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-2

Lecture-9

Welcome to NPTEL MOOCs course on machine learning and deep learning fundamentals and applications. In my last class, I discussed the concept of Bayesian decision theory, I explain how to determine the decision surfaces between the classes. So during the discussion, I considered that the class conditional density follows a Gaussian distribution and based on this, I have determined the location of the decision boundary. The decision boundary is orthogonal to the weight vector. If I consider the diagonal covariance matrix, the covariance matrix is same for all the classes and I am considering the diagonal covariance matrix.

For this case, the decision boundary will be passing through the point x_0 and it is orthogonal to the weight vector.

The weight vector is nothing but the difference between these two means one is μ_i and another one is μ_j. So that concept I have explained in my last class. The second case is if the covariance matrix is not diagonal, non-diagonal, then what will be the location of the decision boundary. So that concept I am going to explain today.

So let me start this class. So how to find the decision boundary between the classes, the condition is the non-diagonal covariance matrix. So this is a Bayesian decision theory, how to determine the decision surfaces. So in my last class, I have shown that discriminate function $g_i(x)$, it is in the form the weight vector is there, w i is the weight vector and x is the feature vector and I have the bias or threshold. So this is the weight vector.

This is the bias. So this weight vector, it is nothing but the inverse of the covariance matrix and the mean of the class omega i and this bias is ln. So we have determined this. If you see my last class, you can see this derivation and for this what we have considered $\mu\_(i\ ).x\_(t\ ) = x.\mu\_(i\ ).\hat{}T.$ So this we have considered.

So this is the expression for the discriminate function and after this what I considered the case number one, case number one is that diagonal covariance matrix. The covariance matrix is same for all the classes and is a diagonal covariance matrix. So corresponding to this, the weight vector we have derived, the weight vector is nothing but the mean of these two classes. We have considered two classes, the classes ω_i and another class is ω_j , two classes we have considered and also we have considered x_0.

x_0 is the point we can determine like this.

This is mu i plus mu j minus sigma square. So, we have determined x naught like this. So in my last class also I have shown how to draw the decision boundary. So suppose this is the feature space, so suppose it is x 1 and x 2, two dimensional feature space. Now corresponding to the first class, this is the mean vector and corresponding to the second class, this is the mean vector mu j.

After this I have to determine the difference between these two means. So that is nothing but this is the vector, this vector is mu i minus mu j and that is nothing but the weight vector. That is nothing but the weight vector and after this I considered the point x_0. So this is suppose x naught. Now how to draw the decision boundary? The decision boundary I have to draw that should be orthogonal to the weight vector.

So you can see this is orthogonal to the weight vector and it should pass through the point, the point is x naught. So this is the decision boundary. So decision boundary, how to determine? So decision boundary is orthogonal to the weight vector. The weight vector is nothing but the difference between two means μ_i and $\mu_j$and also the decision boundary will pass through the point, the point is x naught. So this is true for this case.

The case is we have the diagonal covariance matrix and covariance matrix is same for all the classes. So corresponding to this case, we have this decision boundary. Now let us go to the case number 2. So in the case number 2, we are considering non-diagonal covariance matrix, non-diagonal covariance matrix. So that means in this case we are considering the covariance matrix is same for all the classes.

So it is same for all the classes but it is not diagonal. In case number 1, we consider that covariance matrix is a diagonal matrix. But in this case, we are considering a non-diagonal covariance matrix. So corresponding to this case, decision surface will be $g_{ij}(x) = W^T x - x_0 = 0$ So what is the weight vector? The weight vector in this case, you can get the inverse of the covariance matrix mu i minus mu j and what is x naught? $x_0 = \frac{1}{2}(\mu_i + \mu_j) - ln$

This probability of omega i divided by probability of omega j and mu i minus mu j whole square and it is the inverse of the covariance matrix. So these will be the expression for the weight vector and the vector x naught. Now in this case, the decision boundary or I can say the decision hyperplane is no longer orthogonal to the vector $\mu_i - \mu_j$So you can see in this case that for the non-diagonal covariance matrix, the decision hyperplane is no longer orthogonal to the weight vector. That is actually the weight vector.

So decision boundary will pass through the point, the point is x naught. So x naught already we have defined. But one difference between the previous case, the case number one, that decision hyperplane is no longer orthogonal to the vector mu i minus mu j. So this is for the case number two. Now let us consider the classifier.

That classifier is called minimum distance classifier. And based on this concept, the concept is nothing but the discriminant function. So how to develop the theory for the minimum distance classifier. So let us discuss about the minimum distance classifier in the next slide. So what is the minimum distance classifier?

Minimum distance classifier, minimum distance classifiers.

So we consider the equiprobable classes we are considering with same covariance matrix. So in this case one in equation number one what we have determined, if you remember that is the expression is gi x, that is the expression for the discriminant function, one by two x minus mu i transpose one plus. So we have this expression. So you know this expression. In my last class I derived this expression.

Now we are considering equiprobable classes. So corresponding to this case equiprobable classes with same covariance matrix, I can write this expression gi x is equal to simply one by two x minus mu i transpose and this is a covariance matrix because covariance matrix is same for all the classes x minus mu i. So we are considering the equiprobable classes. So this part is not so important. So we are writing this expression.

Now the first case we are considering, this is the first case. Suppose the sigma that is the covariance matrix is a diagonal covariance matrix. If I consider a diagonal covariance matrix, diagonal covariance matrix we are considering. So for a diagonal covariance matrix because for a classification we have to see the maximum discriminant function. So which one is maximum, that which one is the largest discriminant function out of c number of discriminant function we have to determine.

So this maximum gi x, the maximum discriminant function means minimum Euclidean distance from the respective mean points. So this maximum discriminant function means minimum Euclidean distance from the respective mean points. So what is the Euclidean distance? So Euclidean distance I can write like this dE that is the Euclidean distance is equal to so here actually I am showing the Euclidean norm between x and mu i. So this is the Euclidean distance. So maximum discriminant function means the minimum Euclidean distance from the respective mean points.

So if I consider the constant suppose constant dE constant Euclidean distance suppose we are considering that means we will be getting the curves of circles and if I consider the high dimensional case then it will be hypersphere. If I consider the constant Euclidean distance I will be getting the curves of circles and if I consider the high dimensional case then it is the hypersphere. So this is the case number 1. So now go to the case number 2. The case number 2 what is the case number 2? Non-diagonal covariance matrix is this.

So that means we have to maximize the discriminant function maximizing gi x is equivalent to minimizing the inverse of the covariance matrix, minimizing of this. So minimizing the discriminant function is equivalent to minimizing the inverse of the covariance matrix norm. So that is actually we have to minimize the Mahalanobis distance. So Mahalanobis distance you can write so already I have defined. So it is x minus mu i transpose x minus mu i this is 1 by 2.

So now let us consider the constant Mahalanobis distance constant the Mahalanobis distance dm is equal to c. So corresponding to this I will be getting curves of ellipses and if I consider the high dimensional then I will be getting hyper ellipses. So for constant Mahalanobis distance I will be getting the curves of ellipse or maybe high dimensional I will be getting hyper ellipse or the ellipsoids. So these two cases we are considering. So for maximizing the discriminant function I have to consider minimum Euclidean distance and also the same thing for maximizing the discriminant function gi x I have to minimize the Mahalanobis distance.

So in the first case we have considered the diagonal covariance matrix in the second case we are considering non-diagonal covariance matrix. So these two cases I can show like this. So what is the meaning of this? So suppose I have two classes so this is the centroid of the first class there is a mean of the first class and these are the contours. I am considering some of the contours. So this is my mu 1 and I am considering another class.

So another class suppose this class. So this is mu 2 mu 1 and mu 2. So what is the weight vector? The weight vector is nothing but the difference between these two means. So this is the difference between these two means I am drawing that is the weight vector. Now my decision boundary will be perpendicular or orthogonal to this vector.

So this is my decision boundary. This is the decision boundary. So these are the contours of equal Euclidean distances. These are the contours of Euclidean distances. So you can see this is the bisector we are considering and bisector is the decision boundary.

So this bisector is orthogonal to the weight vector.

The weight vector is nothing but the difference between the two means mu 1 minus mu 2. So these contours if I consider in two dimensional it will be circle otherwise this may be hypersphere. Similarly if I consider the Mohalanobis distance. So suppose this is one class and we are considering some contours and

this is suppose mu 1 and another class we are considering another class is.

So this is my mu 2. Now I have to determine the difference between these two means. So this is the difference between these two means. So in this case because we are considering non-diagonal covariance matrix the decision boundary will not be orthogonal to the weight vector. So maybe the decision boundary maybe something like this.

It is not orthogonal. This is a decision boundary and this is not orthogonal. This is not orthogonal. So if you see we are considering these are the contours.

These contours you can see these are the ellipses.

In two dimensional these are the ellipses. So that means in this case we are considering equal Mahalanobis distance. These are the contours of equal Mahalanobis distance. So I can write the contours of equal Mahalanobis distance. So for two dimensional it is the contours of the ellipse and if I consider the high dimensional then we can consider hyper ellipsoid. For high dimensional case we can consider hyper ellipsoid.

So you can see the pictorial representation of one is the Euclidean distance another one is the Mahalanobis distance. In the first case we consider the diagonal covariance matrix and corresponding to that case you can see we have the orthogonal vector that vector that is the decision boundary is orthogonal to the weight vector. In the second case for the non-diagonal covariance matrix the decision boundary is not orthogonal to the weight vector. So this is the concept of the minimum distance classifier. So let us consider one problem on this Mahalanobis distance and the Euclidean distance and how you can do the classification by considering the minimum distance classifier.

So let us move to the next slide. So suppose this example we are considering. Two classes we are considering and same covariance matrix we are considering for these two classes. The covariance matrix is sigma.

So it is 1.1 suppose 1.1 0.3 0.3 1.9 is the covariance matrix and corresponding to the first class the mean is mu 1 and it is 0 0 and this vector is 0 0 transpose. Second mean is mu 2 this value is 3 3 and transpose we are considering. Now the problem is we have to classify vector.

The vector is 1.0 and 2.2 transpose. So first let us consider the Mahalanobis distance. Mahalanobis distance between the first mean and the Feature vector is suppose x. So corresponding to this you can see Mahalanobis distance is this is the expression for the Mahalanobis distance.

So it is nothing but 1.0 2.2 this will be 0.95 minus 0.15. So you can check these calculations 0.

55 and it is 1.0 and 2.2. So corresponding to this this value will be 2.952. So this is the squared Mahalanobis distance is like this. So this is the distance between mu 1 and x and similarly we can determine the Mahalanobis distance between mu 2 and x.

That also you can determine this is minus 2.

0 minus 0.8 this is 0.95 minus 1.5 minus 1.5 0.55. So it is minus 2.

0 and minus 0.8. So corresponding to this this distance is 3.672. So you can see these two distances we have calculated and based on this this input Feature vector is assigned to the class with the mean vector 0 0 because corresponding to this mean vector I have the minimum distance. So I can write the vector x is assigned to the class with mean vector a mean vector is 0 0 transpose. So based on this Euclidean distance we can do the classification. This is the minimum distance classification.

So corresponding to the first mean your minimum distance is 2.952 corresponding to the second mean the distance is 3.672. So then we have to consider the mean mu 1 that is the 0 0 so that the Feature vector is assigned to the first class. But again if I consider suppose again given vector the vector is the Feature vector and vector is 1.

0 2.2 transpose is close to the second mean with respect to with respect to Euclidean distance. So if I consider Euclidean distance the result will be different. In case of the Euclidean distance the Feature vector x is assigned to the second class. But in this problem we are considering the co-variance matrix.

So that is why we have to consider the Mahalanobis distance. So based on the Mahalanobis distance we can determine the particular class. So we can decide. So that is the concept of the minimum distance classifier. So this is the fundamental concept of the

Bayesian decision theory. So up till now we have discussed how to determine the particular decision boundary.

So in this class I discussed how to determine the location of the decision boundary. I considered two cases in the first case I considered a diagonal co-variance matrix. The co-variance matrix is same for all the classes and corresponding to this I have determined the location of the decision boundary. The decision boundary is orthogonal to the weight vector. Weight vector is nothing but the difference between mu i and mu j the difference between two means and it should pass through the point the point is x naught.

This is corresponding to the case one. In the case number two we considered non-diagonal co-variance matrix. In this case the decision boundary is not orthogonal to the weight vector but it passes through the point x naught. So that is the difference between case one and case two. After this I discussed the concept of minimum distance classifier based on Euclidean distance and also the Mahalanobis distance.

So these concepts are quite important. The concept of the minimum distance classifier the concept of the Bayesian decision theory and how to determine the decision boundary corresponding to the normal distribution. Normal distribution means we are considering that class conditional density follows normal distributions. So let me stop here today. Thank you.