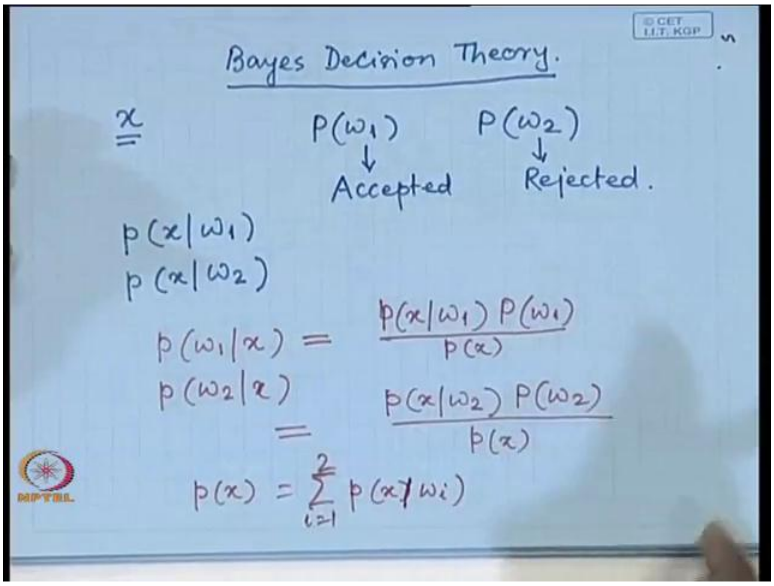
Pattern Recognition Application
Prof. P. K. Biswas
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 6
Bayes Decision Theory
(Contd.)

Good morning, so we will continue with our discussion on Bayes theory, the discussion that we started in our last class.

(Refer Slide Time: 00:33)



So, we were discussing about Bayes decision theory. So, what we talked about yesterday is that, suppose in a particular classification or pattern recognition application domain, we have a set of objects or we have for that particular object, we have an observation say x. So, I have said that x is an observation, and based on this observation we have two classify the object in one of the two classes. So, taking that particular example of a manufacturing industry, we have to classify that object either into the accepted category or in the rejected category.

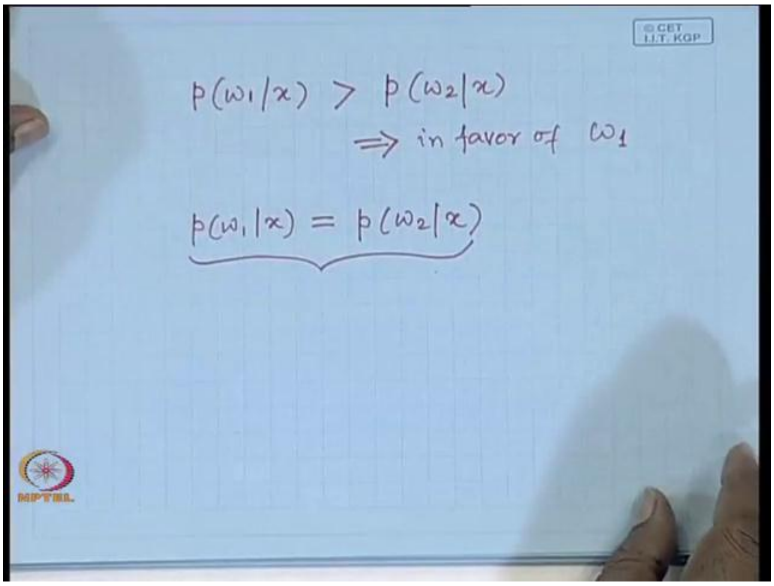
So, for doing this job we have said that we have some a priori probability that is p of omega 1 and p of omega 2. However, omega 1 this category you have said that let us assume that this is the category of accepted objects. And omega 2 is the category of rejected objects. And p of omega 1 is the very a priori probability that the object will be accepted and p of omega 2 is the pro a priori probability that the object will be rejected.

Now, in addition to make our decision we have said that, we have an observation x. And for this observation x, also we have class conditional probability density function that is probability x given omega 1 and also probability x given omega 2. So, these two are the class conditional probability density function which we have to estimate based on the measurements on x or objects which are unknown to belong to class omega 1, and the measurement of x from the objects which are known to belong to class omega 2.

So, from this two we have two estimates this p of x given omega 1 p of x given omega 2 omega 2. But finally, our classification problem is that we have an observation x. And based on this observation x we have to put this object either in category omega 1 or in category omega 2 or effectively what we have to find out is p of omega 1 given x and p of omega 2 given x. So, these are the two probability measures, which we say of the posteriory probability. And based on these two probability measures we have to decide whether the object has to be classified to class omega 1 or the object is to be classified to class omega 2.

So, here p of omega 1 given x, we have said from Bayes rule that this is nothing but p of x given omega 1 into p of omega 1 upon p of x. Similarly, p of omega 2 given x is nothing but p of x given omega 2 into p of omega 2 upon p of x while you find that this p of x appears in the denominator of both these expressions, what this p of x is nothing but p of x given omega i take the summation or i is equal to 1 to 2. So, this is what we get from Bayes rule.

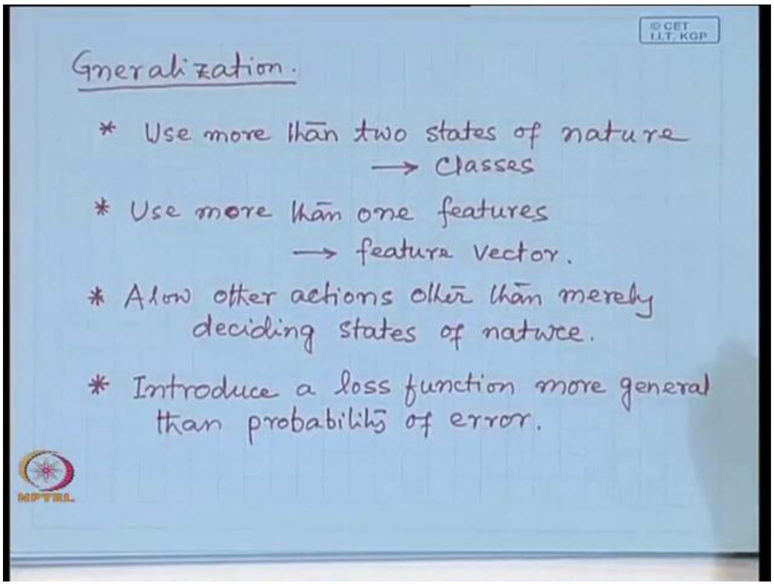
(Refer Slide Time: 05:01)



And then our decision was like this that, If p of omega 1 given x is greater than p of omega 2 given x then we decided in favour of class omega 1. So, in favour of omega 1 that means we decide that the object having this observation x belong to class omega 1. If it is otherwise, that p of omega 2 given x is greater than p of omega 1 given x then we will decide that the object belongs to thus omega 2.

However, we have one condition that if p of omega 1 given x becomes equal to p of omega 2 given x. So, this is the case when we cannot take any decision because the object lies on the decision boundary between the class omega 1 and class omega 2. So, it may be both in class omega 1 as well as in class omega 2. So, this is a case where decision cannot be taken. So, this was the basic Bayes decision theory however, we have taken that we have two class omega 1 and omega 2. And our decision was one of the two decisions either the object is to put in class omega 1 or the object has to be put in class omega 2. Now there can be a generalisation of this Bayes theory.

(Refer Slide Time: 06:47)



So, the generalised Bayes theory can be put something like this. So, instead of classes let us call it the states of nature. So, earlier in the example that you have taken there are only two states of nature. Now, we can generalise it having multiple states of nature. So, the first generalisation is like this that, use more than two states of nature. And then this state of nature in our case, in this classification problem is nothing but the classes. So, that will be our understanding when we talk about when we say that the states of nature it is nothing but the classes. In the earlier case we have taken a single observation x. So, based on this observation x we have tried to decide whether we have to put the object in class omega 1 or we have to put in class omega 2.

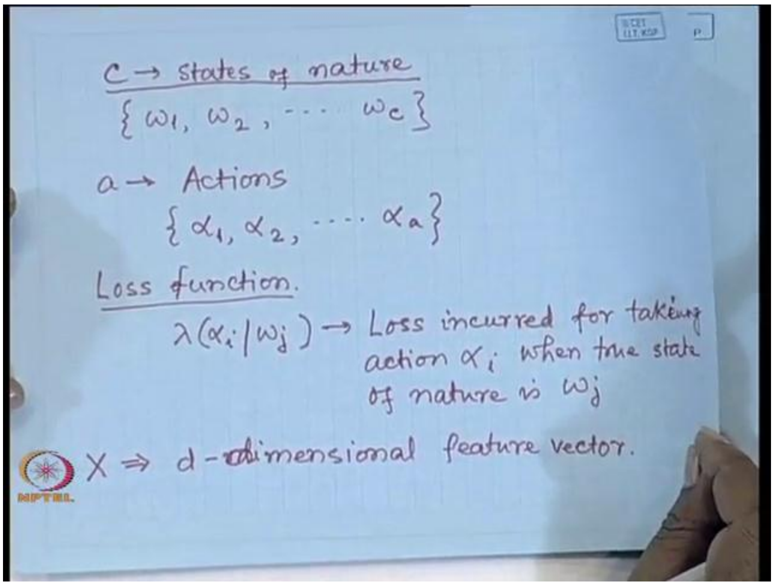
Now, in the generalisation we allow more than one observations that means instead of having a single feature we have allowed feature vector. So, use more than one feature so that means we are going for feature vector. The other generalisation is in the earlier case we had only two actions that is either decide about class omega 1 or decide about class omega 2. Now, we can allow a number of fractions instead of just deciding whether this belongs to class omega 1 or this belongs to class omega 2. So just this case I have shown that if p omega 1 given x is equal to p omega 2 given x I cannot take any decision.

So, this particular fact that I cannot take any decision, I can also define this is as action that I cannot decide to which class does the object belongs that also I can call as an action. This no decision is also action. So, I can allow other actions other than merely deciding states of measure. So, when as I said the states of measure in our cases are classes. So, merely deciding whether the objects belong to class omega 1 or the object belongs to class omega 2. Apart from that I can take other actions as well.

And the fourth generalisation is this that in our case our decision was based on a posteriory probability that if p of omega 1 given x is greater than p of omega 2 given x then we decided to be in favour of omega 1. If it is other way round that is p of omega 2 given x is less than is greater than p of omega 1 given x then we decide in favour of class omega 2.

So, in this case we can have a more generalised criteria based on which we can decide about the states of nature which is called a lost function. So, introduce a lost function, which is more general than probability or let us say probability of error. Because as we said that, if we decide in favour of omega 2 then the probability of error is p of 1 given omega x. And the purpose was to reduce the probability of error in simple ways decision theory. So, as I said that will have more than one states of nature.

(Refer Slide Time: 12:22)



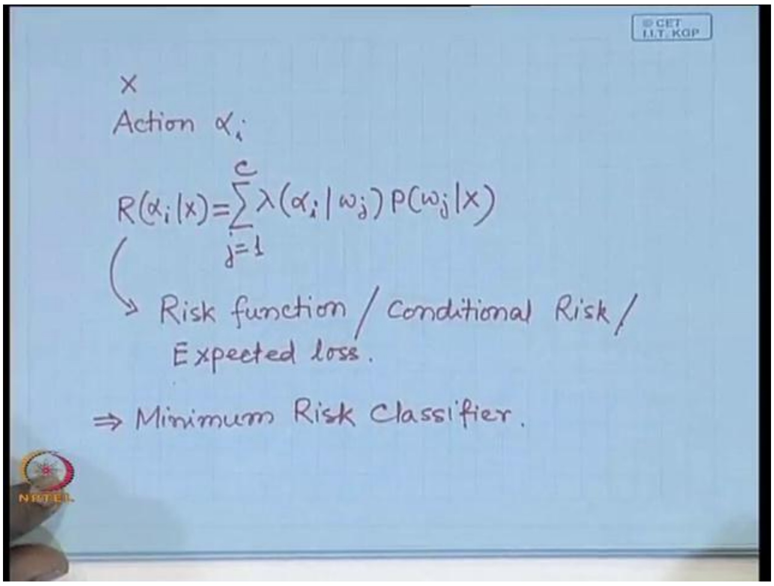
So, let us assume that there is C numbers of states of nature or C number of classes to be more specific in our application. And those classes let us call a set of classes omega 1, omega 2 up to say omega c. So, there is C states of nature. Now, about the actions as you also said that will also allow actions rather than merely deciding about the states of nature.

So, suppose there are a number of actions. So, I have this state of actions which I represent as alpha 1, alpha 2 up to alpha a. So, this is the state of actions that I have apart from simply deciding that whether the object belongs to class omega 1 or it belongs to class omega 2 and so on. So, I have this number of actions and I said that we introduce a lost function which is more general than probability of error. So, this lost function I represent as lambda alpha i given omega j. So, this means that if the actual states of nature is omega j however we take an action alpha i.

Then, the loss incurred while taking this action alpha i when, the actual state of nature is omega j is lambda alpha i units. So, this is loss incurred for taking action alpha i when true state of nature is omega j. And the fourth-generation generalisation is said that instead of considering a single feature, we will consider a feature vector. So, here let us assume that we have a feature vector that instead of a single feature we have a feature vector that is a multiple number of features of multiple observations of various parameters, which is a vector x feature vector x.

And this feature vector x is d-dimensional. So, these are things that we have seen number of states of nature given by omega 1 to omega c. We have a number of phase actions from alpha 1 to alpha a. We have a general lost function which is given by lambda alpha i given omega that means the loss incurred for taking an action alpha i when the true state of nature is omega j. And we consider a feature vector x which is a d-dimensional feature. Now, let us see that how this decision rule in this generalised based theory has to be taken.

(Refer Slide Time: 16:37)



So, suppose we have an object and for that object we have made an observation vector or the feature vector given by x. So, x is the feature vector, which is of dimension d. And for this feature vector we take an action alpha i. So, as we said earlier that the loss incurred for taking an action alpha i while the two state of nature is omega j is given by lambda alpha i omega j.

So, here for this feature vector x from this observation x we have taken an action alpha i, but we do not know what is the true state of nature. It may be omega 1, it may be omega 2, it may be omega 3 and so on. They are true state of nature may be anything. So, for each of these I will incur a loss function. So, if the true state of nature is omega j, I will incur a loss which is given by lambda alpha i given omega j. And if the probability that the two state of nature is omega j given the feature vector x then the average loss or the average risk can be computed like this.

The average risk or expected loss can be R alpha i given x is equal to lambda alpha i given omega j into omega j given x. x is my observation vector, take the summation over j is equal to one to c. Because I do not know what is the true state of nature. So, if the true state of nature is omega j then my lost function is lambda alpha i given omega j. Then to multiply this with the probability of true state of nature being omega j given my observation vector x.

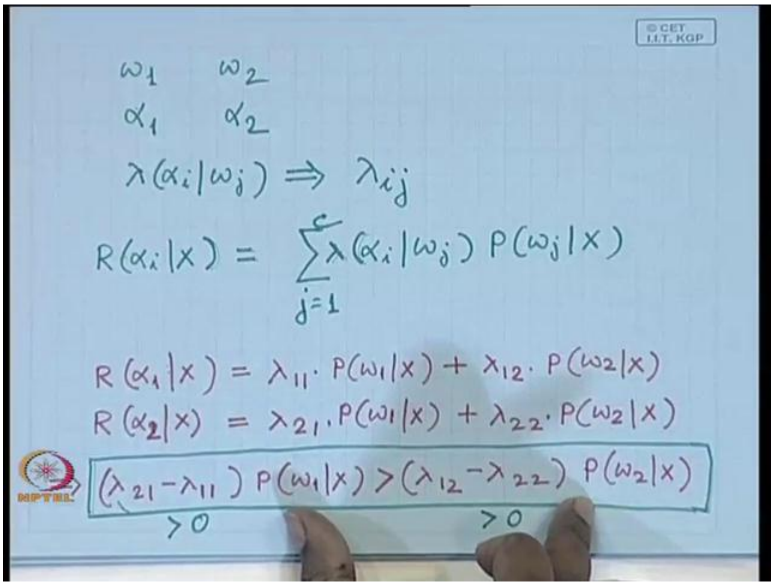
So, this is lambda omega a alpha i given omega j into p given j omega x take the submission over all the states of nature, that is j is equal to one to c that gives you the expected loss. So, which we are calling as R of alpha i given x. So, this expected law is also called a Risk function or we can also call it Conditional risk because it depends upon x. It is also the expected loss. So, any decision or any action alpha i that I had to take that particular alpha, that particular action for which this risk is minimum or the expected loss is minimum.

Unlike, in the previous case where we used only two classes or decision was taken in favour of that class, which gives us minimum error. That is if I decide in favour of omega 1, I make sure that the error is minimum which nothing but p of omega 2 units. Similarly, if I decide in favour of omega 2 my error is p of omega 1 given x which is minimum in that case.

In the generalised case, I had to take that action alpha i for which this risk alpha R alpha x is given. So, accordingly this is also called minimum risk classifier. So, in this generalised case the kind of classifier that I have is a minimum risk classifier. Now, let us see various derivatives of this minimum classifier, it is this minimum classifier which under different conditions leads to different kinds of classifier switcher actually in use.

So, let us see two category cases. Suppose, I have two classes omega 1 and omega 2 or two states of nature omega 1 and omega 2. So, in this case if I assume that the action means saying whether the object belongs to class omega 1. So, alpha 1 means that decision that object belongs to class omega 1, alpha 2 means decision that object belongs to class omega 2.

(Refer Slide Time: 22:23)



So, I have two states of nature one is omega 1 other one is omega 2. As I have this omega 1 and omega 2 and also I have action the alpha 1 and alpha 2 reserve the actions. So, action alpha 1 means the decision that the object belongs to class omega 1, action alpha 2 means deciding that object belongs to class omega 2, so over here now if I write lambda alpha i given omega j as say lambda i j just for simplicity of expression. So, lambda i j means lambda alpha i given here. That is the loss incurred for taking at an action alpha i when the two state of nature is omega j.

So, for this two class problem I can have risk function or of alpha i given x, as we said is nothing but lambda alpha i given omega j into probability omega j given x. We took the summation for j Is equal to one to c for all possible two state nature. So, in our two class problem this expression simply becomes that if I take action alpha 1, so I will have R alpha 1 given x.

This will be lambda 1 1 that means lambda alpha 1 given omega 1 is that into p of omega 1 given x plus lambda 1 2, that means lambda alpha 1 given omega 2 into p of given omega 2 given x. If I expand the lost function the expected loss for taking an action alpha i on and observation vector x then this is the expansion of the expected loss function or the risk function.

Similarly, I have the other option of taking action alpha 2, I can take one of this two actions. So, the risk involved for taking action alpha 2 on observation vector x is nothing but lambda 2 1 into p of omega 1 given x plus lambda 2 2 into p of given omega x. Now, I said that I have to take that action for which the risk is minimum.

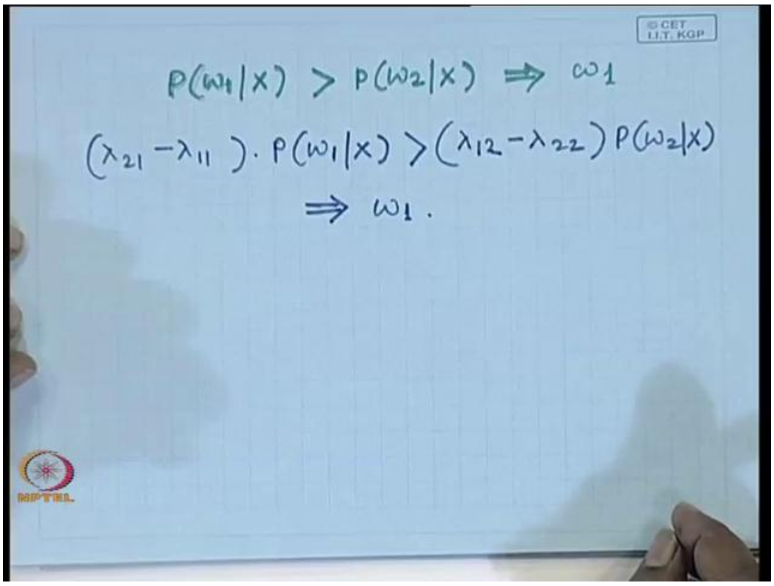
So, if I find after computation of our alpha 1 given x and of alpha 2 given x that are of alpha 1 given x is less than are of alpha 2 given x then have to take action alpha 1. If it is otherwise that are of alpha two given x is less than alpha 1 given x then I have to take action alpha 2. So, under that condition for deciding in favour of class omega 1 or taking an action alpha 1 higher are of alpha one given x will less than that of alpha 2 given x. You can find that this leads to a condition that lambda 2 1 minus lambda 1 1 into p of omega 1 given x. This has to be greater than lambda 1 2 minus lambda 2 2 into p of omega 2 given x.

So, this is the condition if I have to take this decision in favour of class omega 1 or if I have to take action alpha 1. Now, over here you find that this lambda 1 1 means taking an action alpha 1 when the two state of nature is omega 1. And as we are seeing in our case taking an action in alpha 1 means deciding that object belongs to class omega 1 so we are taking the correct decision.

Similarly, lambda 2 2 this is the loss incurred for taking an action alpha 2 when the true state of nature omega 2. So, this is also a case when we are taking a correct decision so this is loss involved for taking correct decision. Whereas, this lambda 2 1 and lambda 1 2 these are the loss for taking wrong decisions because we taking action alpha 1, other two state of nature is actually omega 2. We are taking an action alpha 2 when the true state of nature is actually omega 1.

So, naturally this alpha 2 1 will be much larger than alpha 1 1 because the risk this is the loss for taking correct action and ideally this should be equal to zero because that action is correct. Similarly, that lambda 1 2 is much better than lambda 2 because this is also the loss incurred for taking wrong decision, whereas lambda 2 2 is the loss incurred for taking correct decision. So, you find that both of these that lambda 2 1 minus lambda 1 1 and lambda 1 2 minus lambda 2 2 both of them will be positive or greater than zero. Now I can compare this with the two cases or Bayes decision that you have done earlier or minimum error classification method that I have done earlier.

(Refer Slide Time: 29:04)

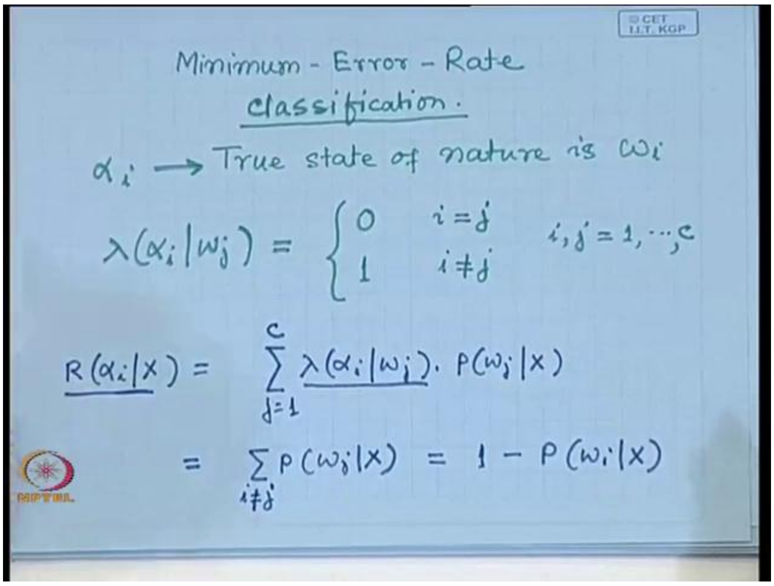


So, our condition was p of omega 1 given x if it is greater than p of omega 2 given x then we have decided in favour of class omega 1. Now, in case of this generalised one by incorporating the risk function that is the minimum risk classifier, what you have do is this, decision rule that we had taken in simple Bayes decision that actually has to be weighted by the loss difference. Because this lambda 2 1 minus lambda 1 1 is nothing but a loss difference for taking wrong decision and for taking a correct decision.

Similarly, lambda 1 2 minus one lambda 2 2 is also a loss preference between taking a wrong decision and taking a correct decision. So, the difference between a generalised case and the specific cases is that here, we had a simple expression p omega 1 given x minus greater than p omega 2 given x leads to the decision of omega 1. In the present term action alpha 1, if it is the reverse then I had to take action alpha 2 whereas, in this minimum classifier simply becomes lambda 2 1 minus lambda 1 1 which is waiting this posteriory probability p of omega 1 given x.

That is greater than lambda 1 2 minus lambda 2 2 into p of omega 2 given x. This actually initiates action alpha 1 or deciding in favour of omega 1. So, this is what we are getting following the minimum risk classification. Now, as I said that there are derivatives of this minimum risk classification. Under different situations I can have different types of classifiers which are actually derived out of this minimum risk classifier.

(Refer Slide Time: 31:45)



So, let us see one such classifier which is minimum error rate classifier. You have any questions so far.

Student: Sir, how will whether lambda 1 ((Refer Time: 32.16)).

Those are predefined the loss function are predefined. Depending on the kind of application that you have, you have to define, what is the amount of loss that you will impose for different wrong decision or for a correct decision?

Student: Sir, what means loss incurred for correct decision?

Ideally it is zero, if I take a correct decision then the loss is actually zero. So, ideally it should be zero. For more generalisations I can still put a loss function which may be very low. Because as we said in our previous class that, if any I take a decision omega 1, but the probabilistic point of view there is also a finite probability that the object may actually belongs to class omega 2. Though, that probability value is very small so ideally the loss involved for taking a correct decision is zero.

But to take care of such cases I can impose a loss even in a correct decision, but that loss value is very low. Because there is always a finite probability that my decision even though I am confident that I am taking a correct decision, but there is a finite probability however, small it is that my decision can be wrong.

So, that is they can take care of by that lambda i i or lambda j j. Even there is a situation where I may incur a loss, even if I am confident that I am taking a correct decision from other circumstances. So, what is this minimum error rate classification? So, we said that if I take an action alpha i that means I am taking a decision that the true state of nature is omega i that is alpha 1 is true state of nature omega 1 alpha 2 is true state of nature omega 2 and so on.

And if I define the loss function like this say lambda alpha i given omega j. So, taking an action alpha i means deciding it state of nature is omega i. So, as I said that if my decision is correct that means if I take action alpha i the true state of nature is also omega i then ideally I should incur a zero loss.

So, if I define this loss function like that, so I define this lambda alpha i given omega j equal to zero whenever, i is equal to j that means I am taking the correct decision. And the loss function I make equal to one whenever i is not equal to j. So, this is how I define my loss function. And this is true for all i and j equal to one to c for all different values of i equal to j. So, whenever i and j are same that means I have made correct decision, I am taking a correct decision, the loss involved is zero. Whenever, I take a wrong decision the loss involved is one.

So, this is how I define my loss function. So, by this definition of loss function let us see what will be the expected loss of the tricks that is R of alpha i given x that will be nothing but you have already said alpha i given omega j into p of omega j given x where j varies from 1 to c. So, this is the loss involved or the risk involved for taking an action alpha i. either equal to zero or equal to one. So, it is equal to zero whenever i is equal to j and it is equal to one whenever i is not equal to j.

So, this term the summation gets simplified to summation of p of omega j given x. However, I had i is equal to j this summation lambda alpha i given omega j is equal to one and wherever i was equal to j this was equal to 0. So, this summation will be wherever i will be equal to j not i equal to j.

Now, for this alpha i, one value of i only one value of j is equal to i, total submission of all these probability values p of omega i given x or p of omega j given x for all values of j is equal to one. Out of that this is the summation where I is not equal to j. So, this is nothing but one minus p of omega i upon x. So, if I want to maximise or minimise this risk function that means this term has been minimised. And if I want to minimise this then this p of omega i has to be maximised.

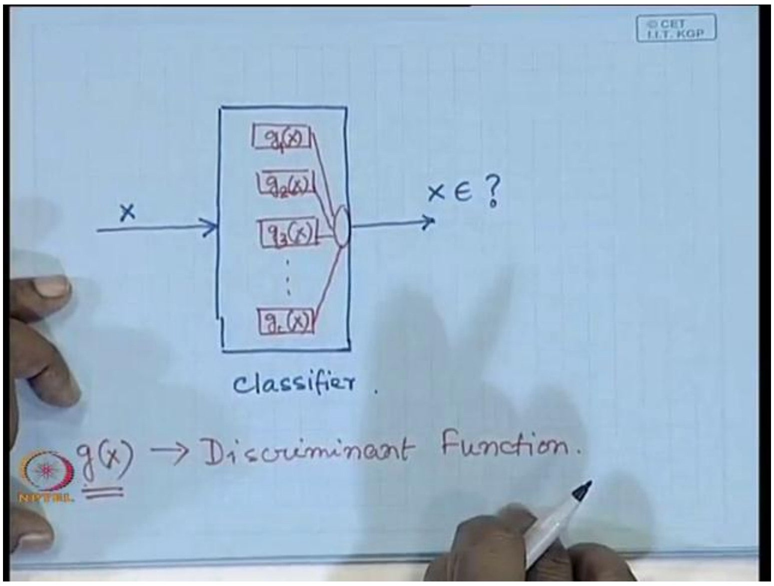
So, you find that I come back to the same decision, similar decision in the generalised case that p of omega i given x, for whichever i this has to be maximum I had to take that decision. And that ensures minimum risk that ensures minimum error affect. So, this minimum risk classifier in this particular case boils down to minimum error rate classifier.

So, that is why I said that starting from there, I can have various derivatives various types of classifiers, but finally, all of them will turn out to be equivalent. But under different situations I can use them differently depending of convenience how we can model a problem in a particular manner? Now, let us come to another concept which is called discriminant function. So, effectively what are the classifiers you have? Say it is something like this ((Refer Time 40.46)).

Probability is not zero, lambda alpha i omega is zero that is why the term is absent from here.

Here we are taking this probability p omega i given x, p omega i given x is not zero. But in this expression this p of omega i given x was to be multiplied by lambda i i. This lambda i i is zero that is why in this expression this p of omega i given x the term corresponding to this is absent. It is only because of this p of omega i is not zero. So, when I have this c class classifier I can put it something like this.

(Refer Slide Time: 41:54)



Say, I have a classifier let us assume that it is a black box. So, this is my classifier box. Input i have an observation vector or feature vector x of dimension d. Then the classifier has to give me a decision that what is the class belongingness of this object having this feature vector x. Now, inside this black box all these different types of calculations are to be done are to be meant. if I go for minimum risk classifier than for every different action this classifier has to find out that what is the corresponding risk. And it has to take give me that action for which the risk is minimum.

If I go for Bayes rule for every different class it has to find out what is the a posteriory probability then whichever class gives me the maximum posteriory probability the classifier will decide in favour of that class. In case of minimum error red classifier for every class the classifier has to decide that what the error for taking a particular action is. And it will decide in favour of that action which gives the minimum error. And you find that I have to compute the number of functions which is equal to number classes or which is equal to number of actions that I have in my classifier.

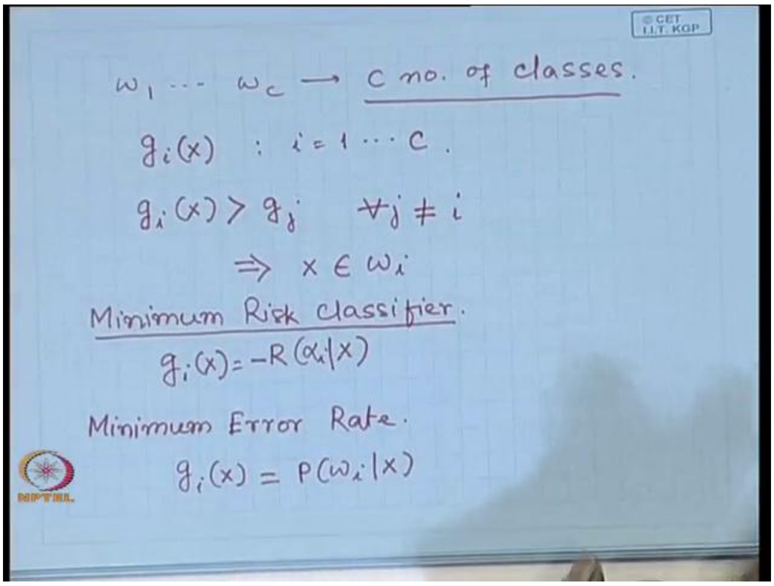
So, many functions are to be completed. Say for every action alpha i in my classifier I have to compute this for every classroom omega i. For every class this omega i I have to compute this p of omega i given x like that. So, it is the number of classes of the number of actions that are higher defined with my classifier where I have to complete. So many functions then whichever functionaries gives either maximum or minimum. I take that particular action of that particular class. So, I can say that in this classifier box I have different modules or functional units which compute a function g of x

So, this computes a function of g of x g 1 of x, this computes g 2 of x, this computes g 3 of x, this computes g c of x on the same feature vector x. Then you takes a decision either following the maximum criteria or following the minimum criteria that to which of these c number of passes this feature vector x should belong.

Now, if I put a generalisation that I will always compute the maximum of these functional values. So out these c number of functional values which ever functional value is maximum I will put x into the corresponding class. So, over there these functions are what are called is discriminate functions. So, this g x is called discriminant function.

So, your function is same I will call it g 1 x when it is computed for class one omega 1 or when it is computed for plus two omega 2 g two x is the same function when this g x is computed for class omega 2 or for action alpha 2 and so on. So, this function is called g is for discriminant function. And whichever class if so maximum value discriminant function I put this object into that corresponding class. Now, let us see what will be the nature of this discriminant class

(Refer Slide Time: 46:50)



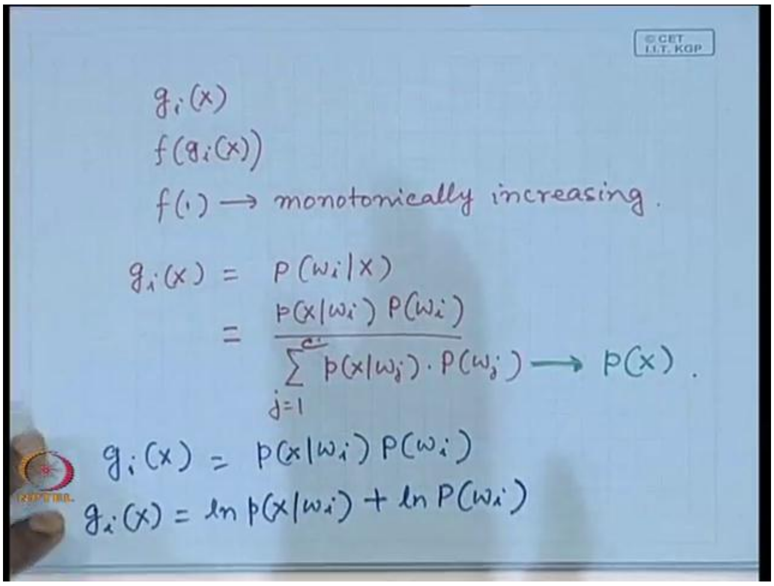
So, when I have c number of classes omega 1 to omega c. So, there is c number of classes. So, I will have c number of the functional value of the discriminant function, so g i x for i varying from one to c. So, if I feel the decision rule in general that g i x is greater than g j x for all j not equal to i. Then I decide that this x belongs to plus omega i.

So, this means that whichever g i x whichever class i gives the value of this discriminant function g i x maximum, because this g j x all j not equal to i means this one is maximum. So, for whichever class this discriminant function gives the maximum value I put x into that corresponding class. So, what will be the nature of this discriminant function under different conditions.

If I go for minimum risk classifier my risk is given by R of alpha i given x for taking and action alpha i. And the distance to be minimum to take that action alpha i, but in terms of discriminant g i x the value of the discriminant function has to be maximum. So, naturally if I want to relate this risk with the discriminant function I have to make g i x which is negative of R of alpha I given x, because whenever this is maximum this is minimum because it is negative. And whenever this R of alpha i given x is minimum by negating this g i because maximum.

Similarly, for minimum error red classification my condition was that this one minus p of omega i given x that has to be minimum that means p of omega i x has to be maximum. So, I can simply equate g i x top of omega i given x. So, for minimum error red classification g i x is simply p of omega i upon x. So, I can have the discriminant functions like this and for multiple number of classes for which ever class the value of the discriminant function is maximum I put x into that particular class. Now, you find that I can define g i f like this, but the choice of the discriminant function g i x is not equal the reason is.

(Refer Slide Time: 50:48)



If I take a function f say I have this g i x. And I take a function which is function of g i x. Now, if this function f which is function of g i x this is monotonically increasing. Then also this f of g i x will solve the same purpose as g i x, because it is monotonically increasing. So, for whichever i g i x is maximum for the same i if you g i x will also be maximum because the function f is monotonically increasing. So, if I can identify g i x then for any monotonically increasing function f of g i x that will also serve the same purpose of discriminant function. So this discriminant function that I said g i x it is not really unique. I can have various types of discriminant function.

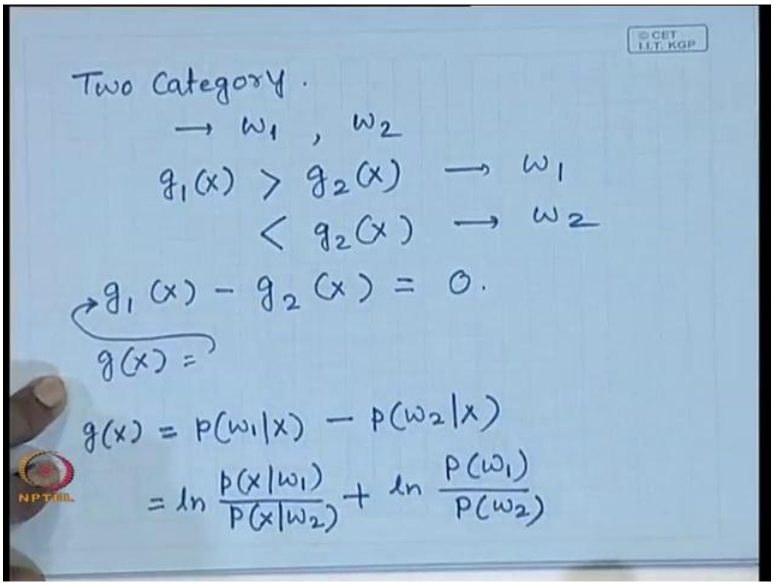
So, only here I have to take is this function f that I have to choose that must monotonically increasing function. And that gives us an advantage in the sense that if somehow I can identify g i x but g i x in its original form if it is not mathematically tractable. I can take another functional of this g i x which can be mathematically tractable, that can be used as a discriminant function.

So, coming to a very simple example. So, coming to this minimum error red classification, we have said that this g i x is nothing but p of omega i given x. If I expand this it simply becomes p of x given omega y into a priory probability p omega i upon summation of p x given omega j into p omega j for j varying from one to six. Now, find that this term p x given p omega j, this is nothing but p of x.

And because this term is appearing in the denominator of all the discriminant functions for every value of i, this will be there in the denominator. So, I can simply remove this when I design my discriminant function. So, I can say that my discriminant function will simply be g i x is equal to p of x given omega i into p of omega i. This will be now g i x.

Now, find that one I define g i x like this, there is a product term. And whenever I have a product it is more difficult to implement as well as analyse rather than if I have a submission. So, as I have my original formation g i x like this. We know that logarithmic function is also monotonically increasing function. So, instead of using this I can use log of this. And that can be my discriminant function. So, instead of using g i f as this, I can use g i x, as this can also be my discriminant function. And here I have avoided this product by summation. So, it becomes mathematically more convenient.

(Refer Slide Time: 55:29)



Now, using this there is, if I go for two categories that is I have the classes omega 1 and omega 2 that means that I have classes omega 1 and classes omega 2. So, when I have these two classes that means I have two discriminant functions. One is g x other one is g x. And my decision rule is if g 1 x is greater than g 2 x, I decide it in favour of omega 1, g 1 x is less then g 2 x I decide in favour of omega 2.

So, what is the decision boundary between the classes omega 1, omega 2? Decision boundary simply where g 1 is equal to g 2 x. So, g 1 x minus g 2 x is to zero, that gives me the decision boundary. So, if g 1 x minus g 2 x is greater than zero, I put it in plus omega 1, if it is less than zero I put is in plus omega 2. So, I can say that instead of taking these two discriminant functions particularly in a two category case, I can have a single discriminant function which is given by g x is equal to this g 2 x minus g 2 x and if this is equal to zero that gives me the decision boundary.

And from here by applying the same concept of logarithm you will find that this discriminating function can now be written as. And if I use g 1 x to be p of omega 1 given x and g 2 x to be p of omega 2 given x then g x becomes p of omega 1 minus p of omega 2 given x. And using the concept of logarithm and by expanding this in terms of a priory probability and class conditional probability, this will simply be written as ln p of x given omega 1 upon p of x given omega 2 plus ln p of omega 1 upon p of omega 2.

So, here I have this priory probabilities as well as class conditional probabilities. So, when a priory probability is same you find that this term become equal to zero. So, only decision is based on your class conditional probability. So, I will stop here today will continue with this discussion in the next class.

Thank you.