Course Name: **Machine Learning and Deep learning - Fundamentals and Applications**

Professor Name: Prof. M. K. Bhuyan

Department Name: Electronics and Electrical Engineering

Institute Name: Indian Institute of Technology, Guwahati

Week-7

Lecture-29

Welcome to NPTEL online course on machine learning and deep learning fundamentals and applications. In my last class, I discussed the concept of linear discriminant analysis that is LDA. In LDA, I have determined a best projection direction for this what I have considered, I maximize between class scatter and minimize within class scatter. And based on this, I have defined the criterion function. And based on this, I can determine the best projection direction. LDA is a supervised technique, because we considered class information.

Today, I am going to discuss the same concept, the LDA linear discriminant analysis. This LDA can be extended for multiple classes, and that is called the multiple discriminant analysis.

So, today I am going to discuss the concept of MDA that is the multiple discriminant analysis for C number of classes. And before going to that, let us consider one example.

So, in this example, I am considering two classes and suppose samples for class ω_1. So, what are the samples? X-$_1$ is a two-dimensional sample. So, the samples are (4, 2), (2, 4), (2, 3), (3, 6) and suppose (4, 4).

So, I have 5 samples. And similarly, suppose I am considering samples for class ω_2.

Samples for class ω_2. So, it is suppose X-$_2$ and again we are considering two-dimensional samples. So, samples are suppose (9, 10), (6, 8), (9, 5), (8, 7) and (10, 8). So, we are considering two classes and we are considering two-dimensional samples for each of the classes. And corresponding to this, you can see the plot of these samples by MATLAB.

So, corresponding to the class ω_1, you can see the red samples and corresponding to the class ω_2, you can see the green samples. So, one sample point is not coming into this figure. So, this red samples that belongs to the class ω_1.

So, you can see the samples are (2, 4) and (2, 3) and (4, 2) and also (3, 6) and also (4, 4). So, these are the samples belonging to the class ω_1.

And these are the samples belonging to the another class that is ω_2. So, corresponding to this case, I have to find the best projection direction I have to determine. So, the problem is I have to compute the LDA, the linear discriminant analysis LDA projection I have to determine. So, we have to compute the LDA projection, what is the best projection I have to determine.

So, corresponding to this, I have to determine the means, the class mean we have to determine.

So, let us see how we can determine the class mean. So, move to the next slide. So, the class mean, class mean we can determine first one is suppose $\bar{\mu}_1$ and that is nothing but $\frac{1}{N_1}\sum_{\bar{x}\in\omega_1}\bar{x}$ and corresponding to this I have 5 samples corresponding to the class ω_1. So, the samples are $(4,2)+(2,4)+(2,3)+(3,6)+(4,4)$. So we can compute this one.

So, this value will be (3, 3.8). This is the mean corresponding to the class ω_1. Similarly, I can determine the mean $\bar{\mu}_2$ and that is nothing but $\frac{1}{N_2}\sum_{\bar{x}\in\omega_2}\bar{x}$ and in this case it is $\frac{1}{5}[(9,10)+(6,8)+(9,5)+(8,7)+(10,8)]$. So, 5 samples we are considering and corresponding to this the class mean is 8.4 and 7.6, this is the second mean $\bar{\mu}_2$. So, we can determine means. After this we can determine the covariance matrix of the first class. So, what is the covariance matrix of the first class? So, this is $S_1=\sum_{\bar{x}\in\omega_1}(\bar{x}-\bar{\mu}_1)(\bar{x}-\bar{\mu}_1)^T$. So, this is the covariance matrix.

So, this can be computed like this $[(4,2)-(3,3.8)]+[(2,4)-(3,3.8)]+[(2,3)-(3,3.8)]+[(3,6)-(3,3.8)]+[(4,4)-(3,3.8)]$. So, finally, I am getting the covariance matrix of the class 1 that is [1 -0.25 -0.25 2.2]. So, this is the covariance matrix of the class 1 the first class. Similarly, we can determine the covariance matrix of the second class.

So, move to the next slide and that covariance matrix for the second class that is $S_2=\sum_{\bar{x}\in\omega_2}(\bar{x}-\bar{\mu}_2)(\bar{x}-\bar{\mu}_2)^T$ and finally, after doing all these calculations you will be getting the covariance matrix for the second class is[2.3 -0.05 -0.05 3.3]. So, you can determine the covariance matrix for the second class. After determining that this covariance matrix we can determine within class scatter matrix. So, what is the within class scatter matrix? So, that is S_w, S_w is nothing, but $S_1 + S_2$.

So, if I add these two covariance matrix. So, we are getting S_w = [3.3 -0.3 -0.3 5.5]. So, this is the within class scatter matrix. After this we can determine the between class scatter matrix, between class scatter matrix that also we can determine and that is nothing, but $S_B = (\vec{\mu}_1 - \vec{\mu}_2)(\vec{\mu}_1 - \vec{\mu}_2)^T$. So, we have these values the $\vec{\mu}_1$ and $\vec{\mu}_2$ and based on these values we can determine the S_B the value of S_B will be [29.16 20.52 20.52 14.44]. So, that you can compute because you have $\vec{\mu}_1$ and $\vec{\mu}_2$ and from this just you can calculate and value this S_B the between class scatter matrix you can compute. After this we have to determine the base projection direction. So, for base projection direction we have to consider the solution of the generalized Eigen value problem.

So, move to the next slide that is the solution we have to consider the solution of the generalized Eigen value problem that already we discussed what is the generalized Eigen value problem for the solution what we need to consider for the solution $S_w^{-1} S_B \vec{w} = \lambda \vec{w}$. So, the we have to find the solution of this.

So, that is nothing but $|S_w^{-1} S_B \vec{w} - \lambda I|$, I is the identity matrix is equal to 0. So, this can be written like this [3.3 -0.3 -0.3 5.5] and we are taking the inverse of this [29.16, 20.52, 20.52, 14.44] - λ and

we have to consider one identity matrix [1 0 0 1] and that should be equal to 0.

So, after computation of this just you have to do the computations. So, I will be getting [$9.2213 - \lambda$ 4.2339 6.489 $2.9794 - \lambda$] we are getting and that is equal to 0. So, after this we have to do the solution of this. So, we will be getting $\lambda^2 - 12.2007\lambda = 0$. So, that means, that $\lambda(\lambda - 12.2007) = 0$. So, corresponding to this I will be get corresponding to this I will be getting two Eigen values. So, one is λ_1, $\lambda_1 = 0$ and another is λ_2. So, $\lambda_2 = 12.2007$. So, I will be getting two Eigen values one is 0 another one is 12.2007 and I have to determine the base projection direction based on this Eigen values. So, move to the next slide. So, we have to compute the LDA projection. So, based on this Eigen values.

So, this is [9.2213 4.2339 6.489 2.9794]$\vec{w}_1$ width. So, 0 that corresponds to the this 0 that corresponds to the Eigen value $\lambda_1 = 0$ and $\vec{w}_1$ $\vec{w}_2$ that is the weight vector because we have to compute the LDA projection the direction we have to determine. And another one is 9.2213. So, that is 12.2007 that corresponds to the second Eigen value. So, corresponding to this two equations I can determine $\vec{w}_1$ that is the one projection direction. So, one projection direction is $\vec{w}_1$ that is [-0.5755 0.8178]. So, this is the weight vector that $\vec{w}_1$ we have computed $\vec{w}_1$ we have computed for the Eigen value λ_1 is equal to 0 and $\vec{w}_2$ also we can compute that is the $\vec{w}_2$ weight vector and that is computed based on λ_2. And this $\vec{w}_2$ that is the direction the projection direction and that would be the optimum direction. Because we are considering the largest Eigen value. So, I have two Eigen values λ_1 is equal to 0 and λ_2 is equal to 12.2007.

So, this larger so the largest Eigen value gives the best projection direction. So, that means the λ_2 gives the best projection direction. So, that means I can write this statement I can write the optimal projection is the one that gives maximum lambda. That means corresponding to the maximum lambda that is the Eigen value I have to find the best projection direction. So, that means the largest Eigen value gives the best projection direction.

So, this we can also obtain directly. So, how to obtain it directly you can see what is the best projection direction I can obtain directly same result I will be getting. So, you can see in this example I am getting this is the best projection direction $\vec{w}_2$ and that I can obtain directly. So, move to the next slide. So, how to obtain it directly? So, because we know that the best projection direction $\vec{w}^* = S_w^{-1}(\vec{\mu}_1 - \vec{\mu}_2)$. So, this equation I know and corresponding to this I can compute this one [3.3 -0.3 -0.3 5.5] inverse and I have to subtract $\vec{\mu}_1 - \vec{\mu}_2 (3,3.38) - (8.4,7.6)$. So, corresponding to this I will be getting the weight vector $\vec{w}^*$ that is the optimum weight vector that is the direction.

So, I am getting the same result as we obtain in the previous slide. So, same result we are getting in the previous slide you can see we obtain $\vec{w}_2$ this is the optimum projection direction and directly also we can compute $\vec{w}^*$ and we are getting the same result. So, this can be shown pictorially. So, how to get the best projection direction? So, move to the next slide. In this case you can see corresponding to the smallest Eigen value I am showing the pink line that is the direction of projection.

So, you can see the projection vector corresponding to the smallest Eigen value and in this case you can see the two classes the samples of the two classes will be overlapping. If I consider that direction of the projection that is the pink direction if I consider and that is corresponding to the smallest Eigen value and that can be shown in this right figure also you can see here.

So, if I consider the class conditional density this pdfs. So, you can see they are overlapping.

So, that means, it corresponds to bad separability. So, this is not a good separation this is a bad separation. So, I have to consider the largest Eigen value and corresponding projection direction I have to consider. So, I can show this one into the next slide. So, here you can see I am considering the largest Eigen value and the corresponding projection direction that is the green colored projection direction and corresponding to this you can see I am getting good separability between the samples of two classes and that you can see from this plot also that you can obtain good separation between the samples between two classes. That means, you are obtaining good separations between the samples of two classes that is the concept of the best projection how can I obtain the best projection I have to consider the largest Eigen value and the corresponding the projection direction I can obtain and if I consider that direction the projection directions I will be getting the best separability between the samples of two classes.

So, this is the concept of the LDA. This concept of the LDA that can be extended to C number of classes the C number of classes that can be extended. So, let us see how it can be extended for C number of classes. So, the LDA for C classes C number of classes. So, earlier we considered only two classes the same principle can be extended to C number of classes and that is called multiple discriminant analysis. So, this is called a multiple discriminant analysis that is LDA for C number of classes.

So, now we have C number of classes we are considering now. So, now we have to obtain C-1 number of projections. So, that means to obtain to obtain C-1 projection directions C-1 projections I have to determine and suppose this is $[y_1, y_2, ..., y_{C-1}]$. So, by means of C-1 projection vector $\vec{w}_i$. So, $\vec{w}_i$ is the projection vector and so you can see to obtain C-1 projections $[y_1, y_2, ..., y_{C-1}]$ by means of by means of C-1 projection vectors projection vectors $\vec{w}_i$. So, this $\vec{w}_i$ can be arranged by columns into a projection matrix.

So, suppose the projection matrix the projection matrix is represented by W that means what we are considering this $\vec{w}_i$ can be arranged by columns into a projection matrix the projection matrix is W. So, that is nothing but we are considering these columns w_1 w_2 all these projection vectors we are considering. So, we have C-1 number of projection vectors. So, corresponding to this suppose $y_i = \vec{w}_i^T \vec{x}$.

So, that means $Y = W^T X$. So, in this case what is X this X is a vector and dimension is m x 1. So, this X is x_1 up to x_m and what is y_{C-1} that is the projection after the projection after the projection I am getting the projection vector. So, this is nothing but y_1 y_2 all these projections we are considering. So, we have C-1 number of projections we have to consider and what is this W W is a matrix that is m x (C-1) and that is nothing but w_1 w_2. So, the columns we are considering the columns are nothing but the projection vectors.

So, this is the weight matrix. So, you can see what we are considering we have this input vector X is the input vector and Y is the projection matrix and this W W is nothing but the projection matrix. So, this is the projection matrix. So, after this what I need to consider I have to consider that projection. So, how to do the projection? So, let us move to the next slide. So, we have n feature vectors and we can stack them into one matrix as follows.

So, $Y = W^T X$ that means, that means we have n feature vectors and we can stack them into one matrix like this $Y = W^T X$. So, where X is a matrix m x n and what are the elements of this matrix? The elements of this matrix are $[x_1^1 x_1^2 \ldots x_1^n]$. So, here this is x_{1m}. So, this is a first feature vector and up to $[x_1^2 x_m^2 \ldots x_m^n]$. So, dimension is m x n and similarly what is the projection matrix? Y is the projection matrix and dimension is (c-1) x n and that matrix will be y_1^1.

So, this column is y_1^1 this is the first projection. The second projection is $y_1^2 y_{C-1}^2$ and like this y_1^n and this is y_{C-1}^n. So, this is the projection matrix we are getting dimension is (C-1) x n and what is the weight matrix? The weight matrix is W and dimension is m x (C-1) that is already I explained. So, we have this the projection vectors $[\vec{w}_1 | \vec{w}_2 | \ldots | \vec{w}_{C-1}]$. So, this is the projection matrix. So, after this what we have to consider in case of the two classes what we determine? We determine within class scatter and similarly in this case also we have to determine the within class scatter.

So, what is the within class scatter? So, what is the within class scatter? Within class scatter is S_w for two classes we computed like this $S_1 + S_2$. So, this can be generalized for C number of classes. So, for C classes we can compute within class scatter $S_w = \sum_{i=1}^{C} S_i$. So, where this $S_i = \sum_{\vec{x} \in \omega_i} (\vec{x} - \vec{\mu}_i)(\vec{x} - \vec{\mu}_i)^T$ and what is this $\vec{\mu}_i$? The $\vec{\mu}_i = \frac{1}{N_i} \sum_{\vec{x} \in \omega_i} \vec{x}$.

So, $\vec{\mu}_i$ is the mean of the class ω_i. So, like this we can determine the within class scatter matrix we can determine. So, this is the within class scatter matrix. So, this within class scatter matrix I can show in this figure this you can see we have determined the within class scatter that is nothing but $S_w = \sum_{i=1}^{C} S_i$.

So, we can determine the within class scatter. So, what is S_i from the previous slide what we have obtained $\sum_{\vec{x} \in \omega_i} (\vec{x} - \vec{\mu}_i)(\vec{x} - \vec{\mu}_i)^T$. So, we can determine S_i like this and what is $\vec{\mu}_i$ that already I have explained in my previous slide. So, this is nothing but a mean corresponding to the class ω_i. So, we are considering all the samples belonging to the class ω_i. So, in this figure you can see so here what is N_i actually N_i is the number of samples corresponding to the class ω_i.

So, it is the number of samples number of samples in class ω_i. So, in this case we are showing this example of two dimensional features. So, in this illustration what we are considering the two dimensional features we are considering that means we are considering m=2 and we are considering the three number of classes. For three classes you can see one is the red one is the green another one is the blue I have shown the means one is μ_1 another one is μ_2 and the μ_3 corresponding to the class 3 the last class 3 classes we are considering. And you can see I am showing the scatter within class scatter S_{w1} S_{w2} and S_{w3}. Now after computing this within class scatter I have to determine the between class scatter.

So, let us move to the next slide. So, how to determine the between class scatter? So, for two classes what we have determined for two classes we have determined the between class scatter like this $(\bar{\mu}_1 - \bar{\mu}_2)(\bar{\mu}_1 - \bar{\mu}_2)^T$. So, for C number of classes that we can also determine the between class scatter. So, we can measure the between class scatter which respect to the mean of all the classes. So, that means for C classes for C number of classes we can measure the between class scatter which respect to the mean of all the classes.

So, I can write like this $S_B = \sum_{i=1}^{C} N_i(\bar{\mu}_1 - \bar{\mu})(\bar{\mu}_1 - \bar{\mu})^T$. So, in this case what is actually $\bar{\mu}$? $\bar{\mu} = \frac{1}{N}\sum_{\forall x}\bar{x}$ that means for all feature vectors irrespective of the classes we are determining the mean. So, that means we are determining the total mean total mean means we are considering for all the classes. So, in the figure you can see I have shown this one this is the mean that mean $\bar{\mu}$ is computed for all the classes.

So, that means for all $\bar{x}$ we are computing the mean and that is the mean of all the classes we are determining. So, that means in this case what is N? N is nothing but that that means all the samples of C classes. So, all the samples of C classes are considered that is the N capital N.

So, we can determine this and the mean the mean of all the classes we can determine and after this this $\bar{\mu}_i$ can be determined this $\bar{\mu}_i = \frac{1}{N_i}\sum_{\bar{x}\in\omega_i}\bar{x}$. So, what is N_i now? N_i is the number of samples number of samples number of data samples in class ω_i. So, you can see here in this figure I am computing the mean for all the classes and based on this mean I am determining the between class scatter. So, we can determine the between class scatter what is the between class scatter that I can say like this it is a distance between the mean of a particular class and the total mean the total mean is $\bar{\mu}$ a mean of all the classes. So, like this you can see S_{B1} S_{B2} S_{B3} we can determine and these are the between class scatters and already I have explained how to determine the within class scatter that is S_{w1} S_{w2} S_{w3} that you can determine.

So, one is the between class scatter another one is the within class scatter. So, in this figure it is clear so, how to determine that within class scatter and the between class scatter. So, for determining the between class scatter we have to determine the mean of all the classes we have to determine that is the μ- we have to determine and after this what we can consider we can define the mean vector of the projected sample that is the projected sample is y. So, the mean vectors mean vectors for the projected samples we can determine the projected sample is y how to determine $\mu_1^{\sim} = \frac{1}{N_i}\sum_{y\in\omega_i} y$ and also we can determine the total mean the mean of all the classes after the projection. So, after the projection we can determine so, that is nothing, but y and this y is for all the classes the all the classes we are considering that is the projected mean of all the classes.

Now, this the scatter matrix for the projected sample y can be determined like this. So, scatter matrix scatter matrix for the projected samples samples y can be determined like this $S_w^{\sim} = \sum_{i=1}^{C} S_i^{\sim} = \sum_{i=1}^{c}\sum_{y\in\omega_i}(y-\mu_i^{\sim})(y-\mu_i^{\sim})^T$. So, for all the classes we have to determine this. So, for all the classes I have to determine the scatter matrix. So, this is the within class scatter matrix and similarly we can determine the between class scatter matrix we can determine that is $N_i(\mu_1 - \mu)(\mu_1 - \mu)^T$.

So, we can determine the between class scatter matrix after the projection. So, for the two classes what we have determined you can see this for two classes we have obtained this one $w^T S_w w$ that is we have expressed the scatter matrix of the projected samples in terms of the original samples. So, that means we obtain like this. So, S_w we obtain like this and S_B also we obtain like this $w^T S_B w$. So, for two classes we have considered like this for two classes. So, what is my objective? My objective is to I have to find the base projection direction that maximize the ratio of the between class to the within class scatter.

So, I am repeating this what is the objective of the LDA or the multiple discriminant analysis I have to find the base projection direction that maximizes the ratio of the between class to within class scatter. So, since the projection is no longer a scalar because now it is C-1 dimension. So, we have to use the determinate of the scatter matrix to obtain a scalar objective function. So, how to obtain the scalar objective function? So, move to the next slide.

So, what will be the objective function corresponding to the C number of classes. So, $S_B^{\sim}$ and $S_w^{\sim}$ that is nothing but $w^T S_B w$ $w^T S_w w$. So, you can see the projection is no longer a scalar because it has C-1 dimension. Then we use the determinate of the scatter matrix to obtain the scalar objective function. And after determining this objective function, we have to find the best projection direction we have to find what we have to find the projection we have to find the projection and that is given by w^* that maximize this ratio that maximizes the ratio.

So, what is this ratio? The ratio is this. For two classes how actually we have obtained the best projection direction we showed the Eigen value problem and that is nothing but $S_w^{-1}S_B\vec{w} = \lambda\vec{w}$ that we considered and where $\lambda = J(\vec{w}) = Scalar$. So, for C number of classes we have C-1 projection vectors. Hence the Eigen value problem can be generalized to the c classes case as follows. So, for C classes what I have to consider for C classes we have to consider like this $S_w^{-1}S_B\vec{w}_i = \lambda_i\vec{w}_i$. So, where $\lambda_i = J(\vec{w}_i) = Scalar$ and we are considering C number of classes $i = 1,2,..,C-1$.

So, that means what we are considering in two classes we are considering the Eigen value problem the solution of the Eigen value problem and λ is a scalar and in case of these C classes we have C-1projection vectors. Hence the Eigen value problem can be generalized to the C classes. So, we have generalized like this so $S_w^{-1}S_B\vec{w}_i = \lambda_i\vec{w}_i$. I have C-1 number of projection direction so it is 1 to up to C-1 projection directions. So, it can be shown that the optimal projection matrix the optimal projection matrix is the optimal projection matrix is projection matrix is w^* that is the optimal projection matrix.

So, it can be shown that the optimal projection matrix is the one whose columns are the Eigen vectors corresponding to the largest Eigen values of the following generalized Eigen value problem. So, that means we have to determine the optimal projection matrix and for this we are considering the generalized Eigen value problem what is the generalized Eigen value problems. The generalized Eigen value problem we have to consider because we have to determine the optimal projection matrix w^* we have to determine and already I told you that which one is the optimal projection matrix that is the one whose columns are the Eigen vectors corresponding to the largest Eigen value of the generalized Eigen value problem. So, this is the generalized Eigen value problem $S_w^{-1}S_Bw^* = \lambda w^*$ this is the generalized Eigen value problem. So, move to the next slide so from the previous slide you can see what is the generalized Eigen value problem that is $S_w^{-1}S_Bw^*$ that is the projection matrix the optimal projection matrix λw^*.

So in this case where $\lambda = J(w^*)$ that is a scalar corresponding to this we can determine the optimal projection matrix that is $\vec{w}_1^*$ and if you see the columns this is the optimal projection vector corresponding to the class ω_1 this is the $\vec{w}_2^*$ is the projection vector corresponding to the class ω_2 and like this we have C-1 number of projection directions and just we are putting in the columns and we are getting the matrix w^* and that is the optimal projection matrix W star is the optimal projection matrix. So, we can see how to determine w^*. So, for C number of classes the principle is same we are extending the concept of the that simple LDA that is for the two classes that can be extended for C number of classes. So, this is the concept of the multiple discriminant analysis. In this class I explained the concept of LDA and I have shown how it can be extended for C number of classes and that is the multiple discriminant analysis.

So, the concept is same I have to increase or I have to maximize between class scatter and I have to minimize within class scatter and that is the fundamental concept of multiple Discriminant analysis. So, let me stop here today. Thank you.