Applied Multivariate Analysis

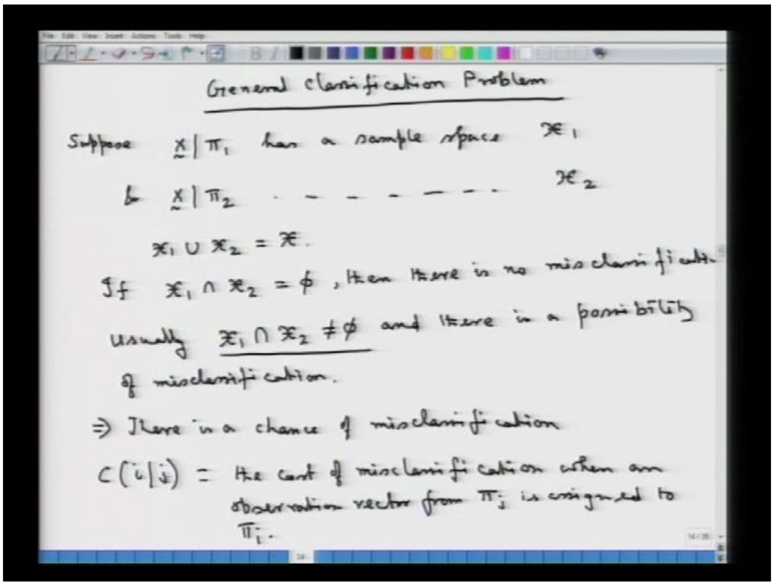
Prof. Amit Mitra
Prof. Sharmishtha Mitra
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture No. # 31

Discriminant Analysis and Classification

So, we have started discussing about the discrimination and classification problem. Let us discuss this in a general classification setup.

(Refer Slide Time: 00:23)



General classification problem as what we are going to look at. Now, we have the following set up that suppose X, the multivariate random vector is from by one population has sample space say script x 1, and X given the second population, we are still looking at two population problem. This has got a sample space similarly say script x 2. Now, x 1 union x 2, the union of the two sample spaces is the entire space of possible x vectors. Now, if we have the following that script x 1 intersection script x 2, if this is equal to a null set, then there is no problem actually, because there will not be misclassification. Then there is no misclassification.

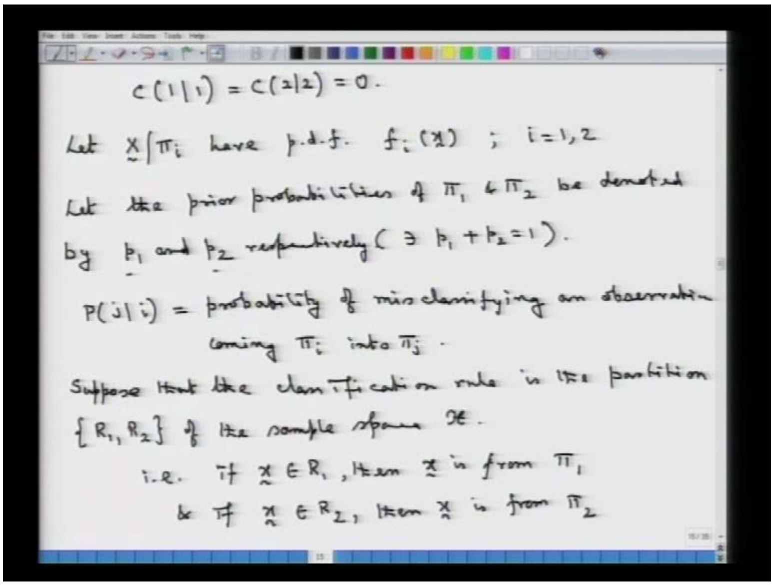
However, for all practical purposes what we usually observe is that this intersection region is not equal to phi. Usually, this script x x 1 intersection script x 2, this is not equal to phi, and there is a possibility of misclassification. (No audio from 02:11 to 02:22) Now, if the problem was that that the two sample spaces script x 1 and script x 2 under the two different populations pi 1 and pi 2, whatever be those multivariate populations. If that was a null set, then whenever we have a particular multivariate observation from this part of this sample space, then we will assign it to pi 1.

And if it is from x 2 space, then we will assign it to pi 2 and that is going to be full proof, because there is no chance of any misclassification looking at the region of its occurrence in the sample space. One can easily say that from which population it is coming. However, for almost all the real life applications it that is not the case and we will definitely be having. So, we have got for practical situations; script x 1 intersection script x 2 to be not equal to phi. Say for example, if pi 1 and pi 2 are two multivariate normal populations, then we have got suppose we have pi 1 population to be a multivariate normal with a mean vector equal to mu 1 and covariance matrix as sigma.

And pi 2 another multivariate normal populations with a mean vector as mu 2 and a covariance matrix as sigma. Then ofcourse, the sample space of the two populations are defiantly not disjoint. The the entire space actually is common and hence there is always a possibility when one is looking at classifying and observation coming from pi 2 or pi 1 to be going into the other way. So, there is possibility of misclassification. Now, since there is a possibility of misclassification, we will look at a cost which may be associated with such a misclassification. So, this implies there is a chance. If we have got x 1 intersection script x 2 to be not equal to phi, there is a chance of misclassification.

Let us now denote by the following quantity that C i j; this is the cost that one would be incurring for wrongly classifying and observation coming from population index by j into population index by i. So, this is the cost of misclassification the cost of misclassification, when an observation vector when an observation vector from pi j is assigned to pi i. Now, it is that we will assume along with this particular term. Now, this since this C i j is a cost of misclassifying an object coming from pi i pi j into pi i population. If we are looking at C i i; that is there is actually no cost of misclassification, because the object is coming from i in such a situation and is being classified through the classified in to the i th population.

(Refer Slide Time: 05:46)



And hence, we will be having C 1 1, that is its no misclassification; that would be equal to C 2 2, which is going to be equal to zero. Now, let us also have a density associated with such pi i populations. Suppose we have X given pi i to have a density, the joint density of x 1, x 2, x n to be denoted by f i x; this is for both the populations 1 and 2. And let also the prior probabilities for the two populations we defined. Let the prior probabilities prior probabilities of pi 1 and pi 2 be denoted by say p 1 and p 2 respectively. So, we have these to be the A priory probabilities for the two population, which are forming actually the possible population set. Now, these are such that we will have p 1 plus p 2; this equal to 1.

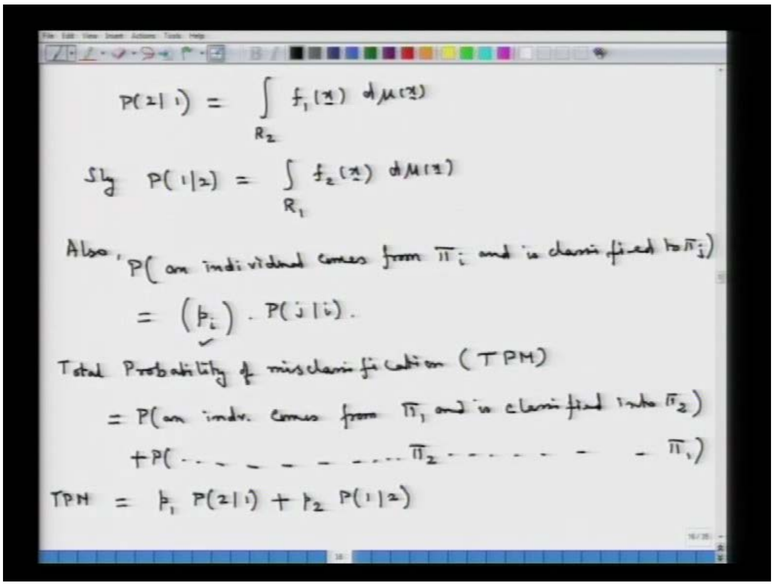
So, what we assume is that these two are the two populations, which make up the universe. So, a particular random vector is either coming from pi 1 or it is coming from pi 2. So, it is basically an exhaustive set. We cannot have a particular observation coming from anything other than this. This is a two class problem, that is why we have chosen that to be p 1 and p 2; the two A priory probabilities. If we have a general more general c c class problem pi 1, pi 2, pi c for example, then we will be having the corresponding A priory probabilities as p 1, p 2, p c such that summation of p i will be equal to 1. Now, along with these definitions, let us also define this quantity, which is P j given i.

So, this is going to denote the probability of misclassifying an observation probability of misclassifying an observation coming from pi i into pi j. So, the notation is very clear that one is looking at P j given i to be the probability of misclassifying an observation, which is coming from pi i into the population pi j. Now, suppose along with this; suppose that the classification rule is given by the following. Classification rule is the partition we had discussed, how we are actually looking at the classification rule as partitions of the sample space is the partition R 1 and R 2. So, this is the partition of the sample space, partition is the partition this of the sample space script x script x.

That is, when we say that this is the classification rule, which is the partition of the sample space script x. What we try to mean is that; if x belongs to R 1, then x is classified or supposed to have come from population, which is index by 1; that is pi 1. So, x is from pi 1. And if x is coming from x is belonging to R 2, then x is from pi 2; because we have made this as the partition this R 1, R 2. So, R 1 intersection R 2 is phi; R 1 union R 2 is script x. So, for any x coming from the sample space, we will see its location with respect to this partition R 1 and R 2. And if x belongs to this R 1 part, then we will say that x is basically coming from pi 1 and hence x is classified into pi 1 population.

If x is belonging to R 2, then x is classified to be coming from the second population, which is pi 2. So, this is the classification rule. So, what are the quantities that we have introduced? (Refer Slide Time: 00:23) We have introduced this cost of misclassification. We have introduced with with the condition that C 1 given 1 and C 2 given 2 both of them are 0's. We have the probability density function under the respective populations; pi i is to be given by f i x's. Their prior probabilities are p 1 and p 2 for the two populations. And P capital P j given i is the probability of misclassifying an observation coming from pi i into pi i j. And along with that we say that we have a classification rule R 1, R 2; which is a partition of the sample space script x and that is what is going to decide the classification problem.

(Refer Slide Time: 11:15)



Now, what is this quantity equal to in the light of what we have discussed? This is the probability of misclassifying an observation coming from the first population pi 1 into the population pi 2. Now, when are we when are we going to assign an observation to pi 2, if it belongs to the partition R 2? So, this expression is given by the integral over the region R 2. Now, the point actually the multidimensional point is coming from pi 1. So, it has got a density, which is f 1 (x) this with respect to the underlying measure d mu x. So, this is the probability of misclassifying an observation coming from pi 1 into pi 2.

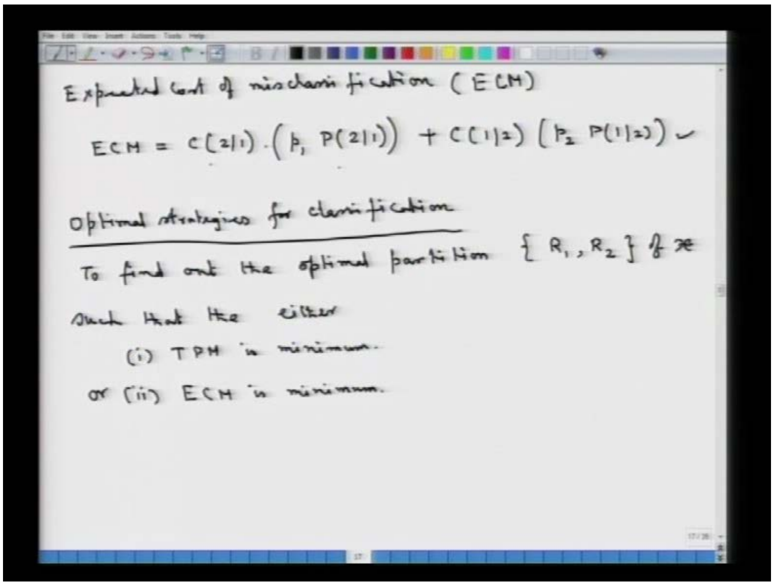
Similarly, if we look at P 1 given 2, this is integral over the R 1 region; because we are classifying it into pi 1 and it is coming from pi 2. So, it has got a density under pi 2 as f 2 (x) ; this with respect to the underlying measure say mu. Also, if we look at this probability, probability that an individual probability that an individual comes from pi i and is classified to pi j; this is, note that this is different from P i given j. What is this? This probability can be; so, this basically is a joint event that an individual is coming from pi i and then it is getting classified into pi j. So, its probability that an individual first comes from pi i into the probability that individual is classified into pi j given; it has already come from pi i.

So, what is this going to be equal to? This is going to be equal to probability that an individual is coming from the first from the pi i population. The probability of which is this; this into the probability that an individual from pi i is classified or misclassified into pi j. So, this is P i I am sorry this is classified into j. So, this is P j given i. So, this is what we have. So, one can actually look at based on these definitions, an important quantity which is called total probability of misclassification, total probability of misclassification in short the TPM. For this classification rule, total probability of misclassification for this classification rule that is the partition R 1, R 2 here.

So, this TPM is going to be given by the some of the probabilities, that an individual is coming from pi 1 getting misclassified into pi 2 plus an individual is coming from pi 2 and is getting misclassified into pi 1. So, this is probability that an individual comes from pi 1 and is classified into pi 2 and is classified into pi 2. So, the total probability of misclassification is given by the some of the two probabilities that an individual comes from pi 1 and is classified or misclassified in to pi 2 plus the probability that an individual is coming from pi 2, the other population and is getting classified into pi 1. So that, this in our previous notation is equal to…Now, this would be individual is coming from pi 1.

So, the probability of that A priory probability is p 1 and then it is getting misclassified into the second population. So, this is the first part of the probability statement; the total probability of misclassification plus it is coming from the second population, which has got a prior probability equal to p 2 and it is getting misclassified into the first population. So, this is our total probability of misclassification, which is of importance. Because when we are looking at finding out optimal classification rule, one looks at one partition R 1, R 2 of the sample space, which we would actually lead us to one that minimizes this total probability of misclassification. Now, here note that when we looking at the total call probability of misclassification, one is not looking at the costs of misclassification. So, one can also look at a quantity, which is called the expected cost of misclassification.

(Refer Slide Time: 16:20)



So, the expected cost of misclassification expected cost of misclassification or ECM. What is this ECM? This is the expected cost of misclassification. Now, the misclassification cost would be given by C 2 1; that multiplied by the probability that it the individual is coming from the first population and getting misclassified into the second population. So, that this is equal to p 1 times probability of misclassifying an observation coming from 1 into 2. So, this is the cost with this probability and then the cost of misclassification is C 1 given 2, if an individual from second is misclassified into 1.

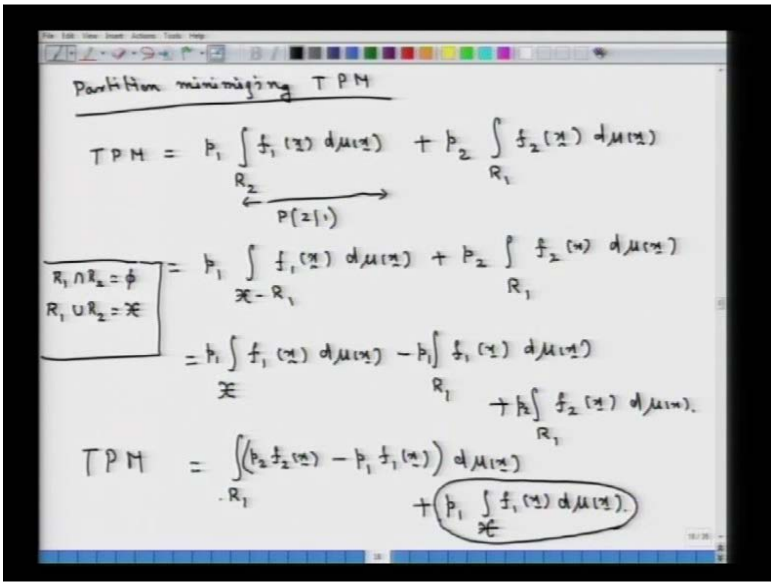
So, that that probability would be given by p 2, which is the prior probability probability of the second population; that multiplied by this P 1 given 2. So, we have two important quantities. One is this expected cost of misclassification (Refer Slide Time: 11:15) and the other is this total probability of misclassification. Now, let us look at optimal strategies optimal strategies for classification rule. Now, this general discussion on finding out the optimal strategies for misclassification or rather the classification rule optimal strategies for classifying it correctly. What one is assuming is pi 1 and pi 2 to be these two populations with the densities given by f 1 (x) and f 2 (x).

So, we are not putting any special structure on those densities. We are looking at a general setup. So, when we talk about optimal strategies for classification, what we are trying to do is to find out the partition or rather the optimal partition optimal partition say

R 1, R 2 of the sample space script x such that the criteriance that we have defined. That is the total probability of misclassification or the expected cost of misclassification. One of them is going to be minimum possible, because the first one when we look at this is the total probability of misclassification.

So, lower the better and ofcourse, when we are also looking at expected cost of misclassification, lower would be the better. So that, we are looking at this optimal partition R 1, R 2 of x such that the total cost… let me write it one by one; such that say we have got either of the two approaches. Either this TPM, the total probability of misclassification is minimum or we can look at an alternate approach and say that know our objective would be to minimize the expected cost of misclassification. So, under anyone of these we will try to find out, what is that partition R 1, R 2; which is going to lead us to the minimum value of the respective quantities.

(Refer Slide Time: 20:04)



Now, let us first look at the optimum partition, which is going to minimize partition minimizing the total probability of misclassification. Let us now derive this particular total probability of misclassification minimizing partition. Now, remember what we had the total probability of misclassification that given by p 1 into this quantity, which is integral over R 2 f 1 (x) d mu x. So, this quantity is nothing but probability that an individual is wrongly classified into the second population given that it is coming from the first population; this multiplied by the A priory probability. So, this plus p 2 times P

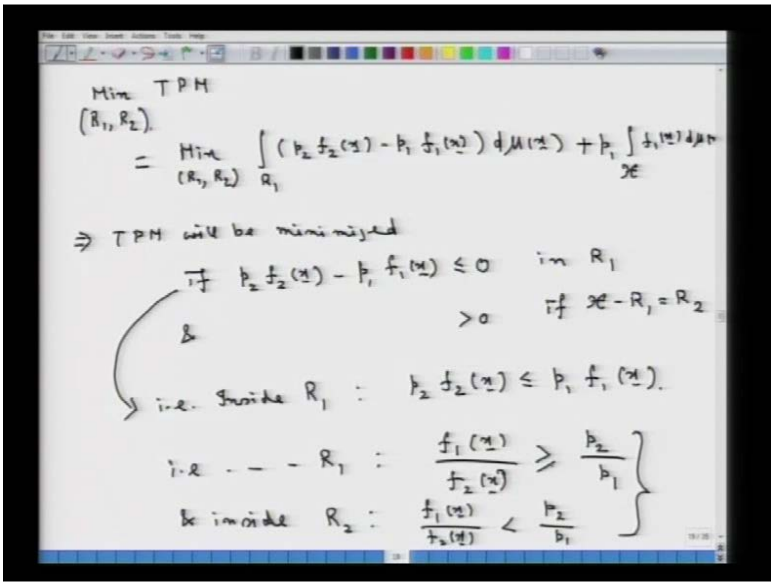
1 given 2. So, that this is now going to be given by integral over the region R 1; it is coming from the second population.

So, it has got a density f 2 (x) ; this this with respect to the underlying measure. Now, let us look at the following that this R 2 region can be written as script x minus R 1, because we have R 1 intersection R 2. So, we remember that R 1, R 2 is a partition. So, R 1 intersection R 2, this is a null set and R 1 union R 2, this is the entire sample space script x. So, if we have that, then this region R 2 is the sample space script x minus R 1 region of this quantity. We do not disturb that at all and leave this second quantity also as it is; this is integral over R 1 f 2 (x) d mu x. Let us now look at what this is equal to? Now, this would be equal to the integral over the script x region minus the integral over this R 1 region.

So, what we have is this integral over x. So, this p 1 constant remains as it is; this is an f 1 (x) d mu x; this minus integral over R 1 f 1 (x) ; this is a p 1 also this constant here; f 1 (x) d mu x plus p 2 times integral over R 1 of f 2 (x) d mu x. Now, let us write it in the way that this is integral over R 1. Let us look at this third term first. So, this is p 2 times f 2 (x), this minus this is integral over the same region R 1 with a minus sign; this integrant from here is p 1 f 1 (x) this times d mu x. And then we have the first term here, which is p 1 times integral over the region script x, the sample space f 1 (x) d mu x. Now, when we look at this particular term here; the third term, which is this quantity.

Now, our objective is to look at this total probability of misclassification, which is given by this. And to find out that partition R 1, R 2 say R 1 star, R 2 star such that this total probability of misclassification is minimum. So, we are looking at the minimum value of this quantity with respect to the partition R 1, R 2. Now, note that this quantity, the one which I have circled. So, this term is independent of the partition, because this is integral over the sample space script x. So, this is independent of this R 1, R 2 partition. And hence, when we are trying to find out the minimum of the total probability of misclassification, we can just look at what is that R 1and hence R 2, which is going to minimize this particular first term.

(Refer Slide Time: 24:32)

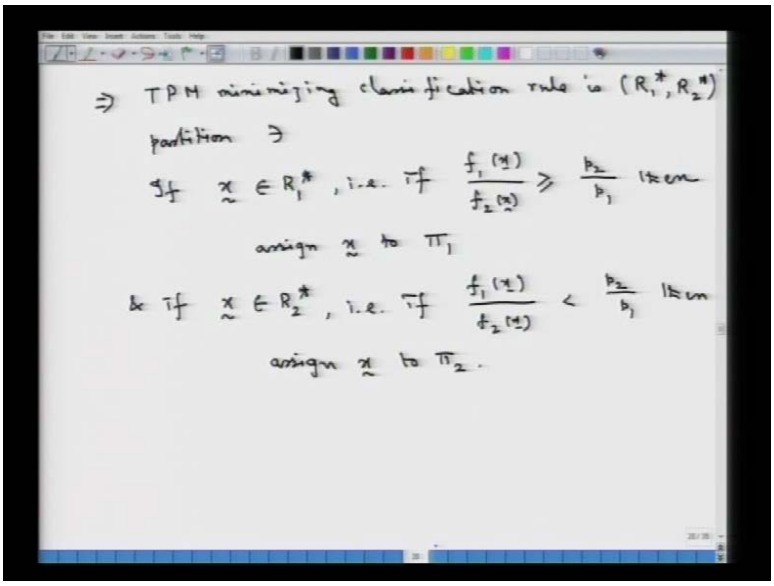


So, we will have minimum total probability of misclassification with respect to the partition R 1, R 2 is what we have looking at. And hence, we are going to look at what is the minimum with respect to the partition R 1, R 2 of the first quantity (Refer Slide Time: 20:04) which is here; which is integral over R 1. Then we have here p 2 f 2 (x) p 2 f 2 (x) this minus p 1 f 1 (x) this into d mu x. So, the minimum of this is minimum of that plus this constant term, (Refer Slide Time: 20:04) which is the last term here; which is p 1 times this p 1 times integral (Refer Slide Time: 20:04) over script x f 1 d mu x f 1 d mu x. So, that we are trying to find out what is that R 1, which minimizes this and that partition would lead us to the partition, which would minimize total probability of misclassification.

That easy to see that when we have over the region R 1; if this integrant is negative, then this quantity is going to be minimized. So, this would imply that the total probability of misclassification will be minimized. This is going to minimized, if we have p 2 f 2 (x) this minus p 1 f 1 (x) to be less than or equal to 0 in the region R 1. And this is greater than zero, if it belongs to script x minus R 1; that is in the region there which is R 2. Now, from here we can say that inside R 1, what we require in order to minimize the total probability of misclassification is that inside R 1, we will have this p 2 f 2 (x) to be less than or equal to p 1 f 1 (x) . That is inside this is region R 1, we will have the quantity that it is f 1 (x) by f 2 (x).

This term is going to be; so, f 1 (x) divided by f 2 (x) that would be greater than or equal to p 1 p 2 by p 1. And inside R 2, we will have this f 1 (x) by f 2 (x) ; this ratio of the two densities to be strictly less than this p 2 by p 1. So, when we have this particular partition that given a particular x 1 would be computing this ratio. And if that is greater than or equal to p 2 by p 1, then that x is going to be assigned to R 1 and if it is otherwise that if this ratio is less than p 2 by p 1, then the corresponding x is going to be assigned to R 2. So, this is the partition which is going to lead us to the optimum partition, which is going to minimize the total probability of misclassification.

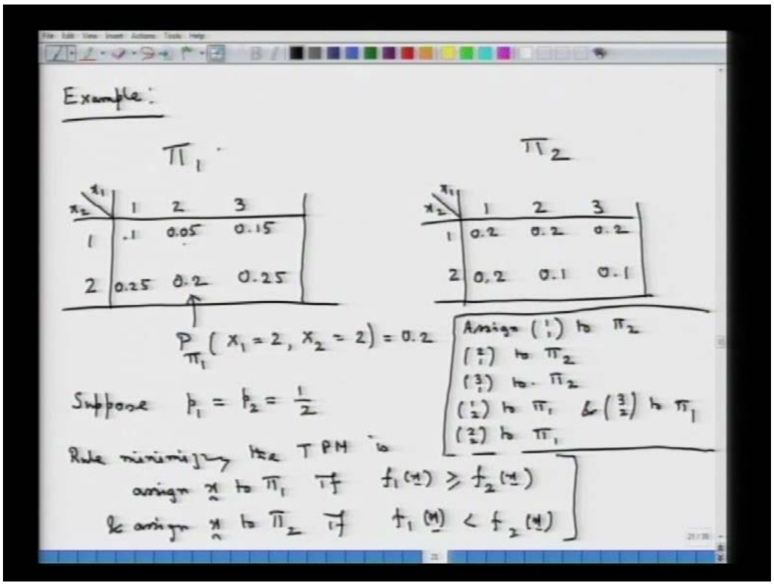
(Refer Slide Time: 28:07)



So, this would imply that TPM minimizing classification rule. TPM minimizing classification rule is say I write it as R 1 star, R 2 star partition R 1 star and R 2 star partition such that we have in R 1 star. So, I am just denoting this (Refer Slide Time: 24:32) optimum partition R 1 and R 2 that we have derived in terms of writing it as R 1 star, R 2 star with in order to just show that it is basically that optimum partition in or inside R 1 star. Well one can also put it in a different way that if x belong to R 1 star; that is, if f 1 (x) by f 2 (x) this is greater than or equal to p 2 by p 1, then assign x 2 the population number 1. Because that is belonging to R 1 star region and R 1 star region is a partition, which is corresponding to the first population that is pi 1.

And if x belongs to R 2 star, which is the complimentary region of R 1 star; that is, if for a particular x we observe that f 1 (x) by f 2 (x), this is strictly less than p 2 by p 1; then assign that particular x to the second population; that is pi 2. So, this is a classification rule, which minimizes the total probability of misclassification. Before we proceed further and to look at the optimum classification rule, which minimizes expected cost of misclassification. Let us now look at a small example of a bivariate discrete to population problem and see how this particular rule the optimum rule, which minimizes the total probability of misclassification work.

(Refer Slide Time: 30:38)



So, let us look at that example. We look at the following example that we have two populations pi 1 and pi 2. So, let me first write the joint distribution corresponding to this pi 1 population. So, it is a bivariate discrete distribution what we have. So, these are the two possible variables x 1 and x 2; x 1 is taking the one of the values from 1, 2 and 3 and similarly x 2 is taking values one of the two values and 1 and 2. And then the probability table corresponding to this bivariate setup is that probability that x 1, x 2 is taking value 1 1; that is, this pair is taking value 1 1. When it is coming from the first population, this probability is 0.1.

Similarly, this probability is using 0.05; this probability is using a 0.15; this is 0.25 say; this is 0.2 and this is 0.25. So, that this have got the following interpretation that this number here. Probability that X 1 random variable is taking the value 2 and X 2 random variable is taking the value 2; this under the pi 1 population. This probability is equal to 0.2. So, it is a standard bivariate discrete distribution setup. So, these are the cell probabilities of these combinations. Now, similarly suppose we have a second population pi 2; for pi 2 population, we have the same variables. So, both are bivariate population. If one is bivariant and other is any other dimension, then there is no problem of any misclassification as such.

So, once again I have the same setup. So, in population number 2 that is pi 2; we once again have 6 cells. So, to say for the six combinations that is possible and the corresponding cell probabilities are 0.2, 0.2, 0.2 for these three; 0.2 also for this one, 0.1 and 0.1. So, we have a discrete population now, which is a special case of the general setup discussion that we had. So, these are the corresponding populations, pi 1 and pi 2. Now, we are going to implement that particular optimum rule, which minimizes the total probability of misclassification. Now, suppose for simplicity, we assume the prior probabilities to be same.

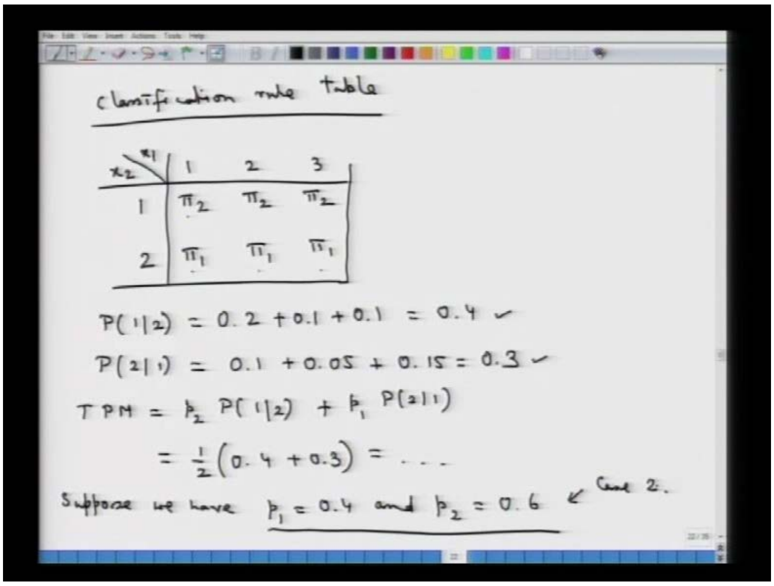
Suppose this p 1 is equal to p 2; that is equal to half, because p 1 plus p 2 is equal to 1. So, suppose we have p 1 and p 2 equal to half. So, that the two populations are equiprobable; that is, A priory probabilities that we had defined in the general setup. They are now equal to 1. Now, in such a situation we will have the rule minimizing the total probability of misclassification is assign let me write it in the next line is to assign x to pi 1 if x belongs to the R 1 region in the previous notation. (Refer Slide Time: 28:07) So, if x is belonging to this R 1 star, that is if f 2 (x) by f 1 (x) by f 2 (x) is greater than or equal to p 2 by p 1.

In the given example what we have p 1 and p 2 to be the same. And hence, the total probability of misclassification minimizing rule in the present setup would be to assign x to pi 1, if f 1 (x) is greater than or equal to f 2 (x). And assign x to pi 2, if it is otherwise; if we have f 1 (x) to be less than this f 2 (x). Now, if this is the total probability of misclassification classifying rule, let us see how to apply it here. It is simple actually; because these are the quantities for the possible six pairs 1 1 1 2 1 1, 2 1, 3 1, 1 2, 2 2 and 3 2. So, these are the possible cells for the two populations and the corresponding probabilities are basically this is the f 1 table so to say and this is the f 2 table so to say.

So, we will look at these pairs 1 1, 1 2, 2 1, 3 1, 3 2 things like that and then we will see which of this is higher; because we have the total probability of misclassification classifying rule to be given by f 1 (x) greater than or equal to f 2. And hence looking at this table, we can frame what is classification rule. Let me write it here; because the table is setting here. So, we will have the following rule. So, the classification problem based on this TPM rule would be assign this pair 1 1; 1 1 is a possibility; assign 1 1. If 1 1 comes to where? Now, we will have to look at whether this is higher than this; because we have got x to be assign to pi 1, if f 1 (x) is greater than or equal to f 2 (x) and this is to be assigned to the other population, if it is otherwise.

So, for this 1 1 combination, we see that this is the f 1 (x) and this is f 2 (x). So, this is to be assigned to pi 2. Similarly, if we look at say 2 1; let us see where 2 1 goes; 2 1has to be assigned. This is the 2 1 pair in f 1; this is 0.5 and this is 0.2 here. So, this 2 1 is assigned to pi 2 also; then 3 1 is another pair; where does it go? This f 1 is 0.15 and this f 2 is 0.2; hence we go into this region. So, this goes to pi 2 once again. Now, if we look at 1 2 1 2 combination, where does this go? this is 0.25 greater than this f 2 (x) and hence this is to be assigned to pi 1. Similarly, if we have this 2 2, where does this go? this 2 2 is 0.2 here this is 0.1 here. So, this goes to pi 1 here and the last pair, which is 3 2; this 3 2 goes to the first population. So, this is the classification rule.

(Refer Slide Time: 37:52)



We can write this in the classification table. Classification rule table is the following that we have this bivariate population set up that the two variables are x 1 and x 2. We have the three values 1, 2, 3; we have two values for x 2. Now if this pair is observed, we have a pi 2 assignment. If this value is observed, we have a pi 2 assignment; this is pi 2 assignment; this is pi 1, pi 1 and pi 1 assignment. So, this is what is the classification rule that has been framed. Now, if a new observation comes looking at what value it is taking on the two random variables x 1 and x 2, we will assign it according to this particular table. Now, given this particular table here; note that there is always a possibility of misclassification.

If an observation has come or rather observed to be 1 2, we are going to assign it to pi 1. However, (Refer Slide Time: 30:38) there is a positive probability that 1, 2 thus occur in the second population. So, there are chances of misclassification and which are given by this; this P 1 given 2, this is what? This is a probability of an observation coming from the second population to be classified in to the first population. So, when I am classifying it into the first population under these pairs. So, under the three pairs here, if an observation really is coming from the second population and we are by mistake putting it into the second population.

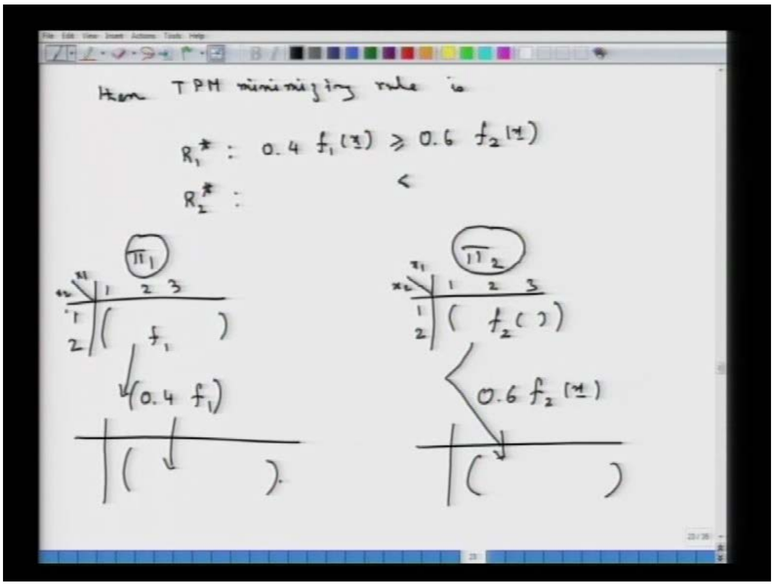
We are going to be penalized and that is a probability of misclassification from the table. We will have to look at, what is this cell probability, what is this cell probability and this cell probability under pi 2? (Refer Slide Time: 30:38) Let us see that the cell probability of this under pi 2; because we are thinking that the observation has really come from pi 2 and we are by mistake misclassifying it into pi 1 population. So, that the values from the previous table in the pi 2 population (Refer Slide Time: 30:38) is 0.4, 0.1, 0.1; 0.2, 0.1 and 0.1. So, that this is equal to 0.4. So, this is the probability of misclassifying an observation coming from the second population into the first population.

Similarly, we can look at P 2 1, which is going to be given by this is the probability that an observation is really coming from 1 and and it has been put into the second population. So, we are going to look at this. We have classified 1 1 into 2; but what is the probability that it is (Refer Slide Time: 30:38) coming from the first population this. So, we will have to look at these three values; because we can have an observation coming from the first population misclassified into two, only if it is under these 3 pairs; that is either 1 1, 2 1, or 3 1. So, that the total the probability of this particular event of one coming from one and getting put in to the second population (Refer Slide Time: 30:38) would be just the some of these three quantities.

So, that what we now have is 0.1 plus 0.05 this plus 0.15. So, that this is equal to this. Now, given that we have obtained for this TPM rule that this P 1 2 is this and P 2 1 is this. The total probability of misclassification, remember that this total probability of misclassification is the minimum possible that one is looking for; because we have obtained the rule this rule classification rule, which was actually minimizing the total probability of misclassification classification. And hence, this minimum total probability of misclassification would be p 2 times P 1 given 2 this plus p 1 times P 2 given 1. Now, we have p 1 p 2 to be equal to half.

So, that both of them are equal. So, this half goes outside and now what we have is 1 given 2 is 0.4 and this is 0.3. So, that what we finally have is this particular quantity; whatever it comes; so, that this term would be given by this. Let us now look at also a different situation. This is different problem; in the first case, we had assumed here (Refer Slide Time: 30:38) that this is the case. So, suppose I name this as case 1; this can be different. Suppose we have different A priory probabilities, the different A priory probabilities now are given by this, say p 1 is given by 0.4 and p 2 is equal to 0.6. So, this A priory probabilities say is corresponding to case 2.

(Refer Slide Time: 43:10)



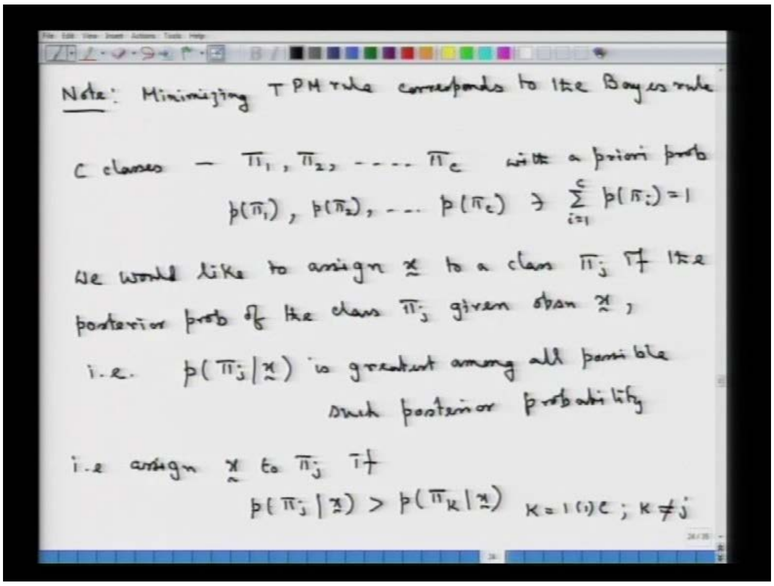
Now, if we have this particular case, then the TPM minimizing rule would be given by the following. The TPM minimizing rule is given by say R 1 star region. In R 1 star region, we will have p 1 times f 1 (x) ; this is greater than or equal to 0.6, which is p 2 times f 2 (x) and in R 2 star, the complementary region this is less than this particular quantity. Now from the given table, we had a pi 1 table, which had these cell probabilities and we had a pi 2 table with once again those cell probabilities. If I remember correctly, this is 1, 2, 3 and 1, 2 here corresponding to the two possible variables x 1, x 2.

The same structure is here x 1, x 2; 1, 2, 3 and 1, 2 here. So, we had those cell probabilities. Now, these are to be treated as f 1 quantities; f 1 for the pair that one chooses and this is the f 2 quantity of these pairs, which is the joint probability statement for x 1 random variable taking a value from here and x 2 random variable taking a value from here. Now, from here if this the f 1, f 2 table, what we will obtain is 0.4 times f 1 table, which would be derived from this particular f 1 table. So, each of the cell probabilities would now be multiplied by 0.04 and we will have this entry is coming here.

And similarly, here what we require is 0.6 times f 2 quantities and then we will have the corresponding table, because f 2 values are the values that we had previously. (Refer Slide Time: 30:38) So, these are f 2 values. So, we have multiplying each of them by 0.6 and we are multiplying each of these by 0.4, because those are the A priory probabilities. So, one can get to this using this f 2 values. Now, as in the previous situation, when we were comparing the cell probabilities in the two populations, this is corresponding to the pi 1 population; this is corresponding to the pi 2 population. So, we will look at whichever cell has got this quantity, which is this quantity here greater than the quantity in pi 2.

If we have for some pairs this to be satisfied, then for those pairs the assignment rule would be pi 1 and for the pairs for which, these values are less than these values; that is this region. We will have the assignment rule to take it to pi 2. So, it is straight forward. This is for a discrete bivariate distribution setup. One can also have similar type of assignment rules, when we consider multivariate distributions; like for example, multivariate normal distributions. We will see those later. But before proceed further, let us look at the relationship between this total probability of misclassification optimizing rule and it Bayes classifier. What is the relationship?

(Refer Slide Time: 46:22)

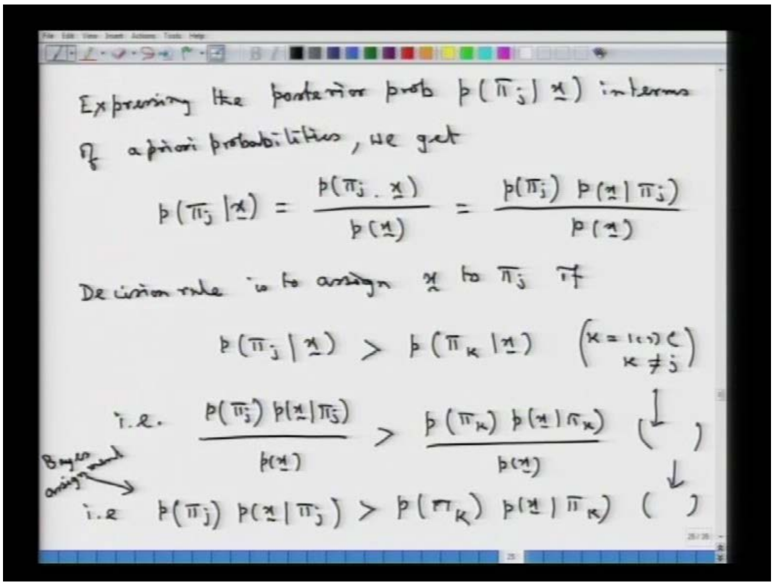


Let us look at this; the relationship between this TPM minimizing rule and the Bayes classifier. I write that this minimizing TPM rule corresponds to the Bayes rule. Now, how is that true? Let us look at a more general case. Suppose we have C classes, say we have pi 1, pi 2, pi c these are the C classes with A priory probabilities with A priory probabilities given by say p pi 1. One can denote that p 1 simply; p pi 1, p pi 2 and p pi c. Now, these quantities are such that some of all these quantities these are such that summation p pi i; this is going to equal to1, for i equal to 1 to up to c; because we are assuming that at the most there are c such populations. Now, we would like to assign… we are looking at the Bayes counterpart.

We would like to assign an observation multivariate observation to a class or a population pi j, if the posterior probability if the posterior probability of the class pi j given x; because we are looking at the posterior probability we would be looking at the posterior probability of the class pi j given x given observation vector x. That is, the posterior probability; let us denote that by pi j given x is greatest among all such posterior probabilities is greatest among all possible such posterior probabilities. So, that would be the principle under which, we are going to frame the Bayes rule. If the posterior probability of a class pi j; posterior probability given this observation x is the maximum among all other posterior probabilities for the remaining of the C minus 1 classes.

Then that observation under the Bayes rule is assign to pi j; that is, assign x to pi j, if we have got the posterior probability for the j th class. This is greater than the posterior probability for the remaining C minus 1 classes. So, that this k is from 1 to up to c; however, this k is not equal to j. So, we are looking at that particular posterior probability and we are going to assign it; assign x to pi j, if this happens. So, whichever wins actually in the posterior probability sense, x is going to be assigned to that particular population. Now, let us write these posterior probabilities in terms of A priory probabilities.

(Refer Slide Time: 50:19)

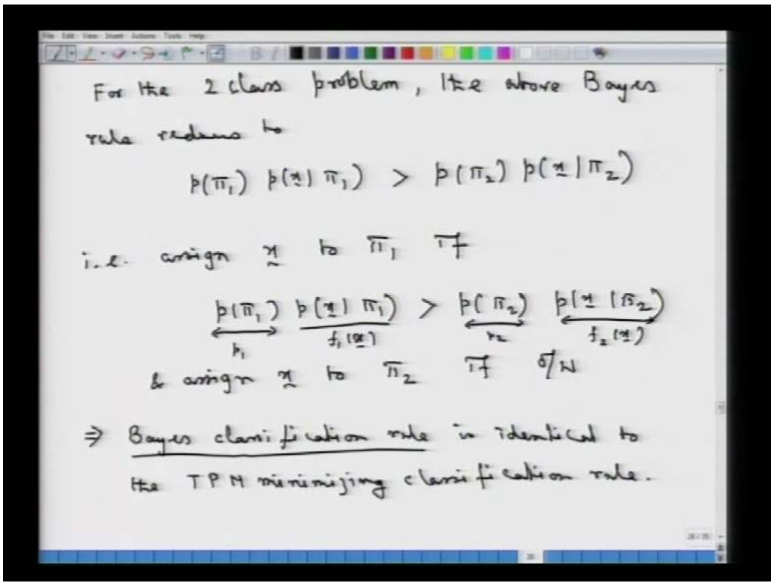


Expressing the posterior probabilities posterior posterior probabilities p pi j given x in terms of A priory probabilities, we get the following. We have this p pi j given x, which is the posterior probability. So, one can write that as p pi j intersection with that x. So, that this is the joint distribution and then we will have that divided by p x. And then, one can write this as probability of the pi j population and the probability that we have x given pi j. So, this would corresponding to the density in such a situation under the population pi j that divided by this.

Now, decision rule what we are getting? The decision rule is to assign x to pi j, if the following thing happens. That is, the posterior probability p j given x is greater than p pi k given x with that k from 1 to up to c and k is not equal to j. That is, note that in the posterior probabilities on the two sides; left side and the right side, we will have the same term p x here common. So, we can one may be write just one step here; p x given pi j this divided by p x; that is a posterior probability is greater than p pi k into p x given pi k that divided by p x; this for the same set of k values.

So, this is a nonzero term what we have here; that is the assignment rule the Bayes assignment rule; Bayes classifier is what is going to have. It is going to be based on this; p pi k that into p x, the density under the k th population with the restriction of k as what we had here only. So, this is what the Bayes rule actually falls down to. Now, for the given two class problem; so, this basically is the Bayes rule assignment. So, this is what you can say is finally the Bayes assignment rule or Bayes classification rule. (Refer Slide Time: 46:22) Now, for the given two class problem what we were looking at?

(Refer Slide Time: 53:24)



For the two class problem, this reduces to… the above Bayes rule reduces to the following; above Bayes rule reduces to what? We will have the two classes. So, that we will have p pi 1 into p x given pi 1 this term; if we have this greater than… look at what we had here (Refer Slide Time: 50:19) for the general case p, we are assigning x to pi j. If we have p pi j, this is the prior probability; this multiplied by the density of x under pi j. So that, for the two class problem, we have these two, pi 1 and pi 2 populations; this is this quantity and then this is x given pi 2. That is, assign now x to pi 1, if this happens; if p pi 1 2 p x given pi 1, that is greater than p x given pi 2 into p x pi 2. So that, that is the density and assign x to pi 2, if it is otherwise to pi 2 if it is otherwise.

Now, this in the previous notation what we had the prior probabilities, we are denoted them by p 1; this as p 2 and this was denoted; this is the density under pi 1. So, this was denoted by f 1 (x) and this was in the previous notation, when we were looking at deriving the TPM minimizing rule is f 2 (x). So, what we have is that this; it is the Bayes classification rule for the two class problem is precisely giving us the rule, which we had obtained; which was minimizing total probability of misclassification. So, that once again what we come back to is look at the rule, that that is what we had. This is the total probability misclassification rule. It is basically the region, wherein we have got this. For a continuous distribution, this equality does not matter actually.

So, we will be having that equality with probability zero for discrete distribution. (Refer Slide Time: 46:22) Some cases can have that ratio to be equal and hence what we have is the Bayes rule the Bayes rule or Bayes classification rule is identical to the total probability of misclassification minimizing classification rule total probability of misclassification minimizing classification rule. So, this this angle of looking at this total probability of misclassification minimizing rule being same as that of Bayes classification rule gives also a theoretical justification for such a rule. We will stop at this point and continue this concept of looking at other type of optimal strategies in the next lecture. Thank you.