Indian Institute of Technology Madras
Presents

NPTEL
NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING

Pattern Recognition

Module 02

Lecture 09

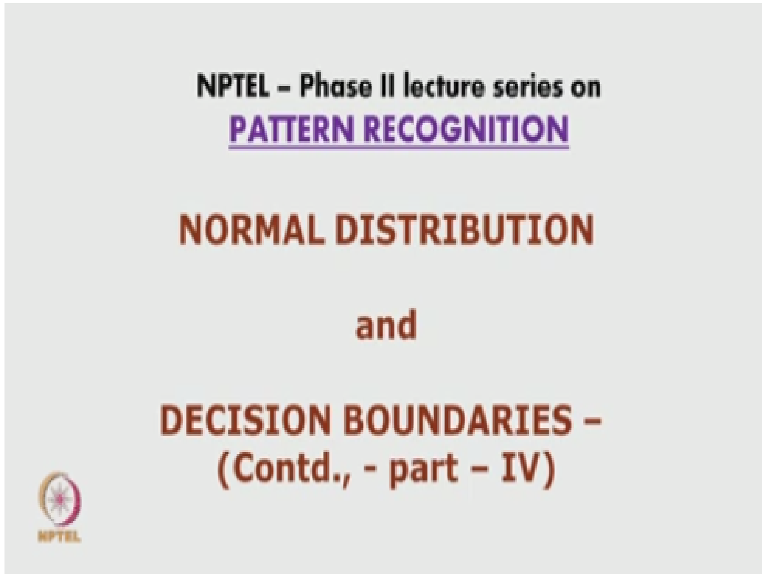
Linear Discriminant Function and Perceptron

Prof. Sukhendu Das
Department of CSSE, IIT Madras

Welcome back to the lecture series of pattern recognition. In the last class we were discussing about how normal distribution, when put under the formulation of the base decision rule or base theorem of posterior probability class assignments, lead us to a distance function which we termed as the Mahalanobis distance. The Mahalanobis distance is actually governed by the covariance matrix.

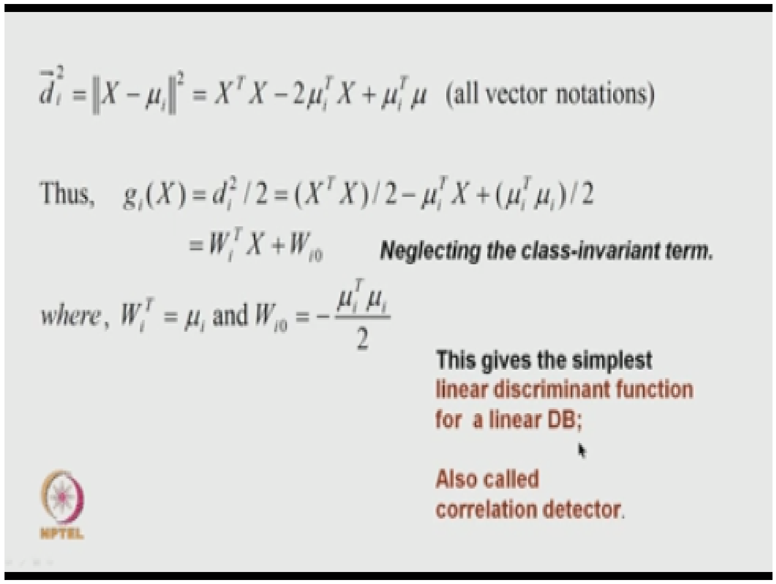
And we took the very simplest form of the covariance matrix equal to an identity matrix in which all the variance and the covariance terms, in fact the variance terms were equal to unity of diagonal covariance terms were equal to 0 and all. And such cases we actually got the Euclidian distance criteria which we actually neglecting the case variance term, we so that we are able to get a linear discriminant function which is actually linear distant boundary now today.

(Refer Slide Time: 01:19)



So let us continue, if you look back into the slide.

(Refer Slide Time: 01:21)



And let us revisit these equations on the top, so what you have is the expression of the distance to the class mean for a particular sample under the assumption that the covariance matrix is equal to an identity matrix, you will get this expression. And considering gi, in some parts of the lecture today we will switch between Gi and gi although both of them almost mean the same, just normalizing factor to take off this factor to out here to divide in the expression by 2 and you get this.

So this is what we got and a linear discriminant function. And this is obtained by ignoring the class invariant term, this is what we have done in the last class. So we have a weight matrix with the class mean i^{th} class mean, and multiplied by the test sample X, and this is actually considered as bias term which is dependent on the class mean. So this is a linear discriminant function which actually gives us, which is also called correlation detector. But we ignore that, so we will proceed with this where we stopped from the last class.

(Refer Slide Time: 02:32)

The discriminant function (DF) for linearly separable classes, can also be expressed as: $g_i(X) = W_i^T X + W_{i0}$

where, W_i is a dx1 vector of weights used for class i.

This function leads to DBs that are hyperplanes. It's a point in 1D, line in 2-D, planar surfaces in 3-D, and hyperplanes in higher dimension.

Henceforth, consider DF to be DB.

In case of 3-D, with a plane passing through the origin, the expression gets the simplest form:

$$(\omega_1 \omega_2 \omega_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

So the linear discriminant function for linearly separable classes can be expressed as this, and we know what are the corresponding terms where this term Wi is basically a dx1 vector of weights used for class i. Well in this particular case we are talking about a D dimensional space that means you have picked up D different features or found the test samples either during training or during testing which we have discriminated earlier in terms of supervised learning.

And we actually see a small example in end of the class today in terms of what can be considered as a training as well. So that is Wi, of course you have the expression of Wi0 earlier. So this function gi(X) and this particular form leads to decision boundaries or DBs which are hyperplanes and higher dimension. It is a point in 1-D, line in 2-D, it is a planar surface in 3-D and called as in general hyperplanes in higher dimension.

So hence forth you will consider although there is a small difference between a discriminant function and a decision boundary we will see these difference and kept how decision boundaries can be considered using discriminant functions we will see. But for the time being we will consider both of them to be identical for, either for a class. Of course, you must remember one particular fact here, that the discriminant function is valid for a particular class I, whereas the decision boundary is between the pay of classes.

I and J, class 1 and 2, class A and B, then is what you consider a decision boundary whereas the discriminant function is for a particular class. Now when we are talking about a decision boundary between two classes, there are often certain classifiers design where you try to classify samples belonging to a particular class with respect to all other classes, that sort of a discrimination or a categorization or classification is often termed as one versus all or one versus the rest.

So in that case you will get a discriminant function or a corresponding decision boundary, but however whether it is one versus one or one versus the rest, the one versus one is what we talked about between two classes I and J, of course I is not equal to J, class A and B let us say, fruits and flowers or between two categories of flowers, two categories of fruits or two categories of bags for where enough samples are taken for categorization and all that.

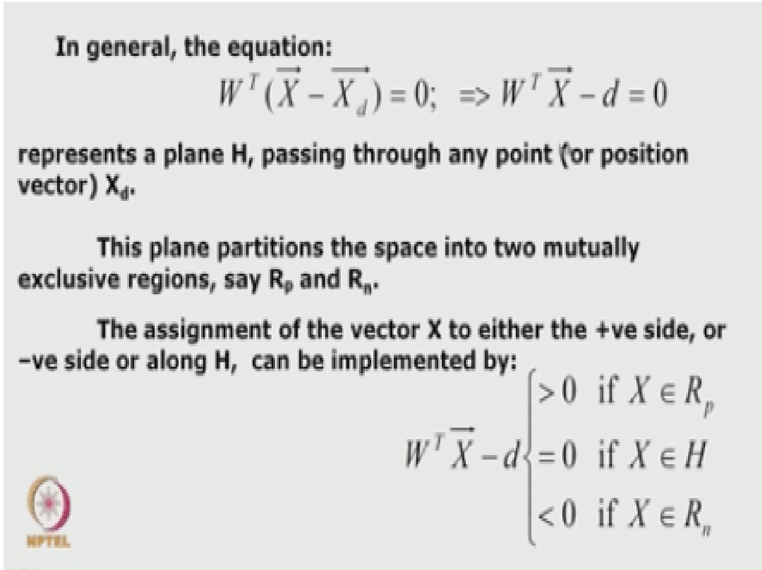
It could be between two different phases or fingerprints in the case of pattern recognition applications. We say that a discriminant function typically is for a class and decision boundaries between pay of classes, but for sometime in the class today, we will not discriminate between the two, but immediately we receive that one of the sides will have an equation where, the DB or the decision boundary is derived from the discriminant function. Let us go ahead.

So in the case of 3-D is basically a plane correct, it is planar surface in 3-D plane, if it is passing through the origin or simplest form of this expression of $Wi^T X$ will be in this particular form. Because there will be three components, both for W and X and this class vast term will be equal to 0, because the plane is passing through the origin. If you substitute all zeros null here, this equation is satisfied.

So this is a very special form of discriminant function or a decision boundary. And we will analyze this expression now in a little bit dated detail and try to see the difference between a discriminant function and decision boundary plus, how do you find weights, how do you find this weights Wis for the purpose of classification. So we will now try to understand, given a set of test samples or training samples sorry.

I repeat again given a set of training samples for you, how do you define these weight which will help you in categorization. It will help you to build up the discriminant function as well as the decision boundary. So we will observe this equation and have some sort of a geometrical graph base representation of using a graph basically two dimensional graph in 2-D space is what we will look. And see the interpretation of and significance of these weights and test samples in the next set of slides.

(Refer Slide Time: 06:45)



So in general the equation is in this particular form. We had this expression so, this is another, I would like to warn you with this D here, do not confuse this D with D dimension of the problem. This is just a notation here, and the expression which we saw in the previous slide, which was in this particular form, Wi^TX+Wi0 is actually being written in this particular form. And you can actually take the weight common out, but W^T multiply and Xd which is actually a vector is actually a scalar distance D.

And we will find the significance of this D very soon. So it represents a plane a hyperplanes passing through an arbitrary position or a point in D dimensional space called Xd. So this plane now H which is represented either basically almost by this equation is passing through this particular point. Of course, we will see representations and diagrams in 2-D, which is easy to visualize for you.

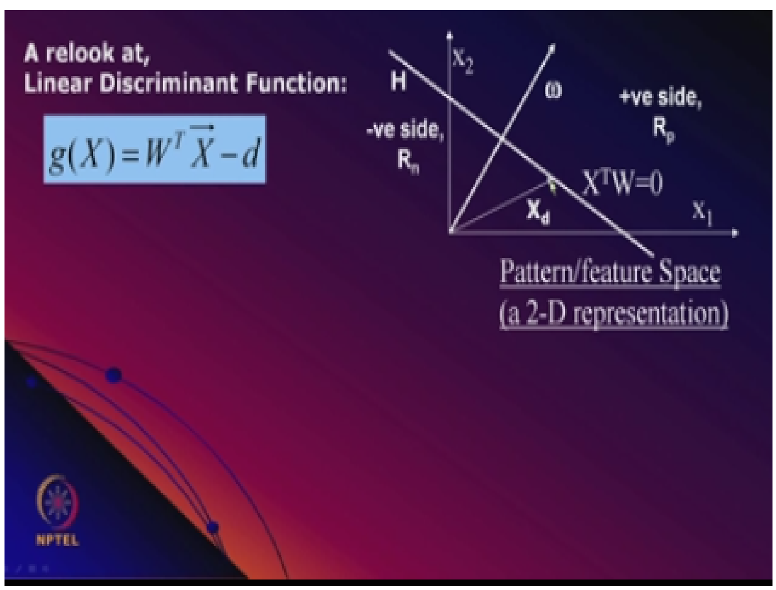
And this particular plane partition the space into two mutual exclusive regions Rp and Rn, both of these two regions are considered to be semi-infinite regions. We talked about this in the earlier class itself. So the classes are meant rule based on the something similar to what we do also in base classifier, although there is no probabilistic function earlier can be simplified using this discriminant function as the following okay.

So if the vector x is lying on the positive side okay which is represented by this Rp ps in Rp indicates the positive side with respect to the plane H what is positive basically means this function is give a positive distance positive value then we assign the sample x to these cluster are the class belonging to the positive side.

If it is negative it belongs to other side Rn negative side of course if the value is 0 then the point is lying on the hyper plane H so it is the simplest criteria here based on which we can do classification provided of course you must remember that this corresponding linear decision linear discriminant function is representing a decision boundary the decision between 2 class gain 1 verse the 1 or 1 vs the rest.

So we will observe this the geometrical interpretation of this with the help of next slides I mean analysis first the linear decision boundary then we will proceed towards nonlinearities in the corresponding discriminate function.

(Refer Slide Time: 09:12)

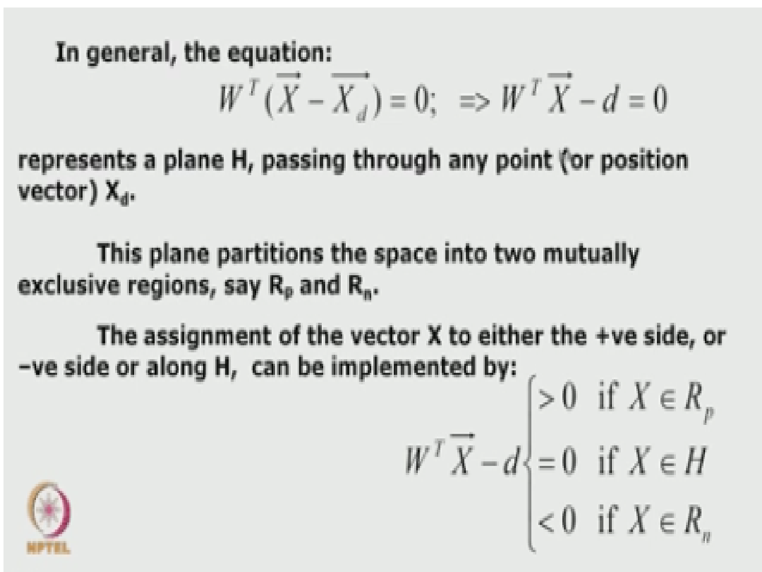


If you look at this slide the same equation which we got for the linear discriminant function remember I did inform you that we will not discriminate now between discriminate function decision boundaries both can be consider for the time being to identical is represented by this equation the background of the slide have purposefully changed because some of these diagrams we need some color markers and color plots.

So it will be visible we will get back to the form later on so look at this particular graph, so this is a 2 dimensional space x1 and x2 are the 2 features for the samples taken so this is 2 dimensional vector x is also 2 dimensional vector which is any point in the space and this H the hyper plane H which is reprinted by this equation g(x) = 0 is represented by this plane H and which is given by basically by this equation here.

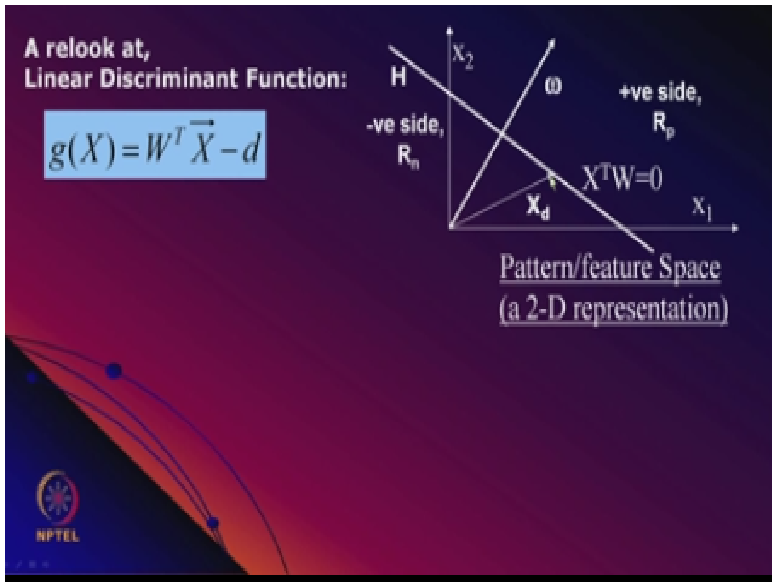
Okay this equation is equal that means if you take up a point x and put it anywhere in this plane H this corresponding equation will be stratified remember this xd we have seen it in the previous slide.

(Refer Slide Time: 10:15)



Let us go back this is the Xd we are talking about it is a point on that plane because the plane H is passing plane and higher times it is a line in 2D it is passing through any point Xd.

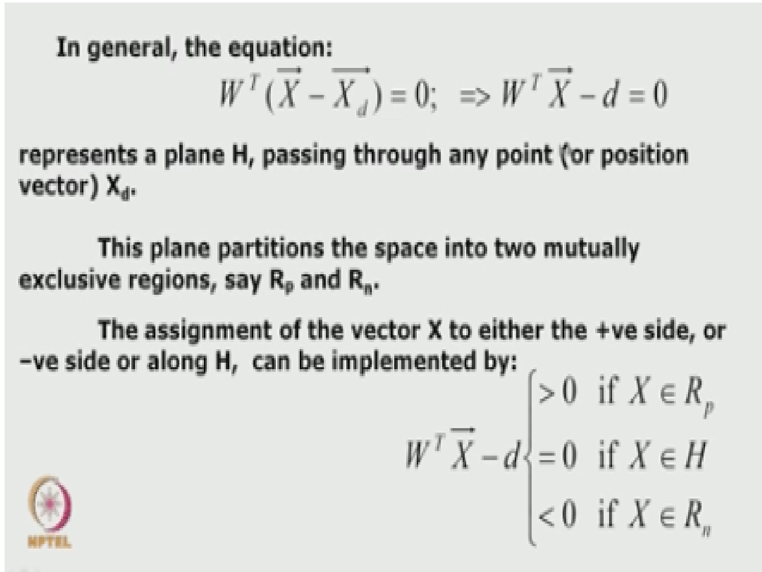
(Refer Slide Time: 10:27)



So this is the point Xd now it is lying on the H this the point Xd which is lying H and the normal to the plane H is basically the vector w, now this w is basically meaning the same w here okay so this not a just a showing a different symbol but it is basically orthogonal to the plane so we at looking at this what we call as the pattern or feature space and look at the 2 dimensional representation of this expression for the hyper plane H.

So it will be aligned in 2 D again I repeat you can visualize this to be a plane in 3D and hyper plane higher dimension okay so you can see this is the positive side Rp this is the negative side Rp what does it mean if you take a sample x and the x sample lies here somewhere in this region which is the positive side of Rp.

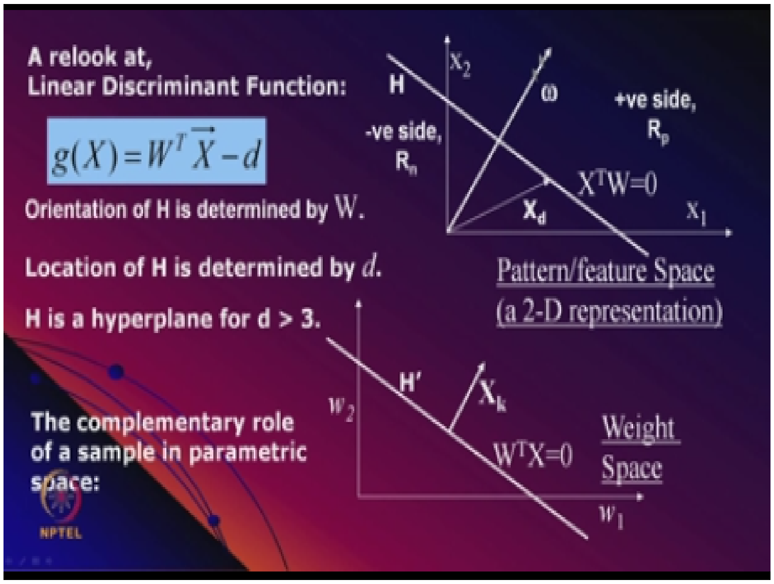
(Refer Slide Time: 11:13)



Then go back to the discriminant function criteria here you will have positive value so we can say that x lies within the positive region on this particular category or class if it lies somewhere here below the H here the negative side Rn to be very precise and specific then you have a negative value of this g and you can say that it belong to the other class however the point lies somewhere H you will have the function going equal to 0.

And then of course in this particular case the plane could pass through the origin so the D is observed in this expression when you write in this particular form when it is a dot product $W^T X$ or $X^T w$ is basically going to mean the same.

(Refer Slide Time: 11:49)



So this is an interpretation of this particular expression where we say that all thought this is a discriminant function and a linear function it is also representing a decision boundary you have a positive side on one side and negative side on the other one which provides you a boundary between two particular classes orientation of H is determinant by this w remember this is w here and the location which is determinant by D which is the something can be consider to be a perpendicular distance from the plane with respect the origin okay.

Which basically projection on this Xd vector on to w okay so that is what and if of courses hyper plane if D is more than 3 look at this complementary role in the parametric space, so what I have drawn in this curve below is a complementary plot of the pattern feature space which we call as the weight space okay see the same equation here but unlike in the previous case where we had 2 access x1, x2 as the 2 feature components along the access insisted of that I have put w1 and w1 which are the components of W the weight matrix.

Again you are operating in 2D then we get the same expression we can get the same line the only difference now is look at the way I have transpose this equation instead of x^T w we have written W^T x so this is weight space that means when you move around the space you are talking about different values are combination of this W1, W2 when you are moving around the pattern features space you are talking about x1, x2 combinations.

So that is the complementary role it is something like you have moved from the pattern feature space to the weight space here the normal to the H is not represented by x instead of the W here the W was dictating the normal or the orientation of the H okay here it is dictated by the sample x so what does this weight space in terms of the decision rule says us now remember this plane is dictated by a sample x.

This plane is dictated by W this plane is dictated X you move around you get different sample points you move around you get different weights this is a negative side for a set of sample this is a positive side for set of samples give away that means if you have decided the weights the plane is decided the question is who give you the weight will that expression is not we have expression because it is dependent on the class means short of the discriminant function okay.

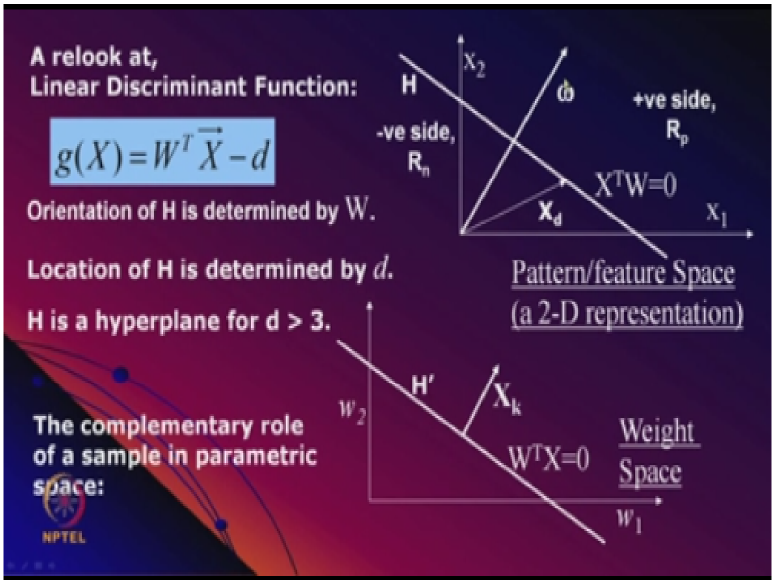
For a decision boundary we will also have the expression of w but so the given w decides orientation and the position of the plane here a sample decides the position of H that is means given a sample now if you have a choose of weights and you are here on the positive side that mean that means set of weights here in the positive side will give a positive value of g the set of weights in te negative side of this hyper plane will give a negative side of the g.

Okay so when you keep choosing different samples you are having different orientations of H if you chosen different weights you have different ordinations of w we will remember this is concept because we are going to use this and extend this idea about and give some idea about design are how to pick up W's in general for a set samples given because sometimes here in training you may not just have 1 set of samples but a set of samples from 2 or different classes.

So in that particular case how do you deicide the W okay so look back in the pattern features space the w decides the hyper plane in the weight space we just complementary okay see the w has become the variables here the variables where x this variables x have become up to the parameter of the plane H so I mean just call you know weight space the independent variables become the parameter in the pattern space where the patterns are independent variables w become the parameter of x.

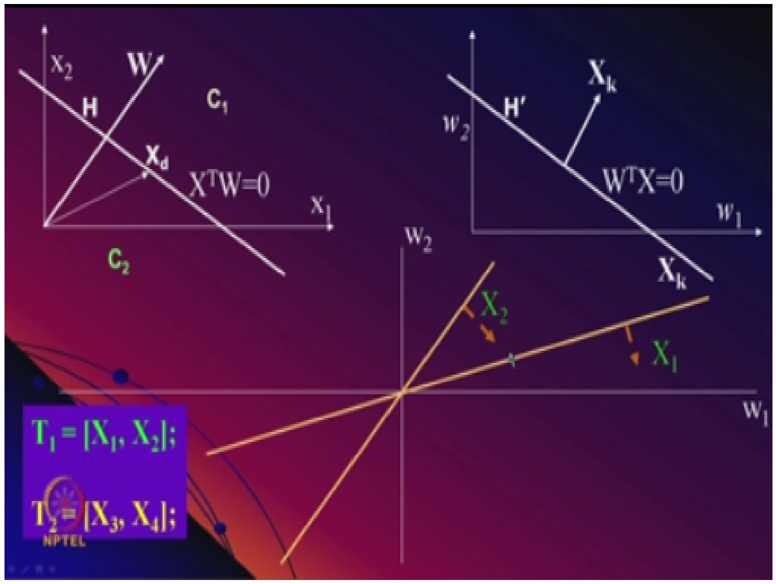
So this complementary role between pattern and weight space you can just keep in mind and these type of concepts are use in many other different applications the typical explanation for those in the working in the imaging domain you can look into concepts of half transformer half space where you actually go from special domain to parametric space, so as a similar analogies are most shown here where the role of the access variable and the parameter okay they are mutually swapped here, so keep this in mind we will probably use one of them but sometimes interchange them so given a point x will be able to show w go back here.

(Refer Slide Time: 16:30)



The point x basically indicates a plane H with the corresponding orientation a point here indicates a weight vector which is basically dictates a plane H here, so you can see the reversal of the role between the parameter and the axis variables in this particular case remember we assume the same expression.

(Refer Slide Time: 16:54)



So these were the two okay so you have the this is the patterns is or features piece this is the weight space we can see that the sample dictates the position on sorry orientation of H here the weight W this is the actual correct notation of W which is normal to this H and weight dictates the position of the hydro planes, so both of these planes are complementary when you are searching for samples you are moving around here we are searching for weight we are moving around here.

Keeping this in mind let us observe the weight space let us observe the weight space what is the impact of different set weights on set if samples which are given to you okay, so for the timing what we are deciding here is let us say for class T_1 this is just a level or a class 1 you are given two set of samples X_1 and X_2 for set of samples derived a class 2 you are given X_3 and X_4 so now what we are trying to see is from the discriminant function we are trying to get a linear decision boundary.

Belonging to which we will split the space into two different classes are two different regions or two different categories and typically in the case of training you often have several samples is actually have several set of samples when an to each class sometimes of course the samples at very less in which the training actually obvious stand actually become little bit more are only one so different but in general sometimes you have lot of samples in the case of machine learning or pattern recognition.

In most applications you have a huge number of train samples so instead of taking a very large for the sack of visualization we are observing this concept in two dimensional space and just considering two samples for each class two samples for each class let us look back into the slide so for class indicated by this level T_1 you have just two samples X_1 and X_2 which are given into the system for the class table two you have two other samples X_3 and X_4, now given this two pairs of samples for each particular or one pair.

For each class one pair of each class and there are two classes or two pairs how do you find a W do you use the equation which was given at the discriminant function directly well may not there are of course the equations of decision boundaries but what we are going to see here is a process which we lead as to another type of classifier which we will just introduce and which gives as a linear decision boundary and during training with two different sample, let us observe as to what are the correct set of weights.

It seems like can we identify some short of region in the weight space where we can get correct classification that is the purpose of explaining this these sort of concept is explain in the book by artificial new network by Sathish Kumar is that the correct name Sathish Kumar's book we will have an explanation of this and actually it comes under the a learning algorithms for a percept trans well in leading towards a discussion about how a percept can be trained and it can provide linear decision boundaries.

We will explore that first and come back to this discriminant function and decision boundaries under the based paradigm may be perhaps in the next class okay if you go back, so what you could do is pick up the sample X_1 and look into this weight space what will you get you will get the corresponding hyper plane H plane depending upon a sample XK let us say this K = 1 I repeat it again let us say K = 1 that means you picked up the first sample and so that first sample will give you.

Some plane in this weight space let us say this plane is this I am not writing the value as this symbol H is so this is some planes line in 2D but for the time way I will always use that from hyper plane or plane CD or high dimensions what you should visualize the graphs are being shown in 2D, so this is geometrical interpretation is what we are looking why this X_1 is similar so if this normal to the hyper plane which is given by the sample X_K, since this plane is being drawn from the sample X_1 the normal to this.

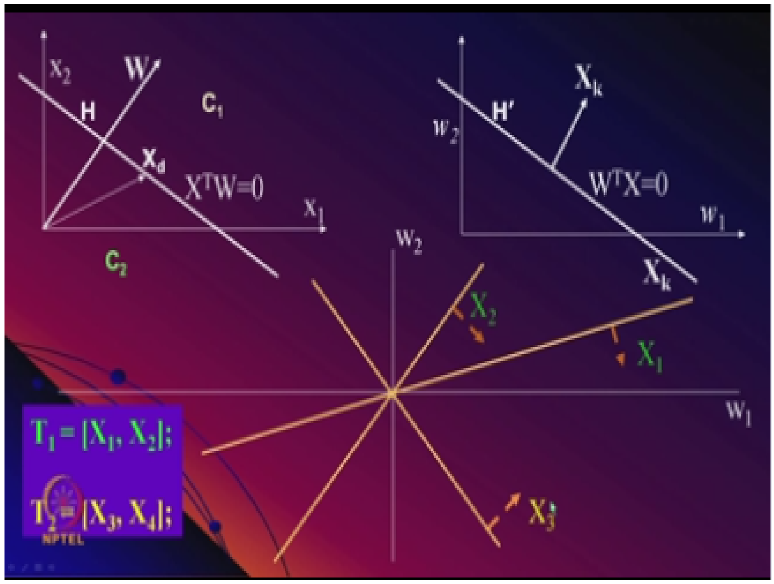
Is actually given by this sample X_1 what we do is pick up the sample X_2 and we will get an another plane so let us say that plane is somewhere here okay so this plane is contributed to by the second sample and that is why the normal to that plane is given by this particular vector we will show that is normal to this plane, so what does this basically tell us in some sense it tell us this positive side of the way vector is shown that it seems that if we choose weights in this region I repeat again the positive side of the plan is given by the direction of the arrow and if we choose weights here you will get the positive value of the G.

A positive value of the G will indicate that you are able to classify the sample correctly using the set of weights which you will choose based on this diagram okay what if you are have negative value of G well then either it is incorrect classification if the sample belongings to the class 1 indicated by the level T_1 in this slide but look at the complementary class or the other class, class 2 indicated by this label T_2 if you pick up samples from those you do not want to discriminant functions G_I to give.

Positive values for those weights we want to have negative values for that remember I am trying to you want to have negative values for that remember I am trying to design a single discriminant function with a certain set of weights I am trying to find out a possible set of values for W which I will decide which will give positive values for class 1 negative values for class 2 so if I pick up samples X_1 and X_2 from class T_1 I should have a positive value of G if I pick up samples X_3 and X_4 form class 2 labeled as T_2 I should have a negative value of G that is my purpose here and to solve that.

What is the appropriate decision boundary which will provide these classifications is the purpose of this representation we are actually not proposing a short of a learning algorithm either for a classifiers statistical classifier and percept and right now but we are trying to give a physical justification of how the weight should be selected so that you can do proper classification and the interpretation of the decision boundary using this a geometrical interpretation using this graph okay.

(Refer Slide Time: 24:00)



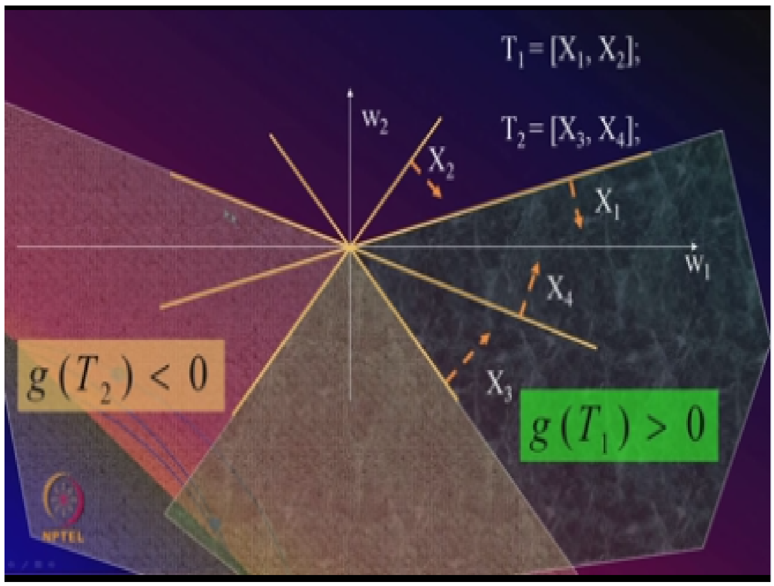
So now so these are the hyper planes corresponding to samples from class 1, we will now select samples from class 2 let us say there are two of these again and draw similar hyper planes nobody can stop us from doing that let us draw one of them, let us say this is the hyper plane corresponding to class x3. Now this vector shows that this is the positive side corresponding to this hyper plane.

What does it indicate, remember the sample x3 belongs to class 2 I must design weights in such a manner that corresponding to samples like x3 I should have a negative value of g, samples belonging to class 1, I repeat that again which are x1, x2 should give a positive value of g samples belonging to class 2 like x3 and x4 should give a negative value that means corresponding to this hyper plane if this is the positive side I should not select weights here in this positive region with respect to this hyper plane if I want a negative value of g for samples belonging to class 2.

If you go back with the similar logic for these two hyper planes belonging to class 1 I should select samples in weights in such a manner that I have a positive value of g corresponding to samples belonging to x2, x2 okay. Now there is another sample left x4 we will draw another one so this is the curve sorry, this is the hyper plane for the sample x4, so if you have more of these samples okay 3, 4 or more you will start getting more and more of these hyper planes.

What will this hyper planes tell us, that please select weights on one side of it width side of it, it should be appropriately chosen in such a manner that weights should actually provide a positive value of g for class 1 and negative values for class 2, I repeat again so corresponding to x1 and x2 I should select weights so that I have positive value of g and corresponding to samples x3 and x4 I should have negative samples. Remember here the samples are just showing the normal to the plane okay, so let us observe this from a side to different angle.

(Refer Slide Time: 26:25)



This is the plot which we have got for the hyper planes, what does it indicate let us take samples from class 1 x1, x2 belonging to class 1 sample here, what are those hyper planes this one corresponding to x1 here this line or this plane and the other one x2. This tells us that you should select weights for which the g of this should become positive okay, what is that region it is the intersection of the positive side with respect to both of this hyper planes which is basically this region.

You can see this textual region in green indicates that if you select weights from here not here not anywhere else, but only in this region if you select weights from this region and substitute into this g then samples belong to class 1 as far as because we have only two samples x1, x2. But it is equivalent, it is equally applicable if you have large number of samples say n of them, the say thing is applicable.

The region will be different, the g(Tn) will be positive I hope this idea is clear, remember if you take this first hyper plane for a sample x1 the semi infinite region is the positive side with respect to the hyper plane which is below this line, for sample x2 it is on the right side let us say this entire region is the positive side.

So you must have an intersection of these two because if you select weights here for sample x1 you will get positive g but not for samples belonging to x2. Similarly here if you select weights you will have a positive value corresponding to sample x2 alright or sample similar to x2 let us say, but not for sample x1 it is guaranteed here that it is in this region you will have a positive value of g if you select weights here, good.

So now if you take up samples on the second class T2 where are those hyper planes well one of them is here corresponding to sample x3, corresponding to sample x4 you have the another one we have drawn them in the previous slide the vector indicating the normal to the plane. Now in the second case for class 2 you have to be a little careful, it is the complementary of class 1, but we are using the same discriminate function g which should actually act something like a decision boundary.

Remember I say right now we are not discriminating we are using the same g and so the g the discriminate function from by selecting a set of weights if it gives positive value for samples belonging to class 1 it should now give for samples from class 2 if I pick up will it give positive or negative, it should give negative values, okay. So I should look at negative region corresponding negative side of the region okay, on negative region to be very precise and correct.

Negative region corresponding to the hyper planes for drawn from samples are obtained from samples belonging to class 2, which are x3 and x4. So let us look at it, so that means with respect to the sample x3 this is the negative region with respect to samples from x4 this is the other negative region. So a similar intersection of these two negative regions will give you another semi infinite region in any dimension.

We are seeing this in 2D you can visualize easily this in 3D and of course in general we have very, very high dimensions depending up the dimension of the feature vector or the pattern space we are talking about. So this is the negative so now what does this indicate on the right side we have g(T1) which is the positive that means samples from class 1 tell us please select weights anywhere here.

Then I will have a correct class assignment from g because it will give you a positive value, the other region on the negative side tells us the g of T2 remember the g is the same that please select samples from this region because if you select samples on the second class I will get a negative value of g which will give you a correct class assignment, same g I want same set of weights two different classes, binary classification as it is called or often referred or it is called a one verses one classification, two classes one and two A and B examples being flowers verses fruit, car verses trucks two different finger prints and faces.

I want to obtain one g which will give a positive value for samples belonging to class 1 or class A and negative value for samples belonging to class 2 or class B, and to do that if you look into the diagram very, very carefully that the intersection of these two spaces is something which is given at the bottom where if you select weights in this well textual area which is as the color codes.

Let us say some where neither this neither this but a sort of intersection of the true this is actually call the solutions space, the solution space for this classification true classes only mind you just to samples and why this is the solution base because it is seems to be the common space only as an inter section of this space g of t2 negative g of t1 positive and the inter section of this is this particular space which is some infinite I am just drawn a line but this is again also it is a mean finite region.

If you select weight there mutually both these constrains will be satisfied for the corresponding class that means if you select samples from class one you will have positive values for g as well as negative values for the samples belong to the other class, if you select anywhere else w\you will get wrong results if you select weight here it will maybe correct for the classification of the samples belong to class 2 but not for t1.

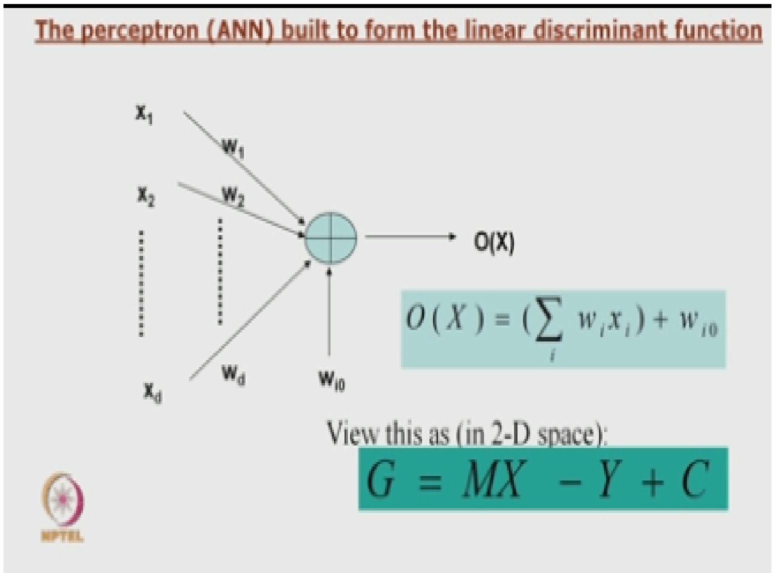
Similarly if you select weights here it will be doing a correct classification for samples belonging to class one but not for 2 and of course no question of selecting samples here so this is the only solution space, I leave it to your imagination that if you have three or more such samples belong the each class you can actually keep on drawing such hipper planes if there are n1 different samples let us say belong to class one you will draw n1 such hipper planes find out that common inter section space g of t1 which is positive then turn to class two let us say you are also given n2 number of samples very lasted of samples draw all those hipper planes find the inter section which will give you g of t2 which is negative that space.

And inter section of these two will actually give you the solution space, you could ask me a question does the solution space always exist, will the solution space exist because if there is no common space shared in the feature space mind you between this g of t1 and g of t2 you will not have it okay we will think of this example okay, in alter class and see for the time mean we look at solutions where you are able to draw a decision boundary with the decision boundary is not very clear here mind it but it is a linear discriminate function each of these are linear discriminate functions each of this hipper planes and you are getting a solution space to it this is not a learning algorithm to learn the weights.

But it tells you gives you an idea of the significant of the weights and the possible solutions space for you to obtain to learn weight that means now if you have the solution space somehow define you can pick up any value in that which will give you the correct result and actually perceptron learns that, perceptron is al linear implementation of a linear discriminate function, its motivation comes from the basic neuron of the human brain in which it receives signals from several other neurons and provide the single output.

And the most simplistic model of such a perceptron is actually a linear discriminate function this is not a course on your network you have to learn many things about that from a separate course but there is a quite significant develop at in terms of applications of new networks for pattern recognition. So will move towards the diagramed perceptron see how that is learned from this solution base given the same set of samples and come back to the discriminate function decision boundary linear case once again?

(Refer Slide Time: 36:09)



Sp as promised in the field of artificial neuron network a perceptron is build to form a linear discriminate function okay, so this let us say a single neuron in your brain which is actually called in the paradime of artificial neural network it is a artificial okay a artificial network is build using a set of perceptron but the single perceptron what it takes, it takes different inputs from a set of components.

So let us say our input is the de-dimension feature vector, x1 x2 up to xd okay do not confuse this d with a distance, this is the d dimension. So there are weights sitting between the input and the perceptron what it basically does these weight is multiplied with x1 w2 is multiplied with x2 w3 with extended on wd is multiplied with xd all of these are summed up at this particular point there is external bias.

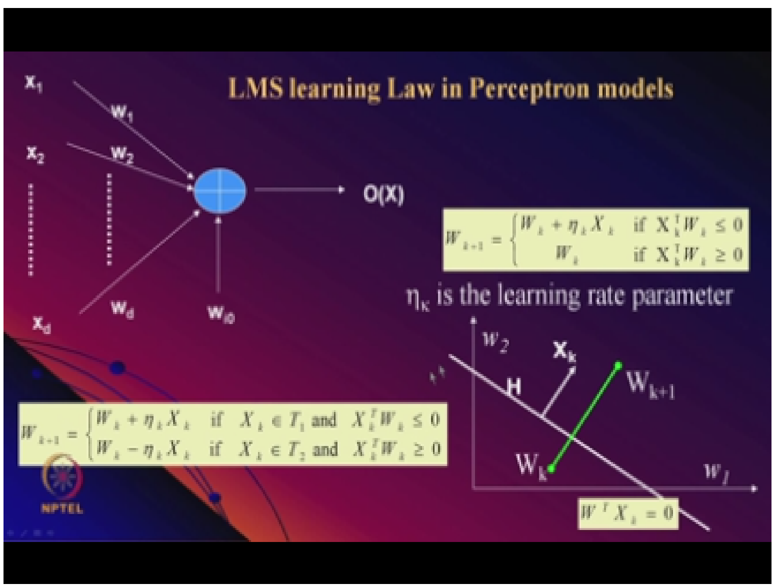
So the output o of x what is this x? Feature vector, with d dimensional feature vector is x so the output of the perceptron can be visualized to implement this. This expression is not in new we have got that in the beginning of our class itself when we talked about the linear discriminate function which we actually got from the morn and a base distance which we got from the morn and a base distance after taking the co variants matrix to be equal to identity matrix.

The same statistical linear discriminate function perceptron also implements that, now the question is the last discussion which we had for the last few minutes of the solutions base the w here you will ask the question in the case of perceptron who decides these weight vector which will go and sit here, how is this result there is a corresponding learning law but can we interpret this process of learning use in the diagram which we have just studied.

And also I would like you to imagine that in the 2d space in very simplistic ANN this ox can be consider to be our g the discriminate function and it gives a simple this is an equation of line everybody knows this, so visualize this to be an equation of line in 2d space to be our linear discriminate function or linear decision boundary.

If you ask me to compact these two early consider that this is just a single scalar quantity c and this I left two component w1 w2 which are m and one, x and y are the two components x1 and x2, this is what we have been doing in the previous slide when we are taking x1 x2 which are equaling to x and y two dimensional space. So we will now move towards what is called I will not derive it completely.

(Refer Slide Time: 38:58)



What is call the least mean square learning law in perceptron models, but we will describe this process of how these weights are assign and the basic idea is something like this. A perceptron learning algorithm starts with the set of random weights w so what you do is you put random values here, and then increment the weight iteratively using this expression well this is not the ultimate expression but look at this what it does, it is actually using the condition here on the right hand side which you are using a little bit earlier in terms of class assignment g of t 1 is positive or negative based on that criteria you where looking at the solution base.

Now what we are doing in studies we are looking at incremental learning of the weight it says that keep the weights and altered if for the sample x provided belongs to the correct class it is giving a positive value if it is giving a negative value well if it is equal you can do either does not matter you are going to increment the old weight using a short of differential term as given as this ε is the learning rate parameter I will tell you what to do with this ε for the timing assume it to be a constant okay.

But it is a function of k that means iteratively when you proceed you are changing this but form the timing assuming it to be constant but what is the significant of this? You are changing the weight remember these are all vector quantities you are adding a vector to another weight vector to get a new weight vector I repeat these was the old vector you start with random weights and incrementally update, that is what the neural network training or the essence of it images and we are trying to pre an analogy with respect to whatever we are learnt in terms of the patterns based for the weight space which we have seen a little bit earlier.

So what it says is now that if the weights which you have point of iteration you do not need to adjust that if it provides correct classification as given here, provided samples for class 1 if it is not of course then it is different. However you want positive values but if the discriminate function g(x) which is given by this expression gives a negative value here on another side of then you change the weight along the direction of the sample vector.

So let us look at the significance of what these means in terms of changing the weight along the direction of the sampling, we will go back to weight space we have the hyper plane H here given by this expression in 2d, 3d or high dimension and the normal to that hyper 10 is given by sample X. if you try to draw the diagram corresponding to this expression what it basically means is if the weight is a as given by Wk then η sk will basically tell you that please change the weight along the direction of the positive side of sk.

And bring it to a point which is given by this Wk 1 and you do this only if you are in the negative zone here in the right space, remember the H is going to partition my feature space into 2 finite regions a positive regions and a negative regions as given here. If you are already here as given by this constraints that means you are on the positive side please do not change your weights keep it Wk1 = Wk here.

However if this conditions of us what does it mean g discriminate function gives the – value here if you are here the weights are not correct you want the positive value of g, so please change w along the direction of x which will take you from negative side towards the positive side, well you may not reach from the negative side to the positive side always in one step jump because if you are far away somewhere here you will probably move towards the positive side but still in negative vision.

Once you are here this constraint will be satisfied you do not change the weights, this is done for one particular sample you will have several set of training samples for both classes and will follow this literally for each sample okay. In general look at this expression now what the 1st line says except that we have put another constrain if the sample belongs to class 1, then if the 2nd condition that means if you are getting the g value positive do not do anything so there is no point writing about it.

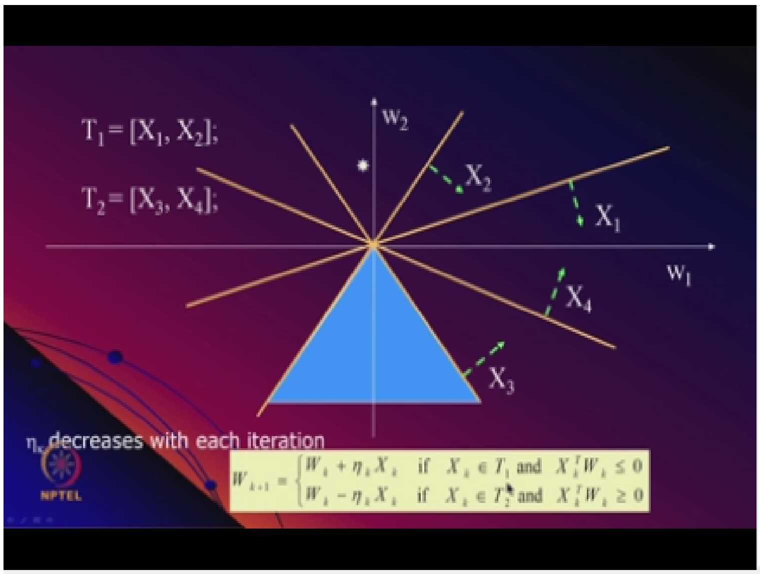
Only the value of g is negative as the 1st constrain is so this is similar to 1st but you are looking at the sample class 1 they need change in the around direction of x. however if you choose sample ground to class 2 then you have look for the complementary nature, if the g is actually positive then you must make a change if it is negative we need not make a change, so what it means here if the sample belongs to class 1 and if the value of g is positive then you do not make a change.

If the sample belongs to class 2 if the value of g is negative then you do not make a change, you do it only when you are in the wrong side, what is the wrong side indicating? 4 samples belonging to calss1 if you get negative value of g then make this change. If samples belongs to class2 and you are on the positive side just the complementary figure think that your Wk is here somewhere, it is on the wrong side, then you need to change it in the negative direction corresponding to the sample x.

If sample is belonging to class 2 okay, so you have to little bit careful here that this is the logic if your sample 1, denominator staple 2 you move it in the reverse direction okay. Now what we will do now is that we will exploit this type of increment learning algorithm with the respect to the previous diagram which we had in couple of slides back where we had the solution space for two class's 2c samples.

And see that if you start with the random value of the weight will you reach the solution space, this diagram as given as static slide in the book by sathish kumar so we will see if this types of incremental learning of the weights helps us to actually obtain a correct value of W and once you get that correct value of W you have designed your classifier.

(Refer Slide Time: 46:40)



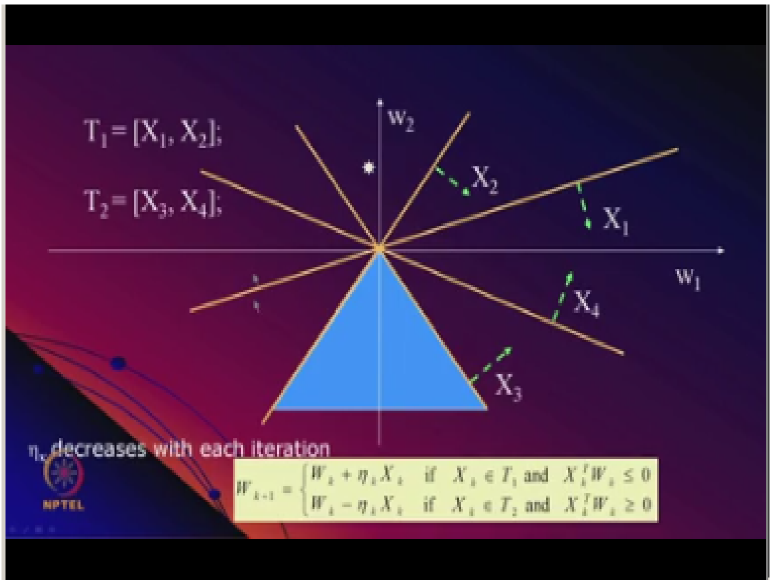
So this is where the both of the things are brought together on the top you have this diagram which shows you that this is the solution space corresponding to the training samples given here which we have worked out, last you want the two samples x1 and x2 you can see the corresponding hyper planes. For class 2 T2 there are 2 samples x3 and x4 for which these two are corresponding hyper planes this one is for x3, and this one is for x4.

And this is now the incremental learning algorithm for learning the weights for the set of training samples, so we are actually looking at the preceptor learning algorithm some sort of a learning algorithm which I will not say the variant of days we had the base classifier which was the statistical classifier working on distributions, we are not going to do that anymore what we are looking at there is just a perceptron giving us a linear discriminate function what should be the weights.

With 2 samples per class it is left to the observer reader or viewer that the number of samples grow very much higher what will be the solutions space and how will this learning algorithm work it terms of weight assignments will it lead us to solution space. What are we doing here look at this, this is the initial range let us say either assigned at random or what we can do is let us say that this is the weight obtained after a few set iteration.

And then in one short this is example may mind you in just one iteration we will see either we are getting closer to the solution space are we even getting into it if possible okay.

(Refer Slide Time: 49:02)



So we want to apply this sequence of steps for these samples what does it basically mean I will pick up sample x1 which will belong to class one tell me will this criteria be satisfied the gx is given by these this is the hyper plane for x1 positive side is here my weight is here will this condition be satisfied yes or no why yes because it is on the negative side of these hyper flame for sometimes so it basically means as you travel along.

The direction which is to the positive side as given by the vector x1 on the normal I will show this moment as a sequence of steps for all this four samples in one shot rather than showing it one of rank and in the visualize then corresponding to sample x2 this is the hyper plane this is the positive side here will this condition be satisfied or not okay so if it is satisfied you will be also moving incrementing.

The weight along this that means you will move the weight along x2 off course you will move it fast along x1 then you move it along x2 so the path of this point will be somewhere as you see my curves moving along x1 then along x2 okay let move towards samples from class 2 there are two samples x3 and x4 so if you look at x3 this is the hyper plane this is the positive side look with respect to w itself right now where ever it is okay.

So this is the positive side now look at samples belonging to class 2 will this condition be satisfied will this condition be satisfied with this sample point with respect to hyper plane the answer is yes so this condition is true for sample beyond to class 2 I have to work on the towards the so it will be pushing it towards the negative side that means it will be negative of x3 so after I move along x1 then along x2.

I will move along negative side of x3 and finally for the sample x4 the same thing happens if the sample is here or somewhere on this region this condition with again they satisfied and I will find move towards the negative side of or the negative direction or towards as pointed by the negative of x4 negative so if you look at the star tic it will move along x1 then along x2 then along x3 and finally on x4.

So if I draw this four strategic approximately peeping the magnitude as shown by the corresponding vectors you will get something like this okay before the strategic you see this is x1 then this x2 then the corresponding two negative this x3 finally x4 okay this is an approximate hand drawn curve not machine computed and this animation shows here once set of interaction for all the samples that is what I say may be you are at the intermediate stage very close to the final solution.

Here you on keep on interacting tell you reach the solutions space because at the solution space none of these conditions if you look back if you are finally here see this is the final position as market by Michael earlier is finally stopped after exploded all the four what will happen at this point for samples beyond to t1 this condition will not be satisfied why because you on the positive side of g with respect to t1 similarly for this point.

Here for samples beyond to class two you are on the negative side of g so this condition will also not be satisfied and you will stop the movement of the vectors but if at any point of interaction you have not reach the solution space one or both of these conditions may be satisfied you keep incrementing the rates as given by the expertise look at their position in one sense for samples belong to class one which is move along the positive direction are x for samples belong to class 2 it says move along the negative direction x correct.

And what happen initially is you start with ransom set of weight you start with random set of weights so it is possible that you are very far away from the solution space the solution space could be somewhere here and the weight that is here you move around you are coming close to the solution space apply the same logic one second and slowly interestingly you will move closer to this solution space which in this diagram if we go back again shows in just one step.

But that will not usually happen number one number two is I have not discussed the significant of which is called the learning rate parameter which decrease with each interaction what is this eta it is sitting here and this learning rate parameter sitting as a coefficient vector x tells us that how much do you want the change will you like to change equal to the magnitude of the vector x well initially yes.

So the eta case starts with the larger value which could be equal to 1 let us say some normalize value of 1 initially but when as the interaction proceeds or the learning algorithms move towards conversions then what will you do is you reduce the value of the learning rate parameter why you need to do that when you are moving towards the solution space you do not want to make large moments.

Because if the solution space is restricted as not the case shown here as I said before if you have large number of samples the solution space we actually may become finite okay and then if you have a large value of it actually jump over the solution space and may not instead of conversion move to the other direction and then you will require larger and larger number of typically sometimes in certain cases of new networks if new networks which is basically called a multi layer perception.

Where what we are learning is single perception you can make a single layer new network by a set of perspective then you can have multiple layers we do not have scope of discussing those in this particular course although I will give you certain expressions of learning in a very general case they are drawn not now which can implement even linear or normal boundaries but let us stake the case of perception for this simple example others are extensions of this you need to reduce the learning rate parameter.

It is actually reduced by a exponential form initially with the large value you bring it down to the smaller value why you do so because when you are approaching to the solution space you want to move such little bit closer to that instead of jumping with the larger value initially when you start with the ransom set of values you want to move with large end like.

The one which is showed here these sort of exhausted moments may initially happen towards the end you will move incremental manner and that is why interpreted learning in vector and parameter brought down off course what may happen is if you are that lucky that you are already there very close to the solution space.

I mean one or two jumps or one or two interaction as it is called you lead the solution space find then what will happen this in equality will not be satisfied you will be not allowed to the system will not move the algorithm will not move the rates along it will get struck in this solution space

Online Video Editing /Post Production

K.R.Mahendra Babu
Soju Francis
S.Pradeepa
S.Subash

Camera

Selvam
Robert Joseph
Karthikeyan
Ram Kumar
Ramganesh
Sathiaraj

Studio Assistants

Krishankumar
Linuselvan
Saranraj

Animations

Anushree Santhosh
Pradeep Valan .S.L

NPTEL Web &Faculty Assistance Team

Allen Jacob Dinesh
Bharathi Balaji
Deepa Venkatraman
Dianis Bertin
Gayathri

Gurumoorthi
Jason Prasad
Jayanthi
Kamala Ramakrishnan
Lakshmi Priya
Malarvizhi
Manikandasivam
Mohana Sundari
Muthu Kumaran
Naveen Kumar
Palani
Salomi
Senthil
Sridharan
Suriyakumari

Administrative Assistant

Janakiraman.K.S

Video Producers

K.R. Ravindranath
Kannan Krishnamurty

IIT Madras Production

Funded By
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in