Applied Multivariate Analysis

Prof. Amit Mitra
Prof. Shramishtha Mitra
Department of Mathematics and Statistics

Indian Institute of Technology, Kanpur

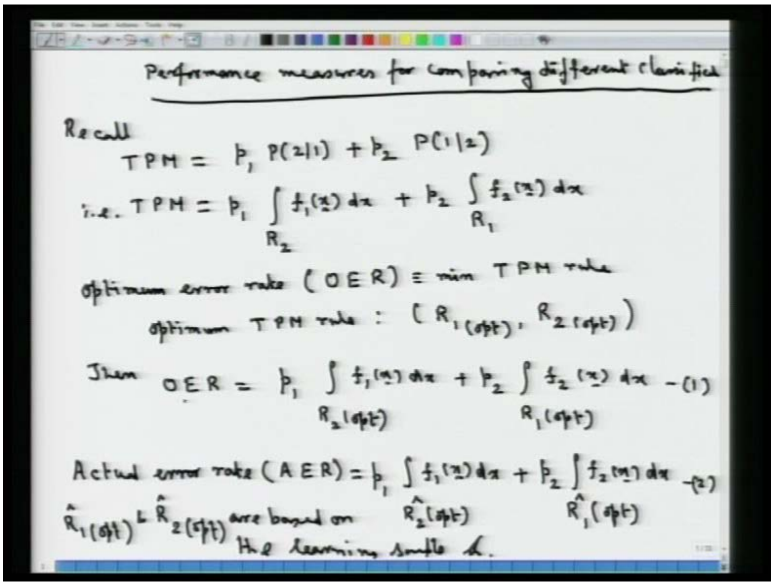
Lecture No. #33

Discriminant Analysis and Classification

So, in the last couple of lectures, what we were looking at was we try to derive classification rules, classification functions; that is partition of the sample space for two populations problem. We had looked at various concepts like what happens to when we look at a deriving a classification rule, when we have got a say for example, a rule which tries to minimize total probability of misclassification. What type of rule to be get when we look at a functions like expected cost of misclassification, and the partition corresponding to one that would minimize an expected class of cost misclassification. We had also seen in the last lecture for specific type of populations namely, normal populations, and we had seen how these optimum rules actually look like for these special type of cases of multivariate normal populations. So, first thing, that we are going to look at today will be performance measures of various classification functions.

So, that gives us some way of comparing various partitions, various classification functions, and then we will move on to looking at a classification problem in a multi population problem. So, we will generalize, whatever we have been drink for two population cases, we will look at generalizing that general c population problem.

(Refer Slide Time: 01:38)



Let us start today, looking at this performance measures thus, we had looking at performance measures for comparing different classifications, we looking at comparing comparing such different classification functions. Let us recall, what we had define earlier, we had a total probability of misclassification, where we are considering still two population problem. So, that was given by p 1. So, it is coming from first population getting misclassified into the second population this class p 2 times p 1 given 2.

Now this, in terms of the regions that we have a obtain. So, this is integration over the R 2 region, the object is coming from the first population. So, it has got density like this, this plus p 2 into integral over R 1. And since it is coming from the second population, it has got a density f 2 this term.

Now, based on this total probability of misclassification, we can look at the following measures, which is called the optimum error rate or O E R. Now, that is what is corresponding to the total probability of misclassification rule, which minimizes that total probability of misclassification. So, this is what is corresponding to the minimum T P M rule.

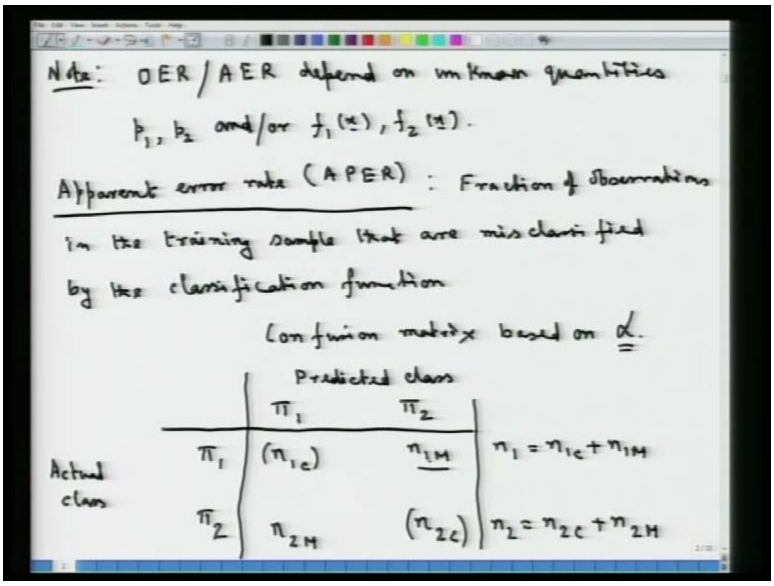
Now, suppose we have got this T P M minimum optimum T P M rule, suppose that is given by this partition R 1 optimum R 2 optimum. So, this gives us the minimum total probability of misclassification, then this O P R expression is given by O E R rather is given by p 1 into integral over R 2 opt, this into f 1 x d x, this plus p 2 times integral over this R 1 opt region of the density for the second population f 2 x dx.

So, this is one such measure for classification, you look at the optimum partitions that you obtain then calculate what is the total probability of misclassification corresponding to that. Now note, that in this particular situation here, when we talk about R 1 opt and R 2 opt, these quantities as we had seen in the earlier lectures they depend on population quantities like unknown mean vector, unknown covariance matrix in case of multivariate normal populations. And hence, for any practical purpose the given data one based on the learning sample, one would be estimating this quantities, and the measure that is based on such quantities is what is call the actual error rate actual error rate or A E R.

Now, this is given by the following expression, that instead of righting it as R 1 opt an R 2 opt, we write it as R 2 opt cap that is a estimate of that particular region based on the given sample data, which is there in the learning sample. So, this is R 1 opt hat f 2 x dx this is a measure. Now this R 1 opt R 2 opt caps and this R 2 opt estimate are based on the learning sample, which was denoted by L. So, these are based on the learning sample which is say script l.

Now once again, if you look carefully at this O E R or A E R. Although A E R has got this estimate here, we still keep it as f 1 x f 2 x.

(Refer Slide Time: 06:20)



So, these quantities are still unknown, now note that O E R or A E R both of this depend on unknown quantities say for example, p 1 p 2 and or as may be the case with O E R or A E R, f 1 x f 2 x. And hence, they as such cannot be applied to compute for a given data what is the O E R or A E R corresponding to partition rule.

Now, a measure that is defined as apparent error rate or A P E R. This is the measure that is going to be based on the sample data. And after the classification has been done, one actually would look at the proportion of misclassifications that is being done based on such measure. Now this is going to be defined, as the fraction of observation fraction of observations in the training sample, because that is what we have in the training sample, that are misclassified by the classification rule classification rule or the classification function. Now, how is that going to be defined, it is based on something which is called a confusion matrix.

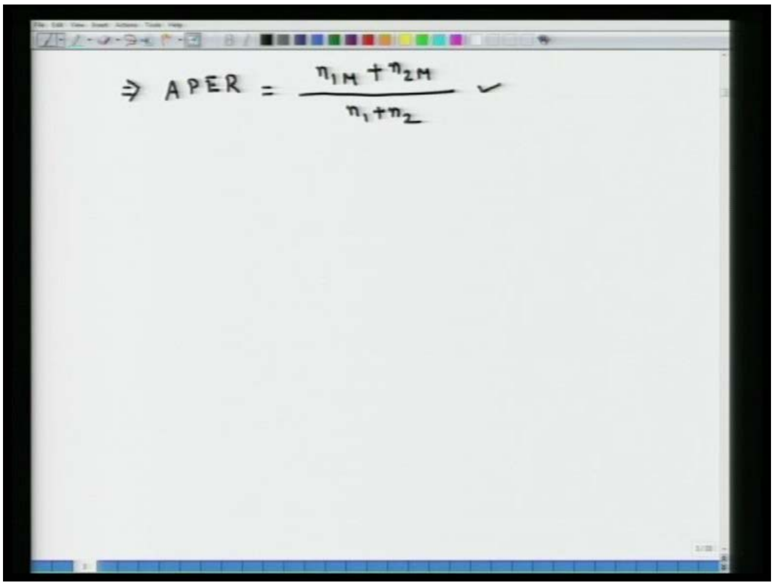
So, one first constructs the confusion matrix after the classification rule is put forward. So, one has got a first start with a learning sample based on the learning sample, learning sample the cases are classified. Then based on those pre classified cases, one has constructed a classification rule. So, the sample classification rule is in place, and using the sample classification rule, one has classified those pre classified examples which where there in the learning sample. And then one looks at this matrix, which is called the confusion matrix. It looks like the following. So, we will be having on one hand predicted class memberships.

So, there can be two such predicted classes pi 1 and pi 2, because we have looking at two class problem. And each observation what we have has got another tag, which is there in the learning sample. So, actual class membership. So, there are two possibilities pi 1 and pi 2. Now, this confusion matrix based on the learning sample that is script l.

Now, suppose in this learning sample, these are pre classified example. And hence, after we use the classification function based on the sample data, when we are going to now classify the feature vectors. We are going to come up with some predicted class memberships. Now, this predicted class membership are going to lead us to this numbers say this is n 1 c n 1 m.

Now, this this n 1 c is the number of observations coming from the first population pi 1 and being correctly classified by the classification rule. So, the predicted class membership is pi 1 corresponding to those observations coming from the first population. So, n 1 c is the number of correctly classified observations, for in the learning sample coming from the first population. Now n 1 m is the number of observations, that are coming from the first population there in the learning sample and by the classification function they are misclassified into this pi 2 population.

(Refer Slide Time: 12:19)



So, this is this notation. Now similarly, we can define this as n 2 m, n 2 m is the number observations which are actually belonging to the second population pi 2, and by the classification function are getting classified into the pi 1 population. And hence, these numbers of this number observations n 2 m are wrongly classified examples, coming from the second population. And similar to this one, we will have a number n 2 c here, which is the number of observation cases coming from the second population, and also being classified to belong to the second population using our classification function.

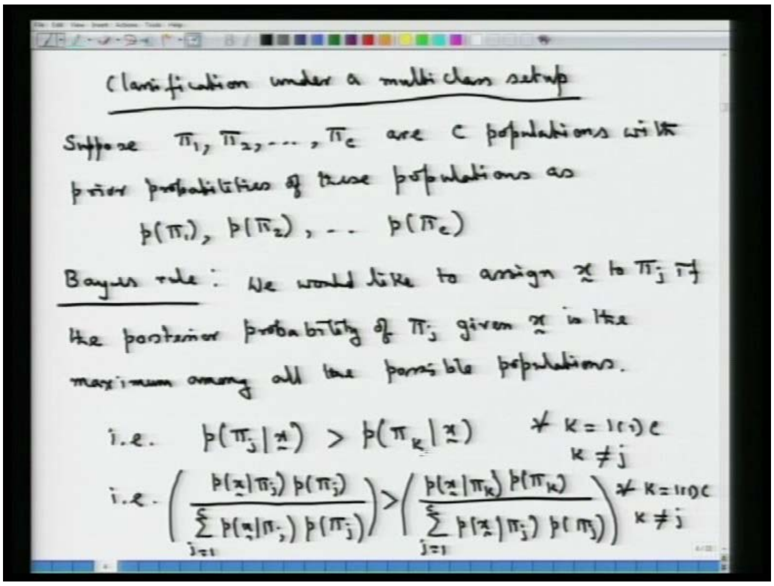
Now, the total cardinality suppose this l learning sample is n. Then, we will be having this, suppose is the sum of this two numbers. So, this is n 1 c plus n 1 m. So, these this n 1 is the number of observations belonging to the first population in the learning set l, and this is n 2 which is the number of observations belonging to the second population in the learning set. And this is the sum of n 2 c plus n to m.

Now, if this is what we get as the confusion matrix which is a derived, when we have a particular classification function in place. Then these observations are misclassified coming from the first population, and these observation n 2 m are the number of observations misclassified coming from the second population. And hence when talk about apparent error rate A P E R, then its fraction of observations in the training sample that misclassified by the classification function. So, this would imply that what we have this A P E R apparent error rate is the fraction of misclassified observations.

So, it is n 1 m plus n 2 m, these are the observations which are misclassified that divided by n 1 plus n 2. So, this is in a perfectly implementable form. So, based on any classification function and the given learning sample l, one can compute this confusion matrix and hence one can get to this apparent error rate.

So, that can be computed for any practical data. So, we end this particular small section on looking at performance measures of for comparing classification functions. Now, we look at the important problem of extending, what we were trying to learn in classification in terms of looking at a multiclass problem.

(Refer Slide Time: 13:19)



So, let us look at that now. So, we are looking at classification under a multiclass setup. So we have, we make provision for more than two populations, suppose we have got now c populations, suppose pi 1, pi 2, pi c are c populations c possible populations with prior probabilities of these populations as say p pi 1, p pi 2, p pi c.

So the problem is simple, that we have got a multivariate normal, multivariate in general not necessarily normal. We have multivariate population and where in the multivariate population, there are c such possible populations. They may differ in the mean vector, they may differ in the covariance matrix or any other measure characterizing that particular population. And then, these populations has got this prior probabilities p pi 1, p pi 2, p pi c, and given an a multidimensional observation, we will have to look at in which class this is going to belong to. So, we are trying to a classify a multivariate observation if is a vector into one of these populations.

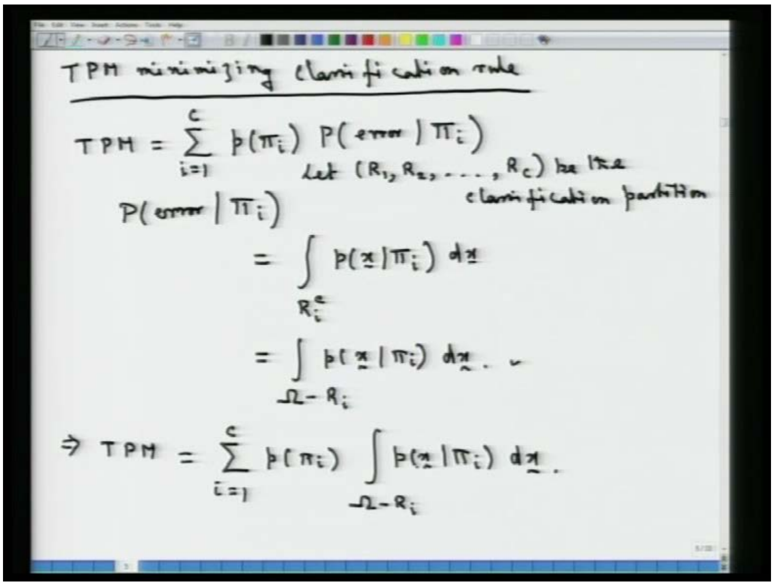
Now, let us look at the three type of classification rules that one can think of or the type of classification rules that we have derive or two population problem. Let us first look at the base rule. Now, the base rule is going to choose that particular population which has got the highest posterior probability.

Now let us write that, we would like to assign an multivariate observation, a multivariate observation x 2 a particular pi c pi j, if the posterior probability of pi j given this x is the maximum among all the possible populations. That is, what we are trying to do here is that let us denote by this quantity p pi j given x to be posterior probability of the population pi j given x is observed. If this is greater than p pi k given x, if this is true for every k equals to 1 to up to c with k not equal to j, then we will assign x the multivariate observation to the population pi j. Because the posterior probability of pi j given x which is this, if that is the maximum among all possible such posterior probabilities for other populations.

So, we have got a k, which is not equal to k. So, this is basically what we have, now this can be return alternatively in the following form that we have got, the using base theorem what we can write straight away is that probability of x given pi j this into probability of pi j. This divided by summation j equal to 1 to up to c p x given pi j, this a multivariate observations p pi j into this multiplied by p pi j. If that is greater than the corresponding quantity on the right hand side. So, this is probability of x given pi k, this into the prior probability of pi k, that divided by summation j equal to 1 to up to c p x given pi j, this into the prior probability of the jth population. If that is true for every k equal to 1 to up to c with k not equal to j.

So, we have got this two give us the base rule, which looks at which population has got the maximum posterior probability given the observation x. So, this is in a nice form, that we have got this to be the base rule, wherein we can infer that this left hand side here what we have is the posterior probability of pi j given x. And the right hand side is posterior probability of pi k for k not equal to j for all other populations.

(Refer Slide Time: 18:19)

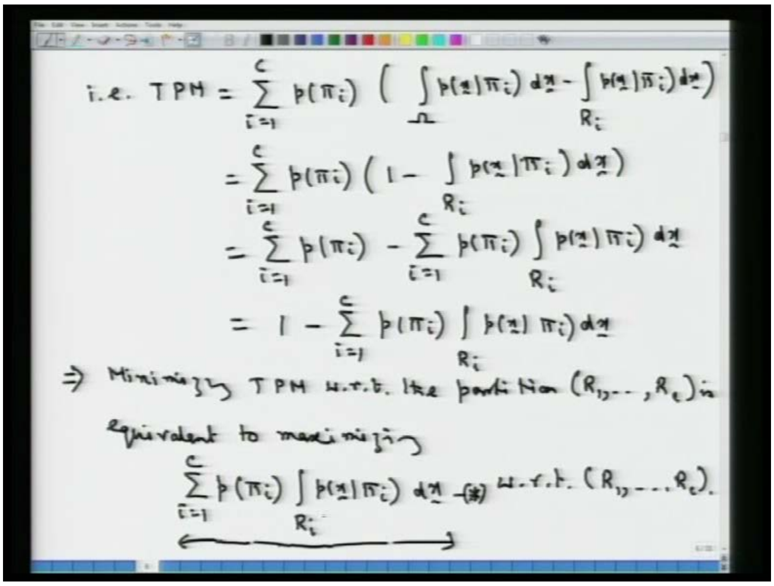


Now, let us look at the total probability of misclassification approach, T P M minimizing classification rule.(No audio from 18:56 to 19:06) Now how is that going to look like in this multiclass problem, the total probability of misclassification the type of concept that we had introduced for the multiclass problem. This will take the following form that it is summation i equal to 1 to up to c, then probability the a priory probability of the pi i population into the probability of committing, and error given an observation is coming from pi i.

Let us see, what are the terms here, because if we are looking at the term by term here. So, the first term is p pi 1 into the probability of classifying that observation which is coming from the first population pi 1, and then putting it into any other population. So, what is this probability of error by the way, this probability of error given pi I, this is going to be given in terms of the partitions that we have suppose we say that, let R 1 R 2 R C be the classification partition; that means, that if x belongs R I, we are going to put it into population number i. That is, for every x that is belonging to this region the class membership is pi 1, for every x that is belonging to R 2 the class membership is going to be pi 2, and for the rest of this also.

So, it is basically that. So, we can write this probability of error, given an observation is coming from pi 1, pi i here. So, that has got a probability which we had denoted earlier p x given pi j. So, this is the density, when we are looking at an observation coming from pi j. Now, this we are actually putting it into some other set here, other than the ith set. So, this is i x given pi I, and then this is integral over the complimentary region of R i. Because if x belongs to R I, then we are going to correctly classify into pi I, otherwise if we are having x belonging to any other R i not equal to that particular term for which we are looking at this. And hence, this is term can be written as omega minus this R i. So, this is p x given pi I d x.

(Refer Slide Time: 22:39)



Now, if this error is given by this particular expression, then this would imply that the total probability of misclassification, for this multiclass problem expression is given by i equal to 1 to up to c p pi I into integral of omega minus R i of p x given pi i.

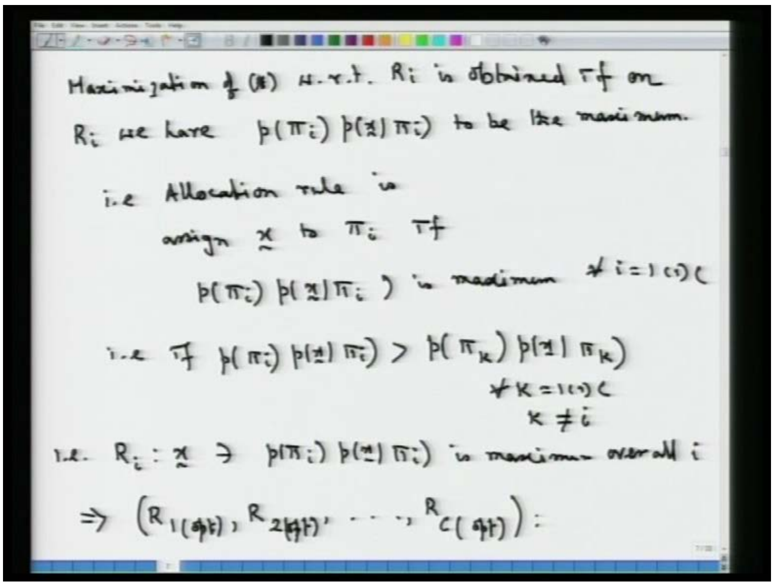
Now, let us see what this term is equal to, this term would be equal to this total probability of misclassification, thus is equal to summation i equal to 1 to up to c p pi i into integral over omega of p x given pi I, the x this minus integral over R i of p x given pi I this dx. Now this expression is equal to 1. So, what we will be having is summation i equal to 1 to c p pi I, this is multiplied by 1 minus integral, we leave it as it is. The second term p x given pi i dx. So, the first term, that we have here after we open the bracket is summation i equal to 1 to up to c p pi I, this minus summation i equal to 1 to up to c p pi I into integral R i p x given pi i to d x.

Now, we note that what is this term equal to, the first term in this expression is equal to 1, because it looks at the prior probability, the sum of the prior probability of all possible c population. And hence, summation of this p pi i terms will be equal to 1. So, it is 1 minus this term here that it is i equal to 1 to c summation p pi i into the term, which will be have their p x given pi i dx. So, this would imply that the rule or rather minimizing the total probability of misclassification is equivalent to maximizing this expression, which is second expression.

So, minimizing total probability of misclassification with respect to the partition R 1, R 2, R C is equivalent to maximizing the quantity, which is summation i equal to 1 to up to c p pi i to integral over R i p x given pi i dx. This with respect to the partition R 1, R 2, R C. Now the minimize this partition, that is going to lead us to minimum total probability of misclassification, thus is same as the partition which would maximize the expression which is given by this.

Now, what is that? So, the optimizing partition, which is going to maximize this particular expression star is going to be, that it is going to be the setup all x's, for which we will have the corresponding term p pi into p x given pi i to be the maximum.

(Refer Slide Time: 25:52)

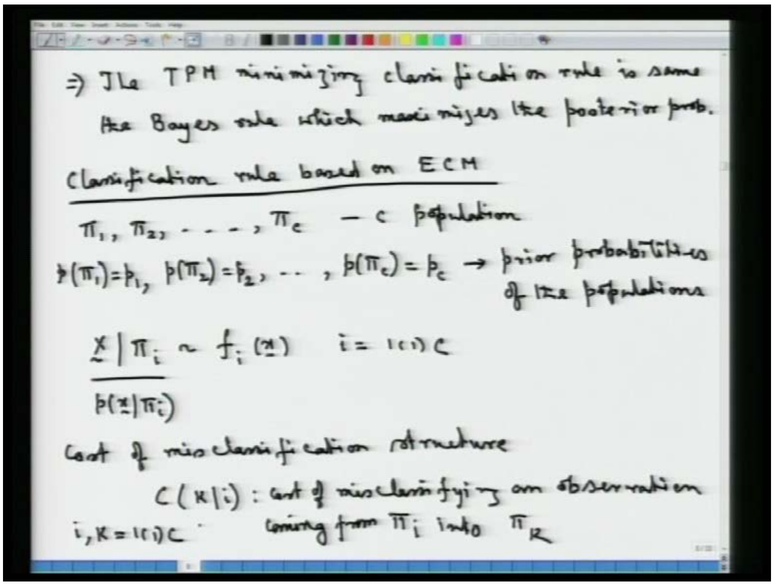


So, what we have here is finally, that the maximization of star with respect to this R i is obtained, if on R i we have the corresponding expression p pi i into p x given pi i to be the maximum this is. So, because we are trying to maximize this particular quantity with respect to R i. So, with respect to each of this R i terms here, we are trying to find out p pi i into p x given pi i. If that is maximum, then the expression is going to be maximized. That is, allocation rule is going to be given by the following, allocation rule is assigned x to pi i, if we have p pi i into p x given pi i is maximum for every i equal to 1 to up to c. That is, if we have got p pi i into p x given pi i to be greater than p pi k into p x given pi k this is true for every k, equal to 1 to up to c where k is not equal to i.

So, this is what is going to give us a partition, that is we have got in other words R i is the region of x's, such that this term p pi i into p x given pi i is maximum over all I, overall say overall I. So, that is basically what is going to lead us to this particular region, and hence if we have this R i, we can construct all the other region. So, we will be having the optimum partition, minimizing the total probability of misclassification. So, this would lead us to this R 1 opt R 2 opt, opt in the sense that it is leading us to the partition which is going to be the partition which would minimize the total probability of misclassification.

Now, what we observe is the following that, when we have obtained this classification rule under the par dime of say the total probability of misclassification, and we had a earlier looked at the rule which was looking at the base rule, which also was essentially the same. So, for a base rule what we had was this condition, this posterior probability to be greater than this posterior probability. However, we have got the both the, a denominators in the left hand side and the right hand side to be the same. And hence that is going to have the base rule is going to be based on this expression being greater than the numerator of the right hand side. And hence, the two rules are basically equivalent, because we have got exactly the same rule here.

(Refer Slide Time: 29:33)



So this implies the total probability of misclassification minimizing partition or minimizing classification rule is same as the base rule, which maximizes the posterior probability. (No audio from 30:00 to 30:11)

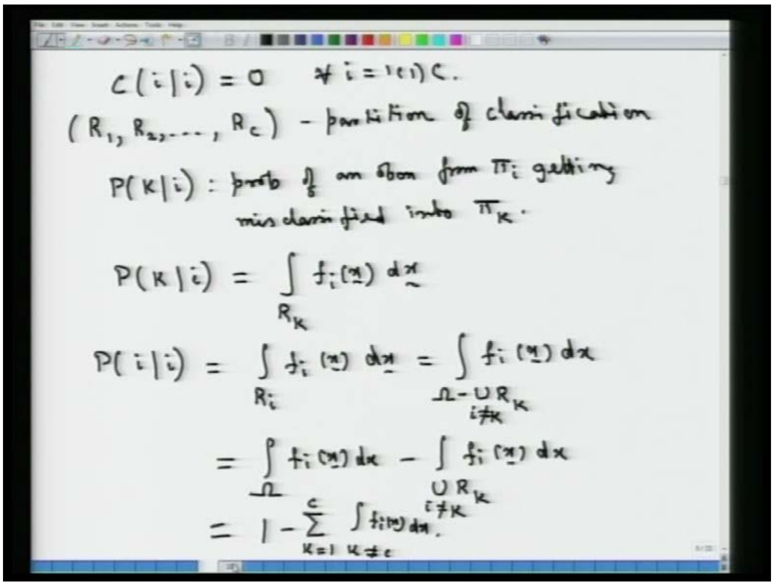
So, this is another justification of looking at either of these two rules. Now next, we are going to look at classification rule based on E C M. Still on a multiclass problem, so we will now look at classification rule or construction of the classification partition based on the expected cost of misclassification. Let us recall, the structure that is what we have, we have got pi 1, pi 2, pi c. These are the c populations, c possible populations into which an observation can belong to, and then we have got these as the prior probabilities, say let us write that is to be equal to p 1 p pi 2 say denote that to be p 2 and p pi c which is prior probability of the cth population given by P C. So, these are prior probabilities of the populations.

Now, prior probabilities of the populations, we will also say that say for example, this the density of x, the density of x given pi i say is given by f i x this is for i equal to 1 to up to c. In our earlier notation, we had perhaps denoted this by p x given pi i. So, that density we are denoting by f i x.

Now, we have got multiclass problem is c class problem. So, let us also look at the cost of misclassification structure, cost structure or cost of misclassification structure. We define that pi c k i. So, this is the cost of misclassifying an observation, coming from the

I th population that is pi i into the k th population, that is pi k. So, we have got this c k i is for i k equal to 1 to up to c.

(Refer Slide Time: 32:58)

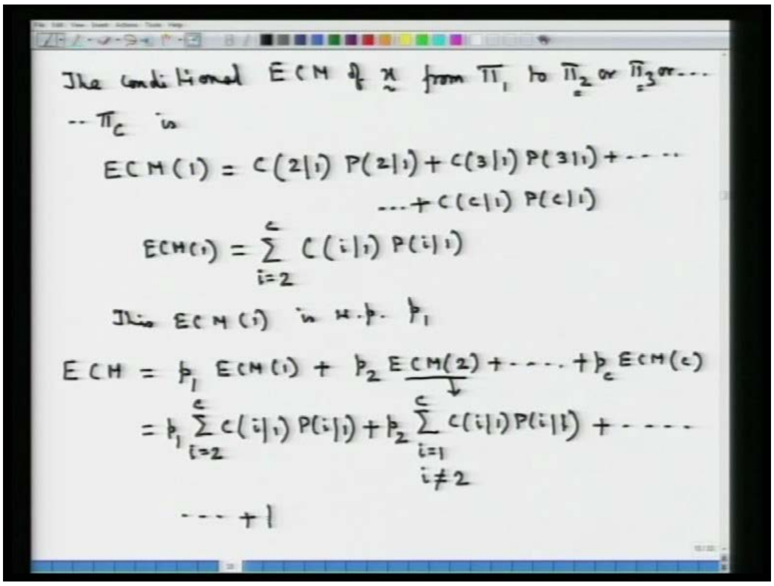


Now, we of course, will be having this term with this c i i, that cost of misclassifying an observation from the pi i into pi i. So, there is no misclassification as such and hence to be equal to 0. This is for every i equal to 1 to up to c. So, once we have this particular term in place, then we will look at constructing the region which is going to look at E C M minimizing rule, but before that we need to actually define how the E C M rule E C M looks like under the presence situation. So, this is the type of partition that we are trying to a get. So, this is the partition of this our classification. Then we can also have a similar notation as to what we had for a two class problem under a cost structure.

So, this is what, this is the probability of let me write that, this is the probability of an observation from pi i getting misclassified into pi k getting misclassified into the population k, that is pi k. Now, what is this equal to p k given i. So, this is going to be the integral over the region R K, because we are putting into the k th population, where in the observation as such is coming from the i th population. So, it has got density f i x dx.

So, we can also see what is p i i equal to. So, this is equal to integral over the region R i f i x dx. So, that you can write this as omega the full space minus union of all other R i's, This i let me write this as k, this i is not equal to k. So, we are looking at the complementary region of R i, that is omega minus union of all other regions, regions which are other than i of this quantity f i x dx. So, this term, the first term would be integral over omega this of f i x d x, which is going to be equal to1. And the next term is over union i not equal to k of this r k regions f i x dx. So, the first term here is going to be equal to 1 and the second term is summation over k equal to 1 to up to c with k not equal to i, because we have taken out that R i region from here. Of these term here, integrals f i x dx. So, these are the quantities which would be require as such to be define the expected cost of misclassification.

(Refer Slide Time: 36:19)



Now, let us build up that expected cost of misclassification in the following way, that we first look at the conditional E C M, the conditional expected cost of misclassification of an observation x from pi 1 say 2 pi 2 or pi 3 or any of the other populations that is pi 1, pi 2, pi 3, pi c, any of these is going to be given by say E C M 1. So, this is the conditional expected cost of misclassification of an observation x, which is coming from pi 1 into any other population other than the first population pi 1.

So, what is this going to be equal to, this is going to be equal to the probability, the cost say first off all suppose let us consider the case that a x is misclassified into pi 2 given it is coming from pi 1. So, what is the cost of that? This is the cost that we are going to incur from the general terminology what we had said was p k i to be I am sorry this c k i the cost of misclassifications structure, c k i is the cost of misclassifying an observation coming from pi i into pi k. And hence, here we are looking at an observation coming from pi 1 and we are putting it into pi 2. And hence, this is cost that we are going to incur and what is probability of misclassifying an observation coming from the first population into the second this is given by p 2 1, where p 2 1 is given by the expression that we have written a for a general k i, k situation here.

So, c to 1 or p to 1 would in particular be integral over R 2 of f 1 x dx. That is how, this term is defined, this plus suppose that observation is still from first. And then, it is classified into the third population. So, what will be having is c 3 given 1. These are all mutually exclusive cases. So, that we will be having this as summation as the expected cost of misclassification, it is a conditional expected cost of misclassification given that it is from 1.

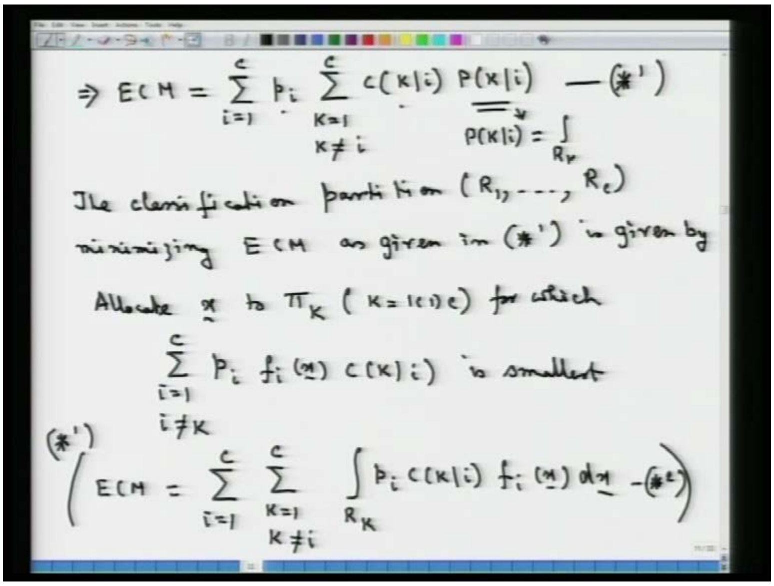
So, that this would be given by this term, this plus the last term would be in this summation will be c given 1. So, this is a c th population. So, this is the misclassifying an observation from the first population into the c th population, this multiplied by the probability of misclassifying this. So, this basically is equal to summation c i 1 into p i 1 summation of i from 2 to up to c. So, this is the E C M 1 term, now this E C M 1 is with a probability p 1. This is E C M 1. So, the expected cost of misclassification E C M 1is with probability p 1, because that is the prior probability for the first population.

So, if we have this conditional E C M of x from pi 1 getting misclassified into pi 2 or pi 3 or pi c to be given by the expression that we written here with a probability p 1, because it is from the first population with a priory probability as p 1. We will have the expected cost of misclassification to be given by the expected cost of misclassification is E C M 1 the conditional 1, that with probability p 1. That times the expected cost of misclassification condition on the observation coming from the second population would similarly be E C M 2 with a probability p 2. And thus we will have to look at all such conditional probabilities, condition cost of misclassification, expected cost of misclassifications, which is going to be given by this E C M c for the c th population.

So, we can write these terms without much of difficulty that this term would be equal to summation i equal to 2 to c. What we have written there, i given 1 into p i given 1, this plus p 2 would be a term. Now, what would be the type of terms that would be there in this conditional E C M of x from pi 2. That is going to be a similar sum, with the sum a little just write it as in the first term here c i given 1 into p i given 1. This summation i is from 1 to up to c, wherein this c is not equal to 2, because we are looking at expected cost of misclassification, their conditional expected cost of misclassification of x from 2. And hence, this would be the expression when we talk about E C M to which is this expression.

Now in a similar manner, we can write the other terms. So, the last term would be equal to summation i equal to 1 to up to c, not c because c is going to be left out here. So, this is up to c minus 1, of the terms which are same as what we have in here.

(Refer Slide Time: 42:10)

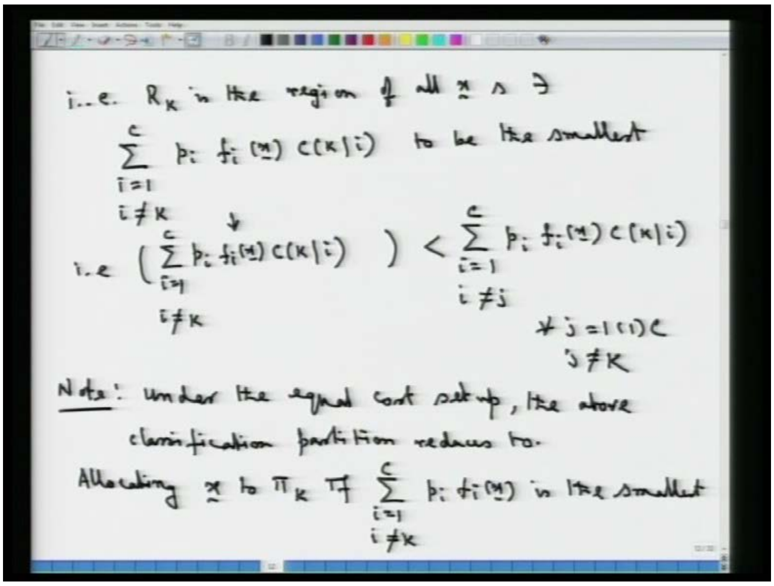


So, we can write this E C M under the present multiclass situation. In the following compact form, that this E C M is equal to summation i equal to 1 to up to c, say that is p i times summation k equal to 1 to up to c, where k will not be equal to c of expressions which are c k given i into p k given i. That is basically, the term that what we have out here. So, the summation is over i for this terms here and then summation over k equal to 1 to c with k not equal to sorry k is not equal to i, because outer sum here is i. So, if we have p 1, then this some k is not equal to 1. So, the summation here for p equal to 1 star from k equal to 2, if we have i equal to 2 that is p 2, we will be looking at summation over k equal to 1 to c without including k equal to 2. And similarly, for the last term if we have here p c, then this particular some would run from k equal to 1 to up to c minus 1.

So, in terms of the partition that we are looking at the classification partition which is R 1, R 2, R C minimizing(No audio from 43:45 to 43:53) minimizing the expected cost of misclassification, expected cost of misclassification as given in this star 1 here, as given in star 1 is given by. We are now looking at regions. So, that we will be allocating x to a particular population pi k, this there are k c such possible populations, k equal to 1 to c for which we will be having the inner sum here, that i equal to 1 to c say with i not equal to k, because we were looking at this pi k for which we will be having this p i into f i x c k given i is smallest, why is that. So, because this particular term if you recall, that that p k i is in terms of this, and hence this expected cost of misclassification. I will just write this expression which would lead one to get to this particular term here, which is summation i equal to 1 to c. So, this is the that is star 1 expression, which is equal to i equal to 1 to c. I will take this p i also inside and write this expression as k equal to 1 to up to c with k not equal to i.

So we will have a p i let me also take it inside. So, we will have any integral here r k f i x dx. So, rather than writing it in this form, I will also take this Constants inside the integral. I will write it as p i into c k given I, this into f i x dx. So, this is an alternate expectation of star 1, in terms of breaking of this particular term here p k i, because this p k i here as we have noted earlier this is integral over the region r k of f i x dx. And hence this star 1 expression is going to be given by this.

(Refer Slide Time: 47:08)



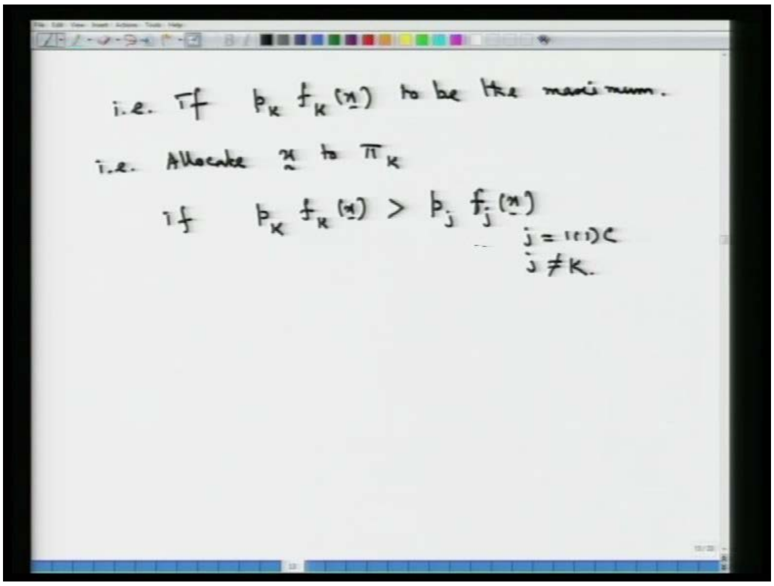
Now, if have looking at the rule the partition R 1, R 2, R C of the sample space which would lead us to minimizing this particular expression. We will look at choosing that particular region are k such that, we will have this term here which is given here to be the smallest among all possible case. And hence, this is what is leading us to the region r k, that is r k is the region of all x's, such that we have got this summation that we have written in the previous slide, that i equal to 1 to c i not equal to k of p i f i x into c k i to b the smallest. That is the expression here which we have is going to be less than the expression for every other k other than the region, wherein we have got this k here. So, that is summation i equal to this is the left hand side only, this is i equal to 1 to up to c with i not equal to k of the expression p i f i x c k given i. If that is less than summation over i equal to 1 to up to c i not equal to j. So, you take out c other term in this here and that would p i f i x c k i, this is for every j equal to 1 to up to c and j is not equal to k. So, we will be looking at all such sums, deleting a particular i, index i equal to 1 to c, and the region r k is going to be the region of all such x's for which this left hand side is less than the right hand side out here.

So, this is what is giving us the rule, which minimizes the expected cost of misclassification. And the allocation rule is what we have written here that to allocate x 2 pi k, if we have got the the summation here to be the smallest among all possible such k terms

Now, we will note that this particular rule here under the equal cost setup. Under the equal cost setup, the above classification partition reduces to what if we look at this particular partition here. It is that we are going to have this to be smallest. Now if all the cost of misclassification are same that is under the equal cost setup. This cost of misclassification is not going to play any rule this can be taken outside.

So, if we have got equal cost setup, the this partition here is going to be reduced to allocating x to pi k. If we have the summation i equal to 1 to c, i not equal to k p i f I x is the smallest, smallest among all such c summations for which we will be having this i not equal to that particular chosen index there.

(Refer Slide Time: 51:05)



So, this is what we are going to get if we assume that we have got equal cost structure. Now, when is this going to be smallest if the term among all the c terms i equal to 1 to c, that is taken out is the largest. That is if we have got the term which is taken out, now what is the term which is taken out if we have looking at i equal to 1 to c, it is a term corresponding to k. That is if we have got p k times f k x to be the maximum.

So, if we have got to be the maximum, we are essentially looking at this rule that we are going to allocate x to pi k. If we have got I said that, this p k f k is a maximum that is if you have f p k times f k x, this is greater than p j times f j x for every j equal to1 to up to c with the j not equal to k. So, we are not looking at that k th product on the right hand side, we are looking at all other c minus 1 products. And hence, we are looking at this all these c terms and then finding out for which of those c possible products. We have got the product to be maximum and then x going to be assigned to pi k corresponding to the population for which this is going to be the maximum.

Now, recall that when we were looking at the total probability of misclassification rule, what was a rule that we had got let us look back. And see what we had got earlier, when we were looking at the total probability of misclassification rule. We had said that the total probability of misclassification rule is one that is going to assign a x to pi i. If the probability product here like this is greater than this type of probability product for all other products other than I. So, this essentially in terms of a our notation, we have in the present case denoted this to be equal to p i and we have denoted this to be f i.

So, what we were trying to say is that, in total probability of misclassification classification minimizing rule, that x is going to be allotted to pi i. If p i f I x is the maximum over all possible such i's, i equal to 1 to see leaving out that particular i which is on the left hand side. Now if we look at the expected cost of misclassification minimizing rule, which we are derived just now, which was this? And that under equal cost setup, we was reduce to allocating x to pi k if the some like this is the smallest or in other words the terms it is left out in this particular sum here is the largest, that is was being allocated to pi k. If the p k f kx is maximum among all possible such products. So, this would imply that under the equal cost , we can just remove this bracket actually. Under the equal cost setup this would imply that under the equal cost setup the E C M minimizing rule is same as that of the T P M minimizing rule.

So, this E C M minimizing rule is same as the T P M minimizing rule. This is what we expect also, because if we considering in the E C M setup, there is no nothing special about the E C M minimization or rather any special structure of the cost is not assumed. Then the rule which would minimize the E C M would naturally be same as the rule which would minimize the total probability of misclassification. And hence, we have also seen that, that is what is happening that if we take in the E C M minimizing rule the cost structures to be identical without having any special a preference about misclassification costs. Then the total probability of misclassification rule can be a derived from the E C M minimizing rule.

So, will stop at this particular point in this lecture. In the next lecture, we will first look at some examples of how to apply for a multiclass problem. This type of concepts of ECM minimizing rule or TPM minimizing rule, and then we will talk about some other important concepts in this classification problem. Thank you.