Pattern Recognition and Application
Prof. P. K. Biswas
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Khargpur

Lecture - 23
Liner Discrimination
(Tutorial)

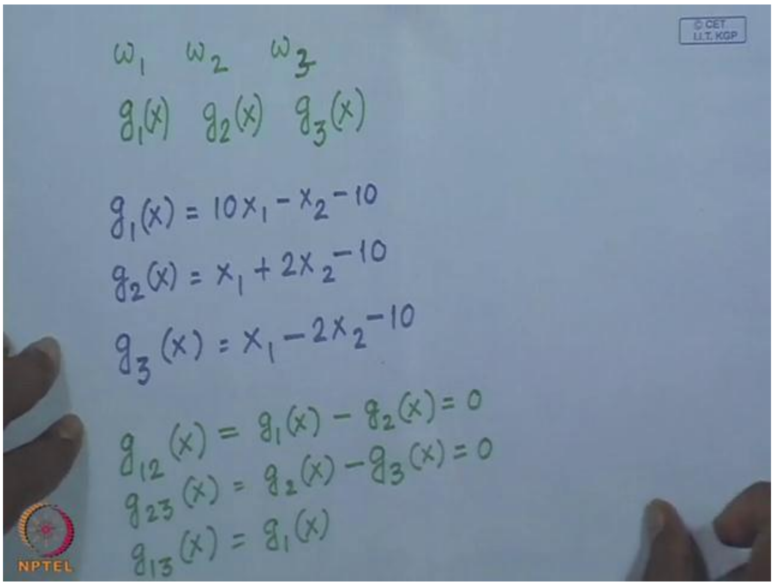
Hello. So, in the last few classes, we have talked about different types of linear discriminators and different types of techniques for designing of linear discriminators. So, we have talked about the statistical model that is starting from the probability distribution of the samples belonging to define classes. How we can design a classifier using a base decision theory? We have also talked about other type of designing techniques for linear discriminators when we have a set of samples, which are given for training of the classifier or design of the linear discriminator. One of them we have discussed is use of perception criteria.

We have also talked about the relaxation criteria for designing of linear discriminators. We have also talked about the mean square error criteria for designing of linear discriminators. Then various such techniques we have discussed in last few classes. We have also talked about a linear machine. What we have said that every class has a discriminate function and for given unknown sample for where when you design the discriminate function, you make use of the samples for which the class belongingness is known so effectively what we are trying to do is we are trying to implement supervised learning technique.

That means training of the classifier is supervised in the sense the classifiers are trained using some samples for which the classes are known. So, you have also talked about a linear machine for every class has a discriminant function. So, for a given unknown sample, what we try to do is we compute the discriminant functional value for that unknown sample for every class. So, whichever class gives maximum value, the unknown sample is classified to that particular class.

So, in today’s class, what we will do is we will take few problems. We will take few problems and try to solve those problems. So, your ideas of designing of such classifiers become clearer. So, let us take the first problem where we have the discriminant functions for every class. From those discriminant functions from that set of discriminant functions, we have to find out what is the class boundary between different classes, and also what is the region in the feature space, which is allotted to different classes. So, we will take the first problem.

(Refer Slide Time: 03:43)



I assume that that I have got 3 classes say the classes are omega 1, omega 2 and omega 3. So, I have got these 3 different classes. For every class, I have a discriminant function that means for class omega 1; I have a discriminant function, which is given by g 1 of X. For omega 2, I have a discriminant function, which is given by g 2 of X and for omega 3, I have a discriminant function, which is given by g 4 of X. So, these are the discriminant function for 3 different classes. For omega 1, I have g 1 X. For omega 2, I have g 2 X. For omega 3, I have g 3 X.

So, naturally our classification is if I find that g 1 X greater g 2 X, then X will be classified to class omega 1. If g 3 X is greater than g 1 X and g 3 X is greater than g 2 X, then X will be classified to class omega 3. If g 2 X is greater g 1 X and g 2 X is greater than g 3 X, then X will be classified to class omega 2. Now, let us see that what are the different functional forms each of these linear discriminating functions have. So, I will put say g 1 X is equal to let us say 10 X 1 minus X 2 minus 10 g 2 X, let us take X 1 plus 2 X 2 minus 10. Suppose that g 3 X is equal to X 1 minus 2 X 2 minus 10, so these are the discriminating functions linear discriminating functions for different classes.

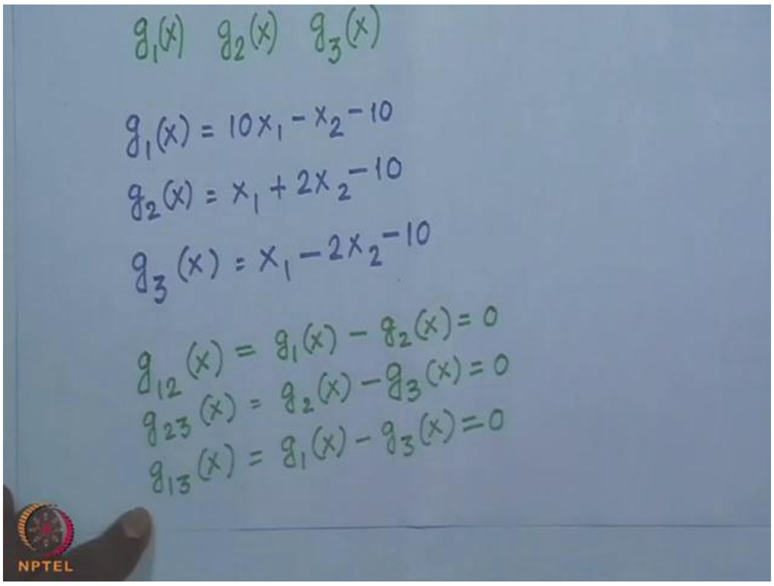
So, in order to classify a sample to class omega 1, what I have to do is I have to find out what is the value of g 1 X. What is the value of g 2 X and I have to find out what is the value of g 3 X. If I find that g 1 X is greater than g 2 X, then obviously between these 2 classes, X belongs to omega 1. If g 1 is greater than g 3 X, then between these 2 classes, X also belongs to class omega 1.

So, undoubtedly, the unknown feature vector X will be classified to class omega 1. Now, if it is so happens that g 1 X is greater than g 2 X, but g 3 X is greater than g 1 X, so those are the different such confusing cases. We will see what happens to all these different confusing cases. So, here what happen I will do is this is the discriminating function for class omega 1. This is the discriminating function for class omega 2. This is the discriminating function for class omega 3.

Now, what I want to do is I want to find out what is the decision boundary between pair of classes say between g 1 omega 1 and omega 2. What is the decision boundary between omega 2 and 3? What is the decision boundary and between omega 1 and omega 3? What is the decision boundary? So, for that, what I need to compute is g 1 2 of X. g 1 2 of X will be given by g 1 X minus g 2 X. If this g 1 2 X minus g 2 X, if this is positive or g 1 2 X is positive, then obviously X belongs to class omega 1.

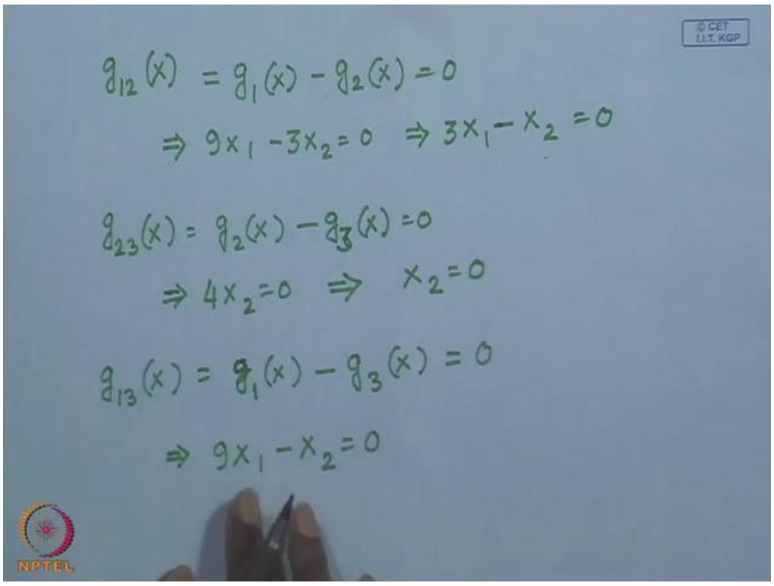
When I compare between classes omega 1 and omega 2, so that decision boundary between the classes omega 1 and omega 2 will be given by g 1 2 X is equal to 0 or g 1 X minus X 2, which is equal to 0. Similarly, I find out what is g 2 3 X that is the decision boundary between class omega 1 and class omega 2. So, I will compute g 2 3 X as g 2 X minus g 3 X. I will equate this to 0 to get the decision boundary between class omega 1 and class between class omega 3 and class omega 2. Similarly, I will compute what is g of 1 3 X.

(Refer Slide Time: 09:43)



It is nothing but g 1 X minus g 3 X. By equating this to 0, I get the decision boundary between class omega 1 and class omega 3. Now, let us see that what each of these functions g 1 2 X, g 2 3 X or g 1 3 X take using these discriminating functional values g 1 X, g 2 X and g 3 X.

(Refer Slide Time: 10:20)

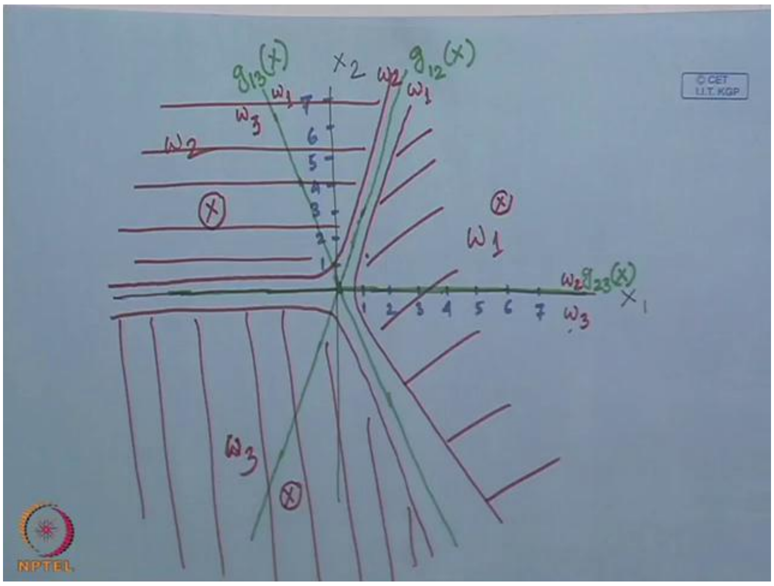


So, from here, you find that my g 1 2 X is nothing but g 1 X minus g 2 X. See if I compute this g 1 X minus g 2 X. That will be 9 X 1 minus 3 X 2. So, this is equal to g 1 X minus g 2 X, which is equal to 0. If I compute this, then as I said that it will be 9 X 1 minus 3 X 2 is equal to 0. It is nothing but X 1 3 X 1 minus X 2 is equal to 0. Similarly, when I compute g 2 3 X, it is nothing but g 2 X minus g 3 X. So, g 2 3 X is g 2 X and this skips g 2 X minus g 3 X. So, this X 1 term gets cancelled and it simply become 4 X 2 is equal to 0.

So, that gives me 4 X 2 is equal to of which is nothing but X 2 is equal to 0. X 2 is equal to X is nothing but X 1 X. Similarly, when I compute g 1 3 X is g 1 X minus g 3 X, so when I do that g 1 X is 10 X 1 minus X 2 minus 10 minus g 3 X is X 1 minus 2 X 2 minus 10. So, when I subtract g 3 X minus g 1 X, what I get is 9 X 1 plus X 2. So, this is 9 X 1. So, this is my g 1 X minus g 3 X. I have to equate this to 0, which gives me 9 X 1 minus X 2 that is equal to 0. So, find the decision boundaries between the pair of glasses.

The decision boundary between omega 2 and omega 1 is given by 2 cross 1 minus X 2 is equal to 0 2 X 1 minus the decision boundary between omega 1 and omega 3. It is given by X 2 equal to 0. It is nothing but X 1 X. g 1 3 X is the decision boundary between class omega 1 and omega 3, which is given by 9 X 1 minus X 2 equal to 0. So, these are the different decision boundaries and what we are talking about. We are talking about; we are considering 2 dimension feature vectors having components X 1 and X 2. So, it is a 2 dimensional space. Now, I plot all these decision boundaries. Let us see what kind of situation that we have the first decision boundary.

(Refer Slide Time: 14:24)



So, I will put it this way. This is say my X 1 X axis. This is my X 2 axis. Now, when you come to this 1 g 1 2 X that is the decision boundary between class omega 1 and 2, which is given by 3 X 1 minus X 2 is equal to 0. So, this decision boundary, I plot the decision boundary. It will look like this.

So, I will put it this way say suppose here I have 1, 2, 3, 4, 5, 6, 7, 1, 2, 3, 4, 5, 6, 7 and so on. So, this g 1 2 X, which is 3 X 1 minus X 2 that will be somewhere here. This is g 1 2 of X coming to this g 2 3 of X, which is X 2 equal to 0 that is nothing but my X 1 axis. So, this X 1 axis, this is g 2 3 of X. That is the decision boundary between the classes, omega 2 and omega 3 and coming to the next. Here, I have made a mistake my discriminating functions g 1 X was 10 X 1 minus X 2 minus 10, g 3 X was X 1 minus 2 X 2 minus 10. So, g 1 3 X which g 1 X minus g 3 X will be 9 X 1 plus X 2, which is equal to 0.

So, this is 9 X 1 plus X 2 not minus. So, if I plot this decision boundary on the same graph, you find that this plot will be something like this, but this is g of 1 3 X. now, when it comes to g 1 2 X, we said that if g 1 2 X is greater than 0, then X belongs to class omega 1. If it is less than 0, then X belongs to class omega 2. Similarly, g 2 3 X, if g 2 3 X or a given X is greater than 0, then that X belongs to class omega 2. If it is less than 0, then that X belongs to class omega 3. Similarly, considering this g 1 3 X if g 1 3 X is greater than 0, then that X belongs to class omega. If it is less than 0, then that X belongs to class omega 3.

So, let us see that what the positive side of these different decision boundaries is. If I take upon a point say 1, 1 and put that into g 1 2 X, so over here g 1 2 X, if I put X 1 equal to 1 X 2 equal to 1, then g 1 2 X is equal to 2. That means this point 1, 1, which is somewhere over here is on the positive side of g 1 2 X. So, this g 1 2 X divides this 2 dimensional space into 2 half space. This side of the space is positive and the other side is negative. This means that when I go or classification, this side of the decision boundary is omega 1 and this side of the decision boundary is omega 2.

Similarly, when I consider this g 2 3 X as g 2 3 X, it is nothing but X 2. So, whenever X 2 is positive that side belongs to omega 2. Whenever X 2 is negative, then that side belongs to omega 3. So, considering this decision boundary, this side belongs to omega

2, this side belongs to omega 3. Similarly, when you consider this 1 3 g 1 3 X, g 1 3 X is nothing but 9 X 1 plus X 2.

So, if I said X 1 equal to 1 X 2 equal to 1, this is nothing but 10, which is positive. So, coming to this g 1 3 X, this side of the decision boundary gives me class omega 1. This side of the decision boundary gives me class omega 3. Now, let us try to combine what we get from these 3 decision boundaries. So, coming to this one between omega 1 and omega 2, this side belongs to omega 1 between omega 3 and omega 1 this side belongs to omega 1. So, you find that so far, as this decision boundary is concerned between class omega 1 and omega 2, this half belongs to class omega 1. Considering omega 1 and omega 3, this half belongs to omega 1.

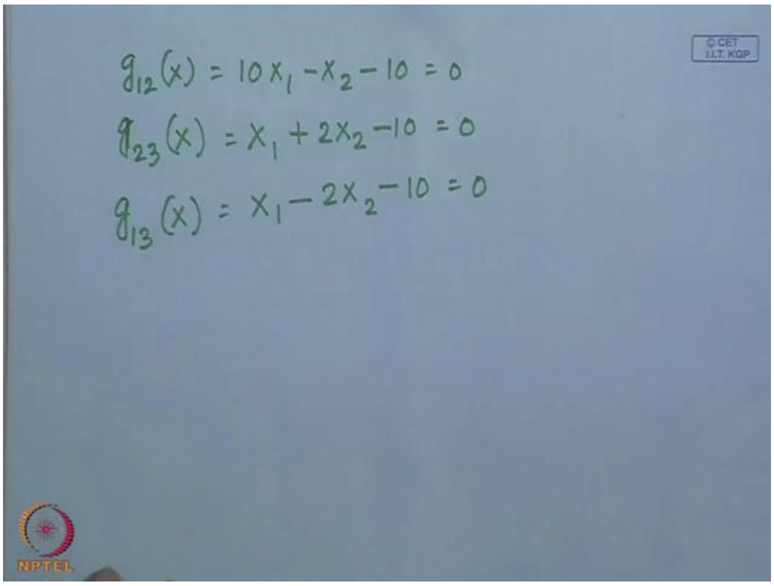
So, finally, the region which will be allocated to class omega 1 is the intersection of these 2 regions. It is nothing but this region which is allocated to class omega 1. Then when I consider omega 2, you find that between omega 1 and omega 2, this half belongs to class omega 2. Between omega 3 and omega 2, it is this half which belongs to class; it is upper half omega 2 and omega 3. It is the upper half which belongs to class omega 2. So, g 1 X tells me that this half goes to omega and g 2 3 omega tells me that this half goes to omega 2.

So, finally, the region which is allotted to omega 2 is nothing but this region. So, this is the region, which goes to class omega 2 and coming to omega 3. This just g 2 3 X tells me that this lower half goes to omega 3 and g 1 3 X gives me tells me that it is this half which goes to omega 3. The intersection of this 2 is nothing but this region, so this is the region which is going to class omega 3.

So, when I have these decision boundaries g 1 3 X, g 1 2 X and g 2 3 X, I have and unknown feature vector, which is lying within this region. Obviously, this unknown feature vector will be classified to class omega 1. If I have an unknown feature vector which is in this region, this will be classified to class omega 2. I have an unknown feature vector, which is here. This unknown feature vector will be classified to class omega 3.

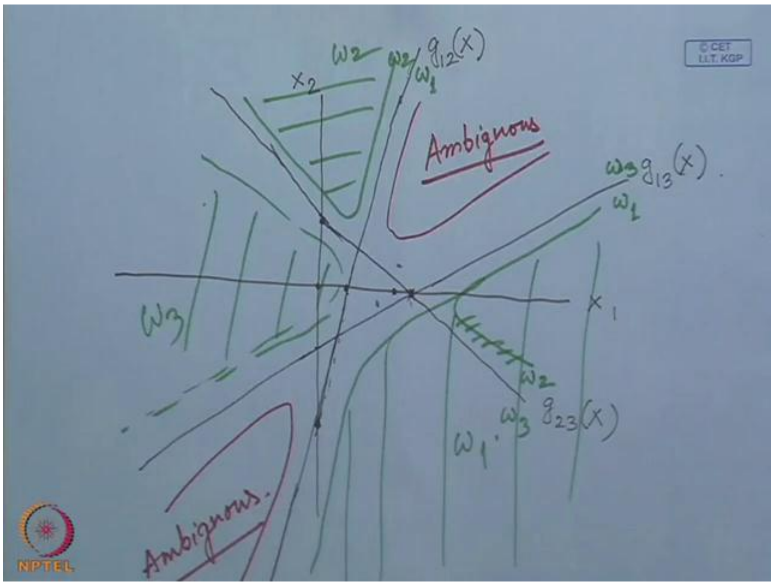
So, from these different discriminating functions, when I get the decision boundaries, the decision boundaries divide the spaces into 3 different sub spaces. Every sub space is allotted to 1 of the classes. So, this was very simple case. Now, let us think of a situation that if these discriminating functions themselves are taken as the decision boundaries. In other words, what I mean to say is given.

(Refer Slide Time: 24:51)



If I take say g 1 2 X is equal to 10 X 1 minus X 2 minus 10 is equal to 0 g 2 3 X is equal to X 1 plus 2 X 2 minus 10 is equal to 0 and say g 1 3 X which is X 1 minus 2 X 2 minus 10 is equal to 0. If I have a situation something like this, then let us see how the different regions will be allotted to different classes. So, what I have to do is I have to again draw the straight lines represented by these 3 different functions. So, let us draw those straight lines.

(Refer Slide Time: 26:01)



So, I have again this axis. This is X 1, this is X 2 here. Find that this particular straight line 10 X 1 minus X 2 is equal to 0. See if I put X 1 is equal to 0, X 2 will be equal to minus 10. So, that will be point somewhere, let us say over here. If I put X 2 is equal to 0, X 1 will be equal to plus 1. So, that will be point somewhere over here.

So, I have this g 1 2 X in this case, which is given by this straight line. This is g 1 2 X. Coming to the next one, g 2 3 X, I put X 1 is equal to 0. X 2 will be equal to what? So, it is 0, 5, which will be a point say somewhere over here. If I put X 2 is equal to 0, X 1 will be equal to that will be a point somewhere over here. I put X 2 equal to 0, and then X 1 is equal to 10. So, that is the point somewhere over here. If I put X 1 equal to 0, X 2 will be equal to 5. So, that will be a point somewhere over here.

So, this is a straight line, which represents g 2 3 of X. Coming to the third one, which is g 1 3 X, which is nothing but X 1 minus 2 X 2 minus 10 is equal to 0. So, here you find that if I put X 1 is equal to 0, X 2 will be equal to minus 5. So, that will be a point somewhere over here. If I put X 2 is equal to 0, X 1 will be equal to 10. So, that is this point itself.

So, I get a straight line, which is this. This straight line represents g 1 3 X. now, let us put the point 0, 0 on each of this too each of this discriminating functions. So, for 0, 0, you find that g 1 2 X is equal to 0 g 2 3 X will also be, g 1 2 x will be negative, g 2 3 will be negative, g 1 3 X that will also be negative. So, that means this origin lies on the negative side of each of these planes.

So, as the origin lies on the negative side that indicates that considering this g 1 2 X, this side of g 1 2 X is omega 1. This is because on this side, I have g 1 2 X is equal to positive. On this side, it is omega 2. Coming to 1 3 X as this is on the negative side, so g 1 3 X will be positive on the other side. So, this side will be your omega 1. This side will represent omega 3. Coming to this g 2 3 X, this side will represent omega 2. This side will represent omega 3. Coming to this g 2 3 x, this side will represent omega 2. This side will represent omega 3.

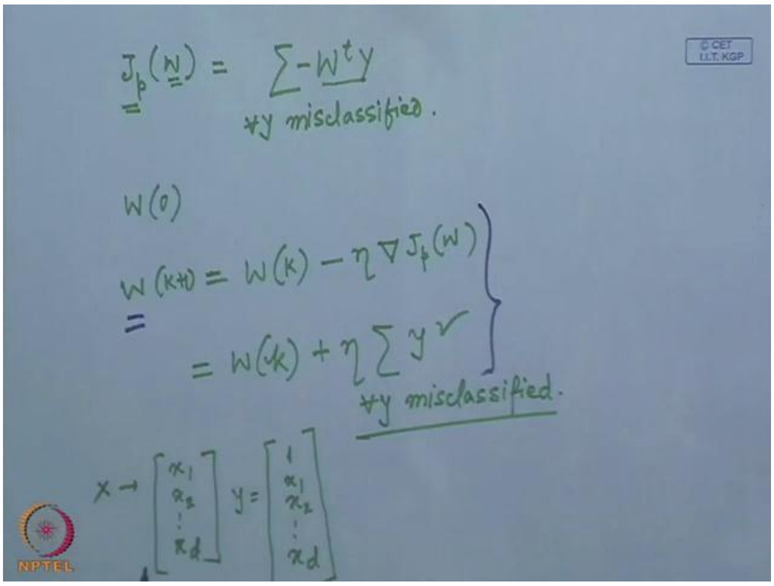
So, with this now, let us try to distribute or alert different regions to different classes. So, here, you will find between class omega 1 and omega 2, this side is omega 1. Between omega 3 and omega 1, this side is omega 1. So, when I take the intersection of this and this, you find that this is the region, which will be allotted to class omega 1; this is the region that will be allotted to class omega 1. Similarly, coming to omega 3, this side between omega 1 and omega 3, this side is omega 3 and between omega 2 and omega 3. So, coming over here, this side goes to omega 3. Between W 2 and W 3, this side goes to W 3.

So, I will have this particular region, which will be allotted to W 3. Coming to W 2, you find that this is the region, which is allotted to W 2. Between W 2 and W 3, this is the region on this side that is allocated to W 2. So, if I take the intersection, then this is the region which will be allocated to W 2. So, I have this region allotted to W1. I have this region allotted to W 2. I have this region allotted to W 3. As I do that, you find that there is a region over here and there is a region over here. These 2 regions cannot be allotted to any of the classes.

So, I have some empty region or these are called ambiguous regions. So, I have this linear discriminator considering the decision boundaries between different classes. If I have a multiclass problem, it is quite possible that in the feature space, I can have few sub spaces. If a feature vector falls that feature vector cannot be classified to any of the classes because those sub spaces are actually ambiguous regions.

So, this is one of the problems that I wanted to discuss to see that how you can solve or how you can identify the regions belonging to different classes when you have linear discriminating functions. So, the second problem that I will try to discuss today is on classified design using perception criteria. So, if you will remember what we said is in case of perception criteria, the criteria function for designing of the classifier was something like this.

(Refer Slide Time: 34:08)



It was J p of say W was given by minus W transpose Y where this summation is taken over all Y V N, which are misclassified. To get the value of W that minimizes this criteria function the perception criteria function, we have used the gradient dissent procedure. The gradient dissent procedure was something like this that initially you assume or you set an weight vector, which is set at random.

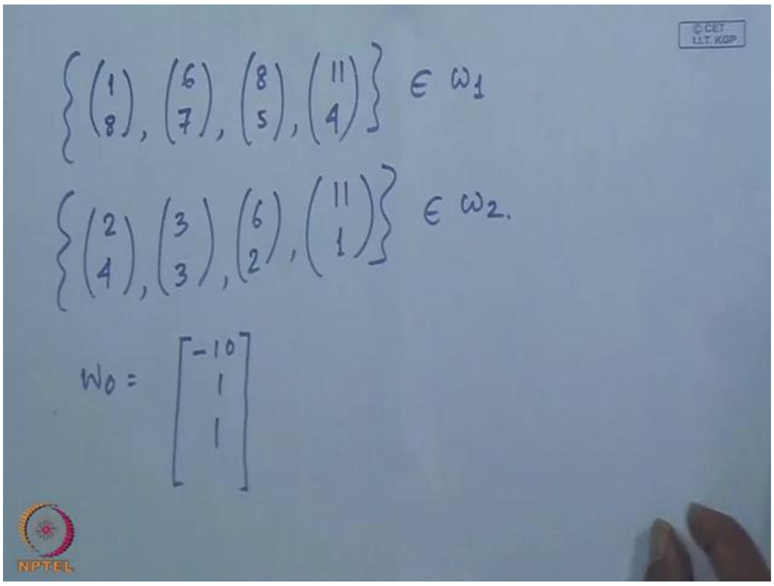
So, initially you assume a weight vector W naught. Then in every iterate iterative step, you set say W K plus 1 in the k theta region, which comes from W K. That is your weight vector at the previous iteration. We are following gradient dissent approach. So, it will be minus eta times grad of J p W where J p W is given like this. If I take gradient of J p W, it will be nothing but minus summation of Y; for all the samples Y which are misclassified.

So, effectively this algorithm comes to W K plus eta times sum of Y for all Y, which are misclassified. How do I get this Y if my original given feature vector is X, I append 1 to this X? So, if this X is say x 1, x 2 up to x d, I get Y by appending 1 to these vectors. So, my Y becomes 1 x 1, x 2 up to x d. Then what we do is if this Y belongs to class say omega 1 Y remains as it is.

If Y belongs to class W 2, then we negate Y, so that whenever a sample is correctly classified, this term is always positive. W transpose Y is always positive if the sample is correctly classified, whereas if the sample is not correctly classified, then W transpose Y will be negative making this minus W transpose Y to be positive. These are the iterative steps, which have to be performed unless the algorithm converges. That means unless I get a W weight vector, which classifies all the tending samples correctly, so this was our perception criteria for design of linear classifiers. This is particularly the one, which is applicable for auto class problem.

So, we are considering 2class problem. Then we have extended this concept to a multi class problem by using a construction technique, which is called keshclark's construction. So, I will consider auto class problem. I will take a problem from auto class case and try to see the different steps of the algorithm how your vector W evolves. Of course, I will not try to solve the entire problem. I will not try to get the solution. I will simply explain the steps. The reason is that it takes a large number of iterations. So, I get a situation something like this.

(Refer Slide Time: 38:40)

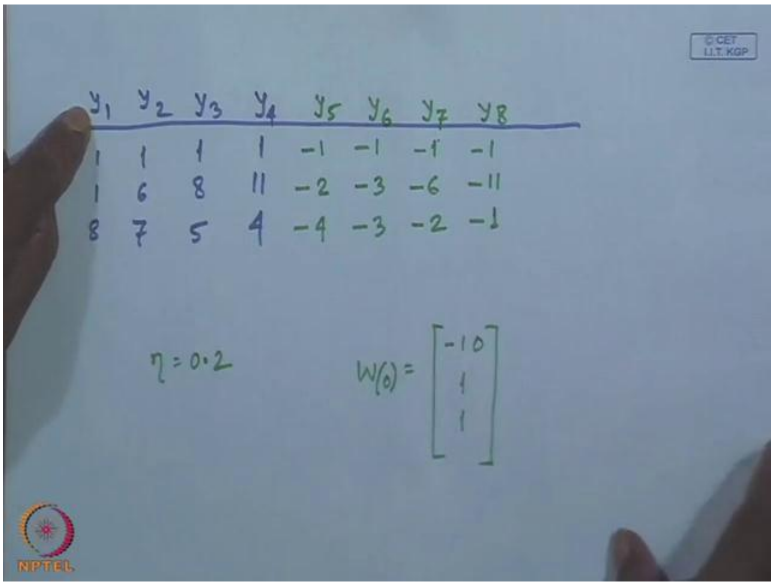


Suppose that you have been given a set of feature vectors say 1, 8, 2 dimensional feature vectors. Then it is 6, 7, 8, 5, 11, and 4. So, this is the set of feature vectors, which are taken from class omega 1. I have another set of feature vectors say 2, 4, maybe say 3, 3, maybe 6 to and suppose 11, 1. These are the feature vectors, which are taken from class omega 2.

So, I have these 2 set of feature vectors, which are my training vectors to train the classifier. So, it is a supervised planning. I also assume that initial weight vector, which we said that the initial weight vector is taken at random. So, I assume that my initial weight vector W 0 is something like minus 10, 1, and 1.

So, this is my initial weight vector. So, considering these vectors to be the vector sets, I have to compute Y out of this vector. So, as we said that we have to append 1 to each of these vectors, after appending 1 all the vectors, which are coming from class omega 2 they have to be negated.

(Refer Slide Time: 40:21)



So, I will get Y 1. I will put it like this. I will get Y 1 which is nothing but 1, 1, eight because I have to append 1 over here. So, it becomes 1, 1, and 8. Y 2 is 1, 6, and 7. So, I will get 1, 6, and 7. Y 3 is 1, 8, and 5. So, Y 3 will be 1, 8, and 5. Y 4 is 11, 4. So, I will have Y 4, which is 1, 11, and 4. So, these are the appended samples or modified samples from class omega 1. In the same manner, I have to get the appended sampler or modified samples from class omega 2. So, I will take these samples from class omega 2. As I have said that after appending 1, these vectors are to be negated. So, my Y 4, Y 5 will be this is 2, 3.

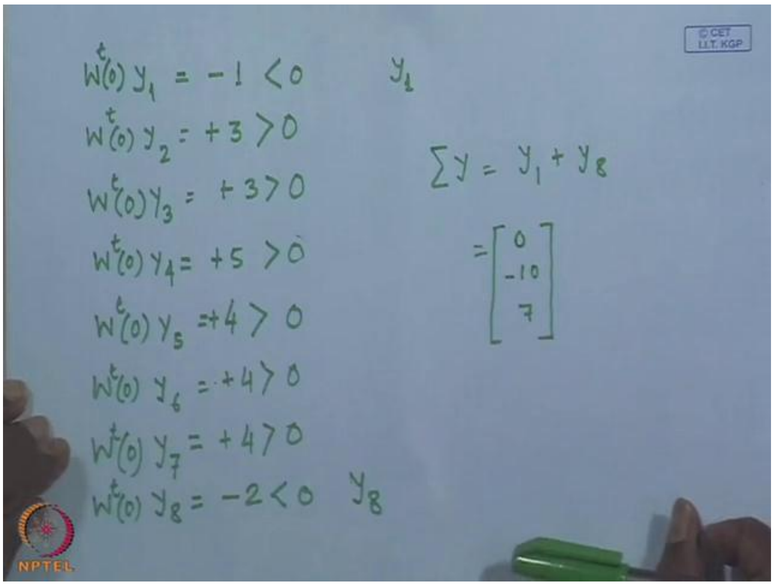
So, I will have 1, 2, 3 and this is 2, 4. So, I will have 1, 2, and 4. Then I will have to negate it. So, it will be minus 1, minus 2, minus 4. So, Y 5 will be minus 1, minus 2, minus 4 in the same manner Y 6, X 6 is 3, 3. So, Y 6 will be minus 1, minus 3, minus 3. Y 7 is 6, 2. So, Y 7 will be minus 1, minus 6, minus 2. Y 8 X 8 was 11, 1. So, Y 8 minus will be minus 1, minus 11, minus 1. So, these are all my modified samples. I also take where this if you look at this algorithm, this eta indicates the rate of convergence.

Let me take this state of convergence eta is equal to say 0.2. As I said my initial weight vector W 0 is minus 10, 1, and 1. So, how the algorithm has to work? I have to take W transpose Y 1 for each of these Y 1s. Then for the Y 1 for each of these Y’s, then for that sample Y, if W transpose Y is positive that sample is correctly classified by this weight vector. If W transpose Y is negative for any of the Y, then that sample Y is misclassified by this weight vector.

So, I have to identify all those samples, Y which are misclassified for which W transpose Y is less than 0. So, all those are misclassified samples are to be used in my gradient dissent procedure for modification of the weight vector W. So, let us see what values of W transpose Y that I get for different samples of Y. So, if I take, you find Y 1. The first component is 1, the second component is 2, and the third component is 8. So, clearly here it is W 0 at the 0th iteration that is the initial value.

So, W 0 transpose Y 1 will be minus 10, plus 1, plus 8. It is nothing but minus 10, plus 1, plus 8. It is nothing but minus that is negative that means Y 1 is misclassified by this W 0 coming to Y 2. W 0 transpose 1 become minus 10, plus 1, plus 7. It is nothing but 3 that is greater than 0. So, W 2 is correctly classified by this sample. So, I do not have to consider W 2 for modification of W. I definitely have 2. Consider Y 1 for modification of W. So, let us see that what this defines values that we get W 0.

(Refer Slide Time: 45:18)

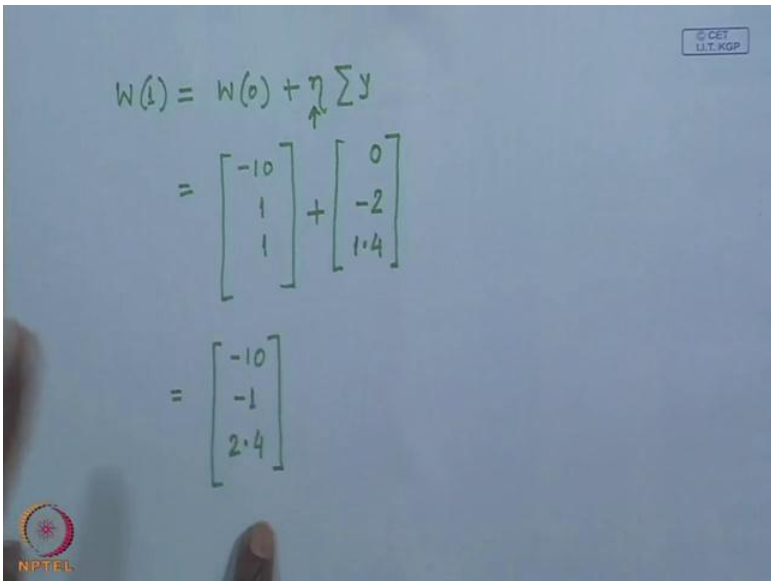


I will get W 0 transpose. If I compute this W 0 transpose Y 1, this has values minus 1 which is less than 0. So, I have to consider Y 1 for modification of my weight vector. When I consider W 0 transpose Y 2, W 0 transpose Y 2 is plus 3 as I just seen which is greater than 0. So, I do not consider Y 2 or modification of the weight vector. Similarly, W 0 transpose Y 3 is also plus 3 greater than 0. I do not have to consider Y 3 medication W 0 transpose 4. It will come out to be plus 5. That is again greater than 0. W 0 transpose Y 5 will be 4. I compute this will again be greater than 0.

W 0 transpose Y 6 will again be plus 4 greater than 0. W 0 transpose Y 7, if you compute, it will be 4 again, which is greater than 0. W 0 transpose Y 8, if you compute this; it will be minus 2, which is less than 0. That means I have to consider this Y 8 for modification of the weight vector W. now, looking at this algorithm, you find that weight of this additional algorithm is W K plus eta times summation of Y is misclassified. So, over here, it is Y 1 and Y 8, which are misclassified.

So, sum of Y of the misclassified samples is nothing but Y 1 plus Y 8. You look at this. A modified sample Y 1 is 1, 1, 8 and Y 8 is minus 1, minus 11, minus 1. So, if I add this to what I get is 0, minus 10, 7. So, this sum of Y is nothing but 0, minus 10 and 7. This is sum of Y, which is misclassified. Now, what I have to compute is by using this, I have to compute the next iterated value of W. It is equal to W 1 because I had W 0. I had chosen W 0.

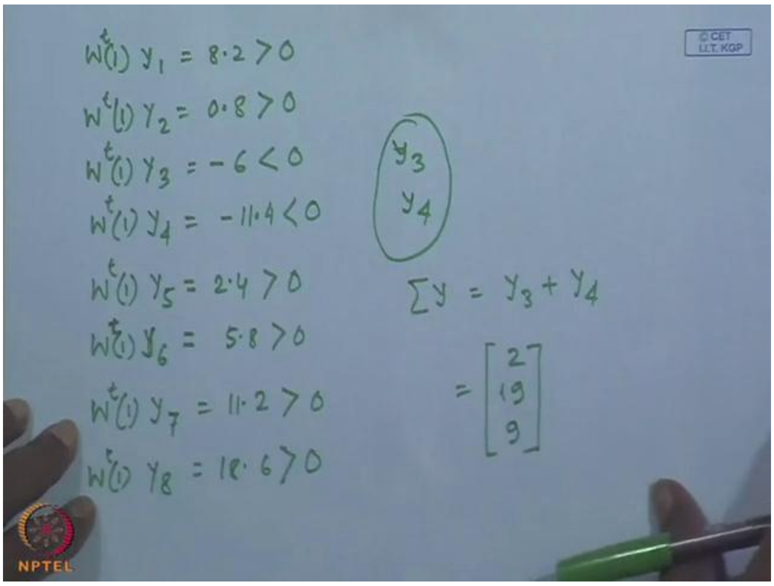
(Refer Slide Time: 48:46)



So, my W 1 becomes W 0 plus eta times sum of Y for misclassified samples where eta is we have assumed eta as 0.2. So, this W 1 simply becomes what was my W 0? W 0 was if you look at here, it was minus 10, 1, 1 plus eta is 0.2 times sum of Y which is 0, minus ten seven so eta times sum of Y will be 0 minus ten into point 2 that is minus point 2 at 7 into 0. 2, is 1.4. So, W 1 simply becomes minus 10, minus 1 and 2.4.

So, you see that how we have modified the weight vector W from W 0 to W 1 using W 0. I have identified all the feature vectors, all the samples Y which are misclassified. Those misclassified samples are used to modify W 0 to get W 1. Now, as I have this W 1, I have to try to classify those entire samples W Y 1 to Y 8 using this W 1. So, I have to compute W 1 transpose Y 1. I have to compute W 1 transpose Y 2, W 1 transpose Y 3 up to W 1 transpose Y 8. I have to identify all those S W from which W 1 transpose Y is negative. These are the samples. These are the Ys, which will be used to update W to get W 2.

(Refer Slide Time: 51:25)

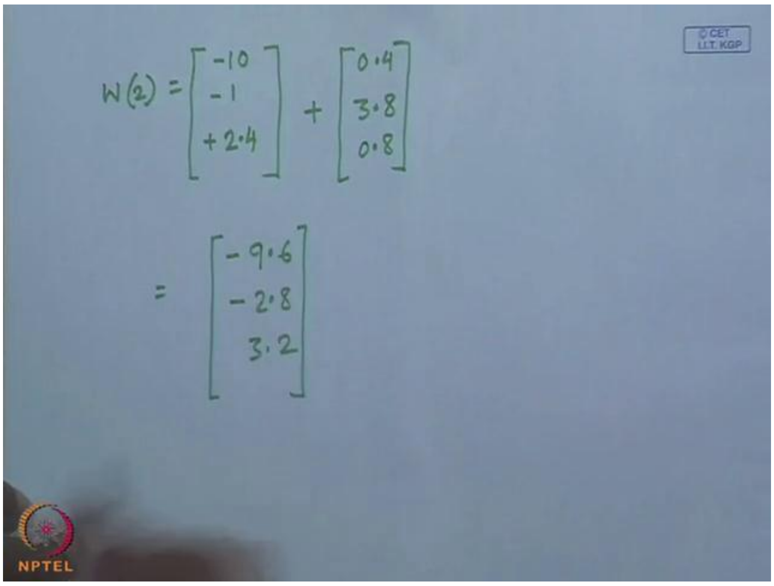


So, if I compute this, you will get that W Y W 1 transpose Y 1 that will be 8.2, which is positive W 1 transpose Y 2, which will be 0.8. It is also positive. W 1 transpose Y 3 that will be minus 6, which is negative that means this Y 3 has to be used or modification of W.

Similarly, it is W 1 transpose Y 4, if I compute it will be minus 11.4, which is again negative. So, Y 4 has to be considered for weight updating for next iteration. W 1 transpose Y 5 will be equal to 2.4. That is greater than 0. W 1 transpose Y 6 will be 5.8. Again, it is greater than 0. W 1 transpose Y 7 will be 11.2 which is again greater than 0. W 1 transpose Y8 will be 18.6. Again, it is greater than 0. So, I have to consider this Y 3 and Y 4 to get my W 2 that is the weight vector to be used at the next iteration.

if I add this Y 3 and Y 4 some of Y is nothing but Y 3 plus Y 4, if you find that y 3 is this 1, 8, 5 and Y 4, 1, 11, 4, I add this 2, I get 2, 9, 2 299. That will be equal to 2, 19 and 9. So, Y W 2 that I have that is the weight vector to be used at the next level of iteration is nothing but W 1 plus eta times sum of Y or Y 3 plus Y 4. So, I had W 1 that is this that is minus 10, minus 1, 2.4.

(Refer Slide Time: 54:06)



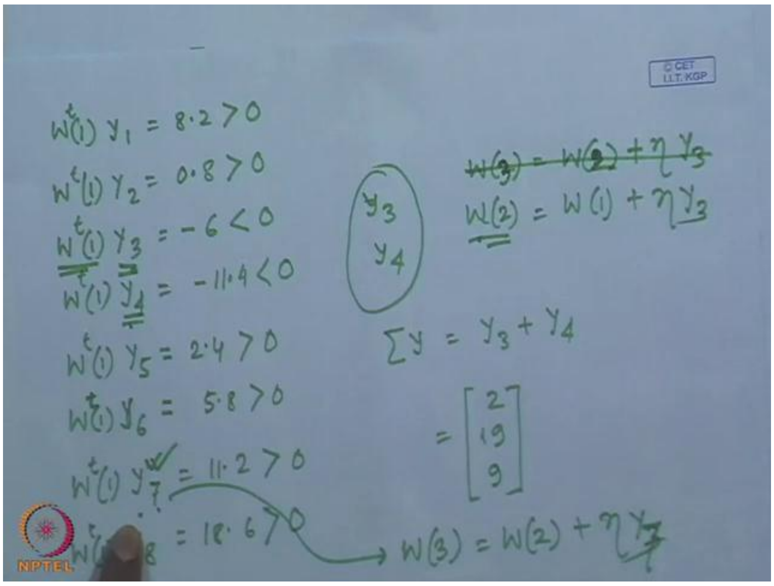
So, from here I get W 2, which will be equal to minus 10 minus 1 minus plus 2.4 plus eta times this. As eta is equal to 0.2, so I will just have plus 2 into 0.2 is 0.4, 19 into 0.2 is 3.8, and 9 nine into 0.2 is 0.18. So, this will be simply minus 9.6, minus 2.8. This is 3.2. So, I get my next weight vector W 2, which is given by this. That is simply minus 9.6; minus 2.8 and this is 3.2. So, this will be simple minus 9.6, minus 2.8. This is 3.2.

So, I get my next weight vector, which is given by this. That is minus 9.6, minus 2.8 and 3.2 again using this W 2. I have to try to see whether the feature vectors are correctly classified or not. If with this W 2, all the feature vectors are correctly classified that means W 2 transpose Y is greater than 0 for all the Y. That means this weight vector W 2 is my correct classifier.

Again, if I get any Y, which is negative for which W transpose Y is negative, those feature vectors are to be used at the next step to get the value of W 3. This is how the perception criteria work. Then we have also talked about the sequential version of the perception criteria.

So, in the original form, what you are doing is in every pass, I am identifying all the feature vectors, which are misclassified by that weight vector. All those feature vectors are added together. That is used for more modification of the weight vector. In the serial version of the perception, criteria algorithm or the perception algorithm, I do not identify all the feature vectors.

(Refer Slide Time: 56:30)



What I do is the moment I find that this W 1 misclassified Y 3, I do not weight for W 4 to be identified. Immediately, I use this Y 3 to get W 2, which will be W 3 plus eta time Y 3. So, immediately I get. This was W 2 is equal to W1 plus eta times Y 3 and maybe this W 2. We will correctly classify Y 4. So, Y 4 will not be misclassified by this, but it may so happen that for Y 7, I find that this feature vector W 2 misclassifies Y 7. So, whenever it misclassifies Y 7, immediately I will compute W 3 to be W 2 plus eta times Y 7.

So, whenever I get a misclassified sample; immediately using that misclassified sample, you modify the weight vector and in a complete pass. If I have W where for the same W Y 1 to Y8, all of them are correctly classified, then I will take that W to be my solution vector. So, I hope with these problems, your concept of the discriminating functions, the decision boundaries, the class regions and the second one have been able to clarify how the perception algorithm works to get the weight vector when supervised learning of a classier is in account.

Thank you.