Multivariate Procedures with R

Prof. Shalabh

Department of Mathematics and Statistics

IIT Kanpur

Week – 10

Lecture – 46

Classification Procedure for Multivariate Normal Distributions

Hello friend, welcome to the course Multivariate Procedure with R. So, you can recall that in the last two lectures we had talked about the linear discriminant analysis and we have transformed our thought process into a mathematical and statistical framework, right. And we have developed the base procedure in the last lecture where we described that how are we going to obtain the classification rule by minimizing the expected loss. And we had assumed that we have certain prior probabilities which are known to us, but we have not assumed any particular form of the probability distribution, but we have used the general term like P1 and P2. But now in this lecture we are going to assume that our observations are coming from a multivariate normal population and these multivariate normal populations are like which are differing only with respect to the mean vectors and they have the equal covariance matrix, right. So, we would like to see if my P1 and P2 are multivariate normal population with different mean vectors and same covariance matrix, then how we are going to create the discriminant rule and how are we going to take a decision that whether to classify an observation into one of the two normal population.

We will be doing it here for two normal population, but this can be extended to more than two normal population without any problem just using the same idea and similar decision rules. So, let us begin our lecture and we try to understand how to do it. Okay, so now we are going to talk about this classification procedure for the multivariate normal distributions, right. So, we are going to assume that our prior probabilities are known to us and we will describe here the general procedure for classifying an observation in the case of two multivariate normal populations, right.

So, these populations were indicated by P1(x), so which is now here indicating the multivariate normal population, say mean vector mu1 and covariance matrix sigma and this P2(x) is the multivariate normal population with mean vector mu2 and the same covariance matrix sigma. So, we can see here that the mean vectors are here different, but the covariance matrix is the same and equal sigma which is a positive definite matrix that we know. So, we know that the probability density function of a multivariate normal with mean vector mu and covariance matrix sigma is given here like this, 1 upon 2 pi raise to the power of P by 2 and the square root of determinant of sigma exponential of - 1 upon 2 X - mu transpose, sigma inverse, X - mu and we have obtained observation on the three variables x1 and x1, x2, x3 as well like this. This is a multivariate observation, right. So, now we try to create the linear discriminant classification rule based on these two populations.

$$\boldsymbol{f(\underline{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})\right]}$$

So, we consider here the ratio of probability density function P1 upon P2 for a given x. We can see here this is here the normal distribution with covariance matrix sigma and in the denominator we have here the second mean vector mu2 and covariance matrix sigma. So, both are the probability density functions of normal mu1 sigma and normal mu2 sigma. So, if you simply try to just simplify it here, you can see here this part will get cancel out and you will obtain here like this expression exponential of x transpose sigma inverse inside parenthesis mu1 - mu2 - 1 upon 2 inside parenthesis mu1 + mu2 transpose, sigma inverse, into mu1 - mu2, right. So, now you can see here this is the ratio of probabilities.

$$\boldsymbol{\frac{P_1(\underline{x})}{P_2(\underline{x})} = \frac{\left(\frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_1)'\Sigma^{-1}(\underline{x}-\underline{\mu}_2)\right]\right)}{\left(\frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_2)'\Sigma^{-1}(\underline{x}-\underline{\mu}_2)\right]\right)}}$$

$$\boldsymbol{= exp\left[\underline{x}'\Sigma^{-1}\left(\underline{\mu}_1-\underline{\mu}_2\right)-\frac{1}{2}\left(\underline{\mu}_1+\underline{\mu}_2\right)'\Sigma^{-1}\left(\underline{\mu}_1-\underline{\mu}_2\right)\right]}$$

So, obviously now what are we going to do? That if you try to observe an observation x, then you try to compute the probability that this observation is coming from the population 1 or the population 2 and you try to classify the observation to the population of which the probability is higher. So, now you are trying to consider here the ratio of two probabilities. So, the region of classification into P1, eta, z for which this P1(x) upon P2(x) is greater than or equal to k, where k is some constant, right. So, now we try to simplify it and yeah, k is I can write down, k is some here, some constant, right, some threshold value, right. And now we try to simplify it and we try to take the natural log of this function, this ratio of identity and if you try to take here like this, so this will become here log of P1(x) upon P2(x) which is greater than or equal to log of k.

$$\boldsymbol{ln}\left(\frac{\boldsymbol{P_1(\underline{x})}}{\boldsymbol{P_2(\underline{x})}}\right) \geq \boldsymbol{ln\ K}$$

$$\underline{\boldsymbol{x}}'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right) - \frac{1}{2}\left(\underline{\boldsymbol{\mu}}_1 + \underline{\boldsymbol{\mu}}_2\right)'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right) \geq \boldsymbol{ln\ K}$$

This again log of k is again a constant, so we do not have to worry. And if you try to take the log, this will be transformed to x transpose sigma inverse mu1 - mu2 - 1 upon 2 mu1 + mu2 transpose sigma inverse into mu1 - mu2 is greater than or equal to log of k. So, log of k again here is a constant. So now, what about the value of here k? So, if I say that this a priori probabilities q1 and q2 are known, then this actually k is q2 upon q1, right. And if I say that suppose if q1 and q2, they are equal that you are trying to say that there is 50 percent probability that the observation is coming from population 1 and 50 percent probability that the observation is coming from population 2, then in that case q1 will become equal to q2 and your k becomes equal to 1 and say log of k will become 0.

$$\underline{\boldsymbol{x}}'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right) \geq \frac{1}{2}\left(\underline{\boldsymbol{\mu}}_1 + \underline{\boldsymbol{\mu}}_2\right)'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right).$$

And in this case, this rule will transform in a very simple format that x transpose sigma inverse mu1 - mu2 is greater than or equal to half mu1 + mu2 transpose sigma inverse mu1 - mu2. And now you can see here that these values are something like some known factor and now the function on the left-hand side, this is only a linear function of X1, X2, Xp. So that is why this is called as a linear discriminant function, right. It is linear because of this X transpose sigma inverse mu1 - mu2 is a linear function of X1, X2, Xp. That is the reason, right.

The other alternative is that in case if this a priori probabilities q1 and q2 are not known, they are unknown, then we may select this log of k equal to 0 on the basis of making the expected losses due to the misclassification to be equal, right. So that and then you will see in the software if they are unknown, the software computes this property also. So that is not a big deal at the moment, right. So now we assume that this observation x has equal probabilities to be classified into one of these two population and the cost of misclassification is also equal. So based on that we have now here two population normal mu1 sigma and normal mu2 sigma.

So, my classification rule becomes here that the observation is going to be classified into the region R1 which belongs to here P1 when x transpose sigma inverse mu1 - mu2 is greater than or equal to 1 upon 2 mu1 + mu2 transpose sigma inverse mu1 - mu2. And similarly, the observation is going to be classified into P2 using the region R 2. If this inequality becomes reverse, so x transpose sigma inverse mu1 - mu2 is less than half mu1 + mu2 transpose sigma inverse mu1 - mu2. So, you can see here this sigma inverse mu1 -

mu2 is constant. So that is why this x transpose sigma inverse mu1 - mu2 is called as linear discriminant function.

$$\boldsymbol{R_1}\text{: } \underline{\boldsymbol{x}}'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right) \geq \frac{\mathbf{1}}{\mathbf{2}}\left(\underline{\boldsymbol{\mu}}_1 + \underline{\boldsymbol{\mu}}_2\right)'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right)$$

$$\boldsymbol{R_2}\text{: } \underline{\boldsymbol{x}}'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right) < \frac{\mathbf{1}}{\mathbf{2}}\left(\underline{\boldsymbol{\mu}}_1 + \underline{\boldsymbol{\mu}}_2\right)'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right).$$

And the quantity on the right-hand side you can see here, this is also a constant assuming that mu1, mu2 and sigma are known. Well, when they are unknown, we will try to replace them by some suitable quantities on the basis of the octane sample, right. Before we try to move forward let me try to just try to give you one more here idea. Can you recall the concept of Mahalanobis distance that we did in the case of testing of hypothesis in the multivariate case? So right, so I am going to now have these things here and based on that I will try to calculate the probabilities of misclassification. So, suppose if I say that let x be a random observation, so there are two possibilities that x is belonging to P1, then the distribution here is normal mu1 sigma and if x is belonging to P2 then the distribution is normal mu2 sigma.

$$\boldsymbol{\Delta}^{\mathbf{2}} = \left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right)'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right).$$

And you can recall that the Mahalanobis distance between the two-population normal mu1 sigma and normal mu2 sigma is given by this quantity delta square which is equal to mu1 - mu2 transpose sigma inverse mu1 - mu2, right. So now we try to create this statistic U which you can see this is nothing but this discriminant rule, right. You can see here if I try to show you this is here the same rule here, right. So now based on this I will not give you here the complete algebra or statistical analysis, but after doing some statistical analysis means I can show very easily that if x belongs to P1 then this statistics U has got normal distribution with mean delta square by 2 and variance delta square and if x belongs to P2 then this statistics U is following a normal distribution with mean - delta square by 2 and variance delta square. So, you can see here that the means of the two distribution they have got the same value, same magnitude, but their directions are opposite.

$$\boldsymbol{U} = \underline{\boldsymbol{x}}'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right) - \frac{1}{2}\left(\underline{\boldsymbol{\mu}}_1 + \underline{\boldsymbol{\mu}}_2\right)'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right)$$

One is + delta square by 2 and another is - delta square by 2, right. So now if you try to understand this thing through this graphic then it is easier for you to understand what is happening and what I am trying to say. Suppose I have here these two normal populations, right, say P1 and P2 which have got here some here mean mu1 and mu2, right. That is a univariate case. So, if you have these two populations and suppose they are well separated, this first population from here to here, second population is from here to here, then in that case there is no misclassification and the observation will be well classified without any problem, right.

If $\underline{X} \in P_1$, then $U \sim N\left(\frac{\Delta^2}{2}, \Delta^2\right)$

If$\underline{X} \in P_2$, then $U \sim N\left(-\frac{\Delta^2}{2}, \Delta^2\right)$

So, these two are well classified population. But on the other hand, if you try to look here in this graphic, now these two populations P1 and P2 they are overlapping in this area, right. And you can see here the way we have computed this distribution of U that is also going to follow our normal distribution. This is the distribution of here U and this is also a distribution of here U. And they are under the two-probability distribution when under P1 and under P2.

So now if you try to see here that U under P1 has got a mean - delta square by 2 and U under P2 has got a mean delta square by 2. And suppose if I say, if you try to look into the first case here, if an observation falls here, I will say okay this belongs to P1, there is no issue. If the observation belongs here in the P2, then I will say that it belongs to P2. Now if you try to look the same thing in this graphic, in the second graphic, let me call graphic number 1 and graphic number 2, in the graphic number 2. Now if you try to see here, if the observation is lying here somewhere I am trying to write here, see here point number 1, then I will say that it belongs to without any problem P1.

So, if the observation, this number 2 is here somewhere, I will say that it belongs to p 2 without any problem. But in case if the observation is belonging somewhere here, suppose this is point number 3, now you cannot say by looking at it whether it belongs to the population P1 or P2. And somewhere it is here 0 and we would like to then compute or estimate the probability of misclassification, right. And based on that we will try to create a decision rule. So, I will try to see here for example if you try to look at point number here 3, then this is belonging to P1 as well as P2.

So, I will try to compute that point number 3 is belonging to population P1 and P2 what are the respective probabilities. And if the probability that point number 3 belonging to P1 is higher, then the probability that point number 3 belongs to P2, then I will try to classify it into P1, right. So, this is how we are trying to do. And now if you try to see here, I have taken here a point C which is here like this in this area. So, I will try to show you that how to compute this probability at a point C in the next slide after I show you that how these pictures will look like in the case of multivariate normal distribution.

So, you can see here that in this graphic number 1, these are very well classified population and these are two normal population which have no region which is overlapping. So, there is no misclassification issue, right. And based on that these are my region R1 and where this R2 and this is here the boundary. So, there is absolutely no issue in this case and the probability that any observation will be classified into P1 and P2, it is 0.5, no issues. This is a very nice picture. But the problem comes here when there is an overlapping of these graphs as in the figure number here too. You can see here now this is the area here where you do not know whether if an observation comes here whether it will belong to the population P1 and P2, right. So, we would like to know how we are going to compute the probability of this misclassification at say point number c, at say point say here c which is here like this thing, right.

If $\underline{X} \in P_1$**, then** $U \sim N\left(\frac{\Delta^2}{2}, \Delta^2\right)$**,**

and $P(2|1) = \int_{\frac{C+\frac{\Delta^2}{2}}{\Delta}}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{y^2}{2}\right) dy.$

If$\underline{X} \in P_2$**, then** $U \sim N\left(-\frac{\Delta^2}{2}, \Delta^2\right)$

and $P(1|2) = \int_{-\infty}^{\frac{C-\frac{\Delta^2}{2}}{\Delta}} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{y^2}{2}\right) dy.$

So, let us try to understand it. So, now we try to understand how are we going to compute the probabilities of misclassification. So, now we have seen that if this random observation is belonging to the population P1 which is a normal mu1 sigma, then this statistic U is following a normal distribution with mean delta square by 2 and variance delta square. And in that case the probability of misclassification, probability 2 given 1 can be computed from the pdf of normal distribution where the range of the integrals will be from the point c + delta square by 2 divided by delta to infinity and the pdf of normal distribution. And similarly, if the observation x is belonging to the second population P2 which is here normal mu2 sigma, then the statistic U is following a normal distribution with here mean here - delta square by 2 and here delta, then in that case this probability of misclassification can be obtained here by here this integral - infinity to c - delta square by 2 and this pdf of normal distribution, right. So now if you try to see once we have given these values, then these probabilities of misclassification P1 given 2 and P2 given 1 can be calculated from the normal probability tables or they can be calculated easily in the R software, right.

So, here this c is known in the case of Bayes solution but here you can estimate them and you will see that when we are trying to implement it in the R software, they will be produced automatically, right. So now we can say that in that we have devised here classification rule assuming that we have equal probabilities of classification and misclassification and there is equal cost involved to be to classify an observation into one of the two normal population mu1 sigma and normal mu2 sigma and this is the decision rule. So, what if you try to see here the quantities here mu1, mu2, sigma, they are unknown to us in real dataset. They are unknown in practice. So, the question is how to implement them in a real dataset and how should I compute this region R1 and R2.

$$\boldsymbol{R_1}\text{: } \underline{\boldsymbol{x}}'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1-\underline{\boldsymbol{\mu}}_2\right)-\frac{\mathbf{1}}{\mathbf{2}}\left(\underline{\boldsymbol{\mu}}_1+\underline{\boldsymbol{\mu}}_2\right)'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1-\underline{\boldsymbol{\mu}}_2\right)\geq \mathbf{0}$$

$$\boldsymbol{R_2}\text{: } \underline{\boldsymbol{x}}'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1-\underline{\boldsymbol{\mu}}_2\right)-\frac{\mathbf{1}}{\mathbf{2}}\left(\underline{\boldsymbol{\mu}}_1+\underline{\boldsymbol{\mu}}_2\right)'\boldsymbol{\Sigma}^{-1}\left(\underline{\boldsymbol{\mu}}_1-\underline{\boldsymbol{\mu}}_2\right)< \mathbf{0}$$

So, the most simple solution is that we can draw a sample and then we will try to estimate the estimators of mu1, mu2 and sigma and we try to replace it here, right. So, because these parameters mu1, mu2 and sigma are unknown in practice, so what we try to do, we get the sample observation and replace the estimated value of this parameter in the place of unknown parameters. So, in order to execute it, we try to obtain here a sample say x1, x2, xn1 from say population 1 which is normal mu1 sigma and the size of the sample here is n1 with this superscript I am trying to indicate this is the population number 1. And similarly, we try to obtain the second sample from the normal mu2 sigma of size n2 like as x1, x2, xn2 and then we are trying to say here this is here, this subscript is indicating here the second population. Then using the first sample we can estimate the population mean vector by the sample mean vector and using the second sample we can estimate the mu2 by the sample mean vector of the second sample observation and we can estimate the sigma matrix by here S like this 1 upon n1 + n2 - 2 and then summation of this xk - x bar from the first sample and the similar quantity from the second sample.

Find $\underline{\hat{\mu}}_1=\underline{\bar{x}}^{(1)}=\dfrac{1}{N_1}\displaystyle\sum_{k=1}^{n_1}\underline{x}_k^{(1)}$, $\underline{\hat{\mu}}_2=\underline{\bar{x}}^{(2)}=\dfrac{1}{N_2}\displaystyle\sum_{k=1}^{n_2}\underline{x}_k^{(2)}$

$$\hat{\boldsymbol{\Sigma}}=\boldsymbol{S}=\frac{1}{n_1+n_2-2}\left[\sum_{k=1}^{n_1}(\underline{x}_k^{(1)}-\underline{\bar{x}}^{(1)})(\underline{x}_k^{(1)}-\underline{\bar{x}}^{(1)})'+\sum_{k=1}^{n_2}(\underline{x}_k^{(2)}-\underline{\bar{x}}^{(2)})(\underline{x}_k^{(2)}-\underline{\bar{x}}^{(2)})'\right]$$

and replace them in place of $\underline{\boldsymbol{\mu}}_1$, $\underline{\boldsymbol{\mu}}_2$ and $\boldsymbol{\Sigma}$.

Then the classification rules become

$$\boldsymbol{R_1}:\underline{\boldsymbol{x}}'\boldsymbol{S}^{-1}(\underline{\bar{\boldsymbol{x}}}^{(1)}-\underline{\bar{\boldsymbol{x}}}^{(2)})-\frac{\mathbf{1}}{\mathbf{2}}(\underline{\bar{\boldsymbol{x}}}^{(1)}+\underline{\bar{\boldsymbol{x}}}^{(2)})'\boldsymbol{S}^{-1}(\underline{\bar{\boldsymbol{x}}}^{(1)}-\underline{\bar{\boldsymbol{x}}}^{(2)})\geq\mathbf{0},$$

$$\boldsymbol{R_2}:\underline{\boldsymbol{x}}'\boldsymbol{S}^{-1}(\underline{\bar{\boldsymbol{x}}}^{(1)}-\underline{\bar{\boldsymbol{x}}}^{(2)})-\frac{\mathbf{1}}{\mathbf{2}}(\underline{\bar{\boldsymbol{x}}}^{(1)}+\underline{\bar{\boldsymbol{x}}}^{(2)})'\boldsymbol{S}^{-1}(\underline{\bar{\boldsymbol{x}}}^{(1)}-\underline{\bar{\boldsymbol{x}}}^{(2)})<\mathbf{0}.$$

Then we try to replace this mu1, mu2 and sigma here by their for example this can be replaced by here mu1 hat, this can be replaced by here mu2 hat, sigma can be replaced by here sigma hat and so on in the entire population of this regions R1 and R2 and we obtain here like this. And then the classification rule becomes here like this x transpose as inverse x1 bar - x bar 2 - 1 upon 2 x bar 1 + x bar 2 transpose s inverse x bar 1 - x bar 2 which is greater than equal to 0 then the observation will be classified into R1 and if this is less than 0 then the observation is going to be classified into R2. So now based on that we can classify an observation when we have a sample of observations and now we would like to introduce here one more concept which is called as APER or apparent error rate and a confusion matrix. So, this is useful when we are trying to judge the goodness of my classification. So, this goodness of my classification is just by the number of cases which are classified correctly or misclassified.

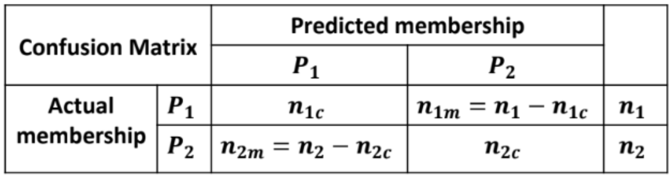
So, based on that we have definition of apparent error rate which is shortly called as APER. This is the fraction of training sample observation which are misclassified by the sample classification function. So, sample classification function is your linear discriminant function that you have observed. Now you try to take a sample and try to classify the observation using your obtained discriminant function and try to see that in how many cases it has classified correctly or incorrectly because that you can measure because you already have a population where you know that which of the observation is coming from which of the population. So, this APER does not depend on the form of the parent populations and can be calculated easily for any classification procedure be it base procedure, be it based on multivariate normal or say anything else.

And this APER is calculated from the confusion matrix. Confusion matrix is a very simple thing. It is a summary of actual versus predicted results of a classification problem. How? Let me try to show you here. Suppose there are n1 observation from P1 and N2 observation from P2 second population.

And now this linear discriminant function is exposed to this n1 and n2 observation. So, some of the observation will be correctly classified, some of the observation will be incorrectly classified and based on that suppose n1c is the number of observations from P1 which are correctly classified into P1. Similarly, this n2c is the number of observations from P2 and which are correctly classified to P2. And similarly, we have number of observations which are misclassified. So, this n1m means misclassified and c means correctly classified.

These are the number of observations from P1 which are misclassified to P2 and similarly n2m is the number of observations from P2 which are misclassified into P1. So, all these observations can be represented in the form of this 2 by 2 matrix which is called here as a confusion matrix which is here like this. There are two ways the predicted membership and the actual membership in the two population P1 and P2. So now you can write down here this n1c, n2c, n1m, n2m here like this and you can see here that n1c + n1m will be equal to n1 and n2m + n2c is equal to n2 because you have the sample of size n1 which is classified or misclassified. And similarly, you have a sample of size n2 which is correctly classified or misclassified.

And based on these numbers we try to compute here the apparent error rate APERS N1m + n2m divided by n1 + n2. So, this is the proportion of the items in the training set that are misclassified. So based on that we try to take a decision whether my classification is good or bad. So now we come to an end to this lecture. So, you can see that we have introduced here that how are you going to create a decision rule when we have a sample from the multivariate normal population.



Confusion Matrix		Predicted membership		
		P_1	P_2	
Actual membership	P_1	n_{1c}	$n_{1m} = n_1 - n_{1c}$	n_1
	P_2	$n_{2m} = n_2 - n_{2c}$	n_{2c}	n_2

The condition here is that these two populations are differing only with respect to the mean vector and their covariance matrix are the same. So, the next question comes that if they are unequal then what are you going to do? But it is not very difficult although I am not covering here but if you try to look into the books this method has been given and even if you have two covariance matrix sigma 1 and sigma 2 some solutions have been suggested. When you try to do it in the R software then you know that it is a very simple thing that you have to simply say variance equal to true or false. Those types of things you already have done in the case of say test of hypothesis. So, you do not have to worry for this procedure but my objective was to give you here a background that whatever R is doing, what R is doing in the background? So now you will understand that okay what should be the outcome? This part is indicating or calculating this rule and based on that these are the observation which we are going to obtain.

So, in the next lecture I will try to implement these concepts in the R software and I will try to show you how you can obtain the linear discriminant function and how you can classify an observation into one of the two populations inside the R software. Definitely as you know in order to understand that thing it is very important for you that you are well aware of this base procedure and the content what I have covered in this lecture for the normal population. So, I would request you that you try to look into the books, try to go through with these concepts, try to understand them and try to revise these topics so that you can understand the content in the next lecture easily and well. So, you try to practice it and I will see you in the next lecture till then goodbye.