I am not doing it that means I am assuming that the misclassification costs are same I will if I get time I will discuss about this cost of misclassification but here you can assume that the cost of misclassification is same.

When it is different then how does one look at this problem this is given in you can look at Fukunaga book on statistical pattern recognition where the expressions for cost of misclassification they are also given there you can look at Fukunaga book on statistical pattern recognition introduction to statistical pattern recognition I think it is the academic press k. Fukunaga now let us see how to minimize this what is the meaning of minimization the meaning of minimization is we need to get hold of that specific Ω 1 and Ω 2 for which this value is less than every other such decision the meaning of minimization is the meaning of minimization is to find D Ω 1 0Ω 2 0.

Such that Ɛ of Ω 1 0Ω 2 0 is less than or equal to Ɛ of Ω 1 Ω 2 for all DΩ 1 Ω 2 belonging to we need to find Ω 1 0 Ω 2 0 such that the error corresponding to Ω 10 Ω 2 0 0 is for actually optimal for optimal Ɛ Ω 1 0 Ω 2 0the error is less than all errors less than or equal to then we say that this is optimal such decision rule then we say that this D Ω 1 0 Ω 2 0 is optimal decision rule then D Ω 1 0Ω 2 0 is optimal decision rule optimal decision rule now how does one find it.

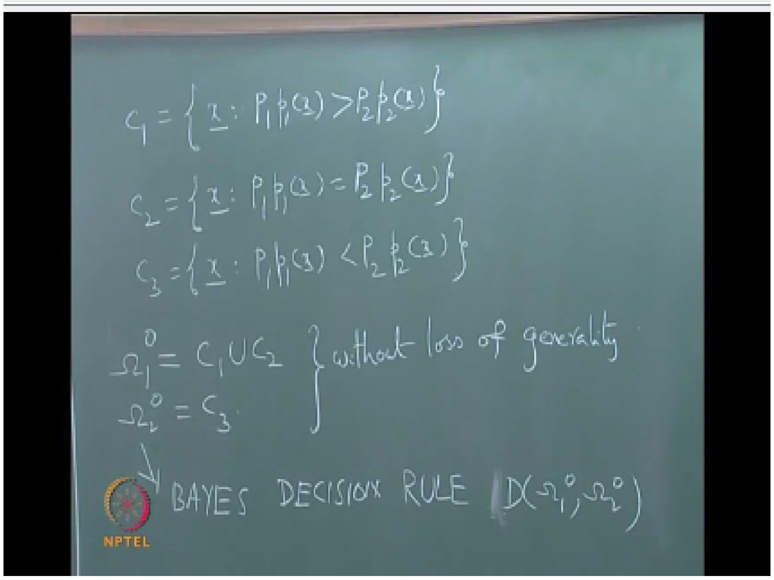
(Refer Slide Time: 21:07)



So this is the expression so what I will do is I will just add I will just add P 1 P1 X DX / Ω 1 and I subtract the same quantity. I subtract the same quantity then what will happen look at this term and this let me just write it as 1 this is P 1 P 1 X / Ω 2 this is P 1 P 1 X over Ω 1 so if you just take the addition this is just going to be over Ω 1 Union Ω 2 okay now let us look at this and this + this is integral over Ω 1this is integral over Ω 1 P 2 P 2- P 1 P 1 X DX so what is this note that P 1 is probability density function defined over the whole space Ω so the integral is 1 so this value is actually P 1 + so similarly instead of adding and subtracting P 1 P 1 X or Ω 1 if you add and subtract P 2 P 2 X 4 Ω 2then what you are going to get is this will be P 2 + integral over Ω 2this will be this will be P 2 +integral over Ω 2 capital P 1multiplied by small P1.

This is capital P 1 and this is capital P 2 multiplied by small P 2 X is P 1 P 1 - P 2 P DX now we will add this and this if we add a and B then Ɛ and Ɛ this will be 2 Ɛ P 1 + P 2 is 1 P 1+ P 2 is 1 + you have this and you have this so we need to minimize this we need to minimize Ɛ Ω 1 Ω 2 is same as we need to minimize 2 Ɛ Ω 1Ω 2 1 is the constant so we need to minimize this as well as this we need to minimize this as well as this right now let me write one set let me just call it as C 1 C 1.

(Refer Slide Time: 26:18)



Is the set of all X and c2 is and I am adding one more set c 3 is now. We need to minimize this and we need to minimize this that means we need to choose Ω 1 and Ω 2 in such a way that this whole expression is minimized and this whole expression is minimized now how does one choose Ω 1 and Ω 2 now let us look at the meaning of minimization this particular quantity P 2 P 2 - P 1 P 1 depending on where X is located it may be greater than 0 or equal to 0 or less than 0 now since we need to minimize this whole.

What one can do is that let us look at all those places where this is less than 0and add them up let me tell you once again let us look at all those places where P 2 P 2 - P 1 P 1 is less than0 and we will add then you will get the maximum I mean you will get the minimum possible such quantity here is this clear similarly let us look at all the places where p1, p1 is less than p2, p2 here and integration means basically summation will add all possible such negative quantities here and as well as here then that will make this thing to be the minimum most.

That will make this one to be the minimum most so that means let us find out let us look at the set C 1 P 1P 1 is strictly greater than P 2 P 2 P 1P 1 is strictly greater than P 2 P 2then if we take C 1 here this whole summation will be minimized let us look at C 3 P 1 P 1 is strictly less than P 2P 2 he then we can take c3 to be Ω 2 but then what about c2 p1, p1 is equal to p2p2 note that if p1,p1 is equal to p2, p2 whether you keep it here or here the difference P 1 P 1 - P 2 P 2corresponding to that that will be 0.

So it is not adding to anything are you understanding this is not adding to anything so it does not matter whether you keep it with the first set R the second set are a part of c1 you keep it with first set a part of C I am a part of c2 you keep it with first set and the rest you keep it with the second set no problem you can put the whole off c2 either withc1 R with c3 are a part of c2 you can put it with the first set and that the rest you can put it with the second set in any way you can just do it is not going to make any difference to the quantity of error value.

So the optimal set Ω 1 0 Ω 1 optimal without loss of generality let us take it to be c1 union C 2 and Ω 2 0 let us take it to be C 3 this is without loss of generality without loss of generality that means there are many optimal decision rules there can be many optimal decision rules all of them giving the same value of the error probability we can take any one of them and every such decision rule is actually called as based decision rule.

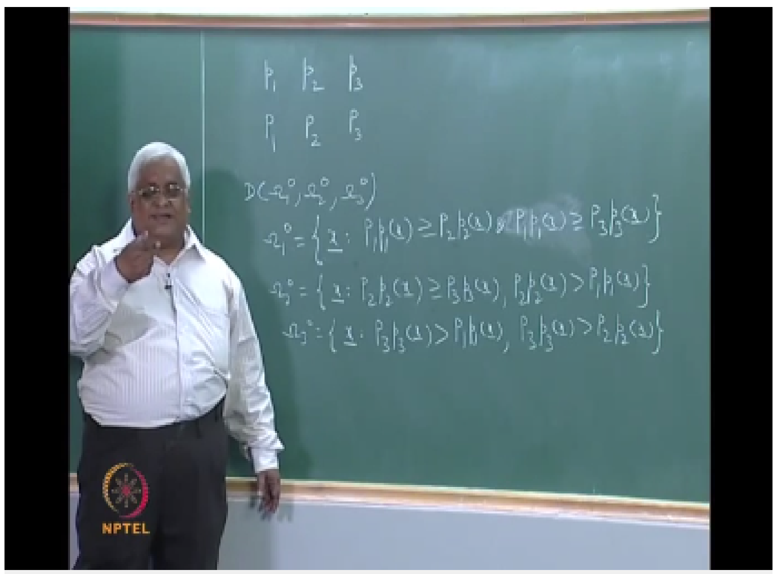
This decision rule is based decision rule and it provides you optimal decision optimal from the point of view of minimizing the probability of misclassification minimizing over what minimizing over every decision rule whatever decision rule that you will take its probability of misclassification.

Will be greater than or equal to that of the base probability of misclassification whatever decision rule you take note that our Ω1 and Ω2 I have not put any condition any decision rule you can take no problem whatever decision rule that you take it is probability of misclassification be greater than or equal to that of the base decision rule this is base decision rule I mean these are the places where P1 P 1 is equal to P 2 P 2 that is basically the decision boundary between.

The classes one side it will be greater than another side it will be less than the boundary will be p1, p1 is equal top2 p2 okay, most of the cases the boundary region will be p1, p1 = p2, p2 and these are the this is based decision rule this is the based shin rule D Ω 1 is 0 Ω 2 0 based decision rule this is d Ω 1 0 Ω2 0 now note that here this decision rule is given for two classes you can have the corresponding thing 4 3 4 5 in fact any number of classes in fact in this one there is also some set of exercises okay, you may take this thing as an exercise to see whether you can solve.

It or not the exercise is for 3 class classification problem for 3 class classification problem for 3 class classification problem you derive the decision rule which minimizes the probability of misclassification you derive the decision rule which minimizes the probability of misclassification and the decision rule will be like this for 3 class classification problem the decision rule will be like this.

(Refer Slide Time: 35:35)



If you have three classes your conditional probability density functions will be P1 P 2 P 3and your prior probabilities will be P 1capital P 1 capital P capital P 3 and the decision rule the optimal decision rule if X is belonging to this you will put it in class 1 X is belonging to this put it in class 2 X is belonging to Ω 3 put it in class 3 so what is your Ω 1 Ω 1 is the set of all X P 1 P 1 X is greater than or equal to P 2 P 2 X + 3 3P 3 XP 1 P 1 X is greater than or equal to P2 P 2 X maybe I will just write clearly and okay, now Ω 2 0 the set of all X for which P 2 P 2 X is greater than or equal to P 3 P 3 X and P 2 P 2 X is strictly greater than P 1 P 1 X and Ω 3 0 is the set of all .

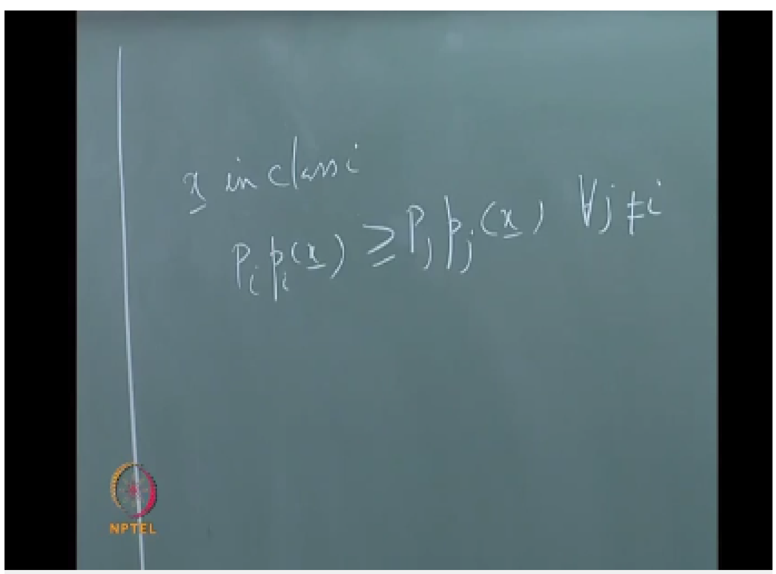
So it basically follows whatever we have got in the two class case he lost P 1 P1 X is greater than P 2 P 2 greater than P 3 P 3 whereas since I would like to make the Union to be the whole space so I included the equality part here I include the equality part but I think P1 P 1 if it is equal to P

2 P 2 and it is also equal to P 3 P 3 that is here I include the Equality part here and for P2 P 2 and P 3 P 3 this equal to I included here and here there is strictly inequality so that I made this thing the Union to be the whole space the Union tome the whole space generally.

What 1 you will see in the books is that at least for one of the classes the strict inequality should hold that is what you will see in books that means at least for one of the classes strict inequality is holding means wherever all of them are same they are not putting it into any class whenever all of them are same then they are not putting them into any class that is fine there is any principle there is nothing wrong therein principle there is nothing wrong because they have not put the condition.

That is exhaustive you need not have to classify each and every point if there is some such confusion they have just kept it like this here I have put all the points in either first class or second class or third class so except for that small thing which is of negligible difference this is the rule the same rule the similar rule you can extend it to M number of classes which is what I have shown in the slide to you for a capital M number of classes the basic thing is that you put.

(Refer Slide Time: 40:08)



It in you put X in class I If PI x is greater than or equal to pj ,pj x for all J not is equal to I this is the general thing that you will find in books P IPI X greater than or equal to pj, pj x for all J not is equal to I we are all we have our background is basically computer science we must have looked at many cases where you have ties right whenever you have in many algorithms many places you must have got some ties so for ties generally.

We have some sort of a compromise okay, here also you need to have some such compromise because ties are not making any difference to the misclassification probability so wherever you would like to put them you just put it does no t matter okay I think I will stop here.

End of Module 02-Lecture 01

Online Video Editing /Post Production

M.Karthikeyan
M.V.Ramachandran
P.Baskar

Camera

G.Ramesh
K.Athaullah
K.R.Mahendrababu
K.Vidhya
S.Pradeepa
D.Sabapathi
Soju Francis
Selvam
Sridharan

Studio Assistants

Linuselvan
Krishankumar
A.Saravanan

Additional Post –Production

Kannan Krishnamurthy& Team

Animations

Dvijavanthi

NPTEL Web &Faculty Assistance Team

Allen Jacob Dinesh
Ashok Kumar
Banu.p
Deepa Venkatraman
Dinesh Babu.K.M
Karthick.B
Karthikeyan.A
Lavanya.K
Manikandan.A

Manikandasivam.G
Nandakumar.L
Prasanna Kumar.G
Pradeep Valan.G
Rekha.C
Salomi.J
Santhosh Kumar Singh.P
Saravanakumar.P
Saravanakumar.R
Satishkumar.G
Senthilmurugan.K
Shobana.S
Sivakumar.S
Soundhar Raja Pandian.R
Suman Dominic.J
Udayakumar.C
Vijaya.K.R
Vijayalakshmi
Vinolin Antony Joans

Administrative Assistant

K.S. Janakiraman

Principal Project Officer
Usha Nagarajan
Video Producers
K.R.Ravindranath
Kannan Krishnamurty

IIT Madras Production

Funded By
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in