Pattern Recognition and Applications

Prof. P. K. Biswas
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 5
Bayes Decision Theory

Good morning. So, today we will start our discussion on pattern recognition problems. And today particularly will talk about Bayes decision theory which is the basis of statistical pattern recognition. Now, before I come to the pattern recognition problem let us just have a quick recapitulation of what we have done over last three or four classes. So, our last few classes what we have said is given an object we can find out different types of features of that object.

The features may be derived from the boundary features which are shaped descriptors or shape features. Similarly, the features may also be derived from the region enclosed by the boundary and those features can be say texture features, or they can be colour features they can be intensity features and so on. We have also said that none of this feature on its own can describe an object or a shape.

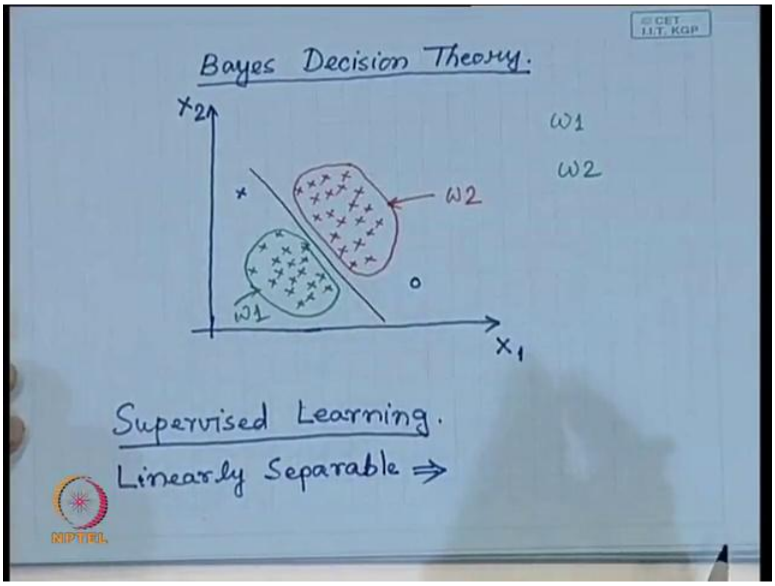
Rather many of the features taken together they can describe an object to some degree of accuracy. So, instead of considering a single feature we have to consider a feature vector, but the components of this feature vector will be from different features may be from boundary features may be from shape features may be from region features, like colour texture and all that. They had to be concatenated in a particular order and whichever order we concatenate them throughout our problem that is modelling of the pattern as well as recognition of the pattern we have to make use of the same order.

So, when I put all these different features in a particular order what I get a feature vector. So, the dimensionality of the feature vector will be dependent upon the kind of problem that we have. In some cases may be just two dimension sufficient in some cases three dimension, in some cases the dimensionality of the feature vector can be more than 100 maybe even 500 or so...

So, as more and more dimension or more and more features, you add to the feature vector the description becomes more or more unique. There is to a great extent the accuracy of the description increases as we increase the dimensionality of the feature vector. So, to what dimension of the feature vector we have to go for in order to tackle a pattern recognition problem that depends upon what is the problem that you have attacked. What is the complexity of the pattern.

Now, whatever it is once we describe or once we represent a pattern by a pattern vector by a feature vector then this entire pattern is mapped to a single point in or feature space. So, if I have a two-dimensional feature then a pattern will be mapped to a point in the two dimensional feature spaces, if we have three dimensional feature vectors for a pattern and the pattern is mapped to a point in the three-dimensional features space. If we have n dimensional feature vector then the same pattern will be mapped to a point in or n dimensional feature space. So, it is nothing but whenever we are finding out the feature vector we are basically mapping that pattern to a point in the feature space, so taking a very simple example suppose.

(Refer Slide Time: 04:01)



Bayes Decision Theory.

Supervised Learning.

Linearly Separable ⇒

We have two-dimensional feature vectors. I am taking this example of two-dimension illustration rather in two dimensions, because I can very easily plot or demonstrate what happens in two dimensions. As the dimensionality increases the complexity also increases. So, I may not be able to plot them on a two-dimensional space. So, that is the only reason I am illustrating this with the help of two dimensional feature space. So, suppose my two-dimensional feature vector has got two components one is X 1 and other one is X 2. So, these are the two components of my feature vector now different patterns will be mapped to different points in this two-dimensional feature space.

So, suppose I have got patterns belonging to two different categories. One of the categories I may call as omega 1 that is one category and the other category maybe omega 2, now because I am taking patterns from these two categories omega 1 and omega 2. So, all the patterns which are taken from class omega 1 they will be put in they will be mapped in the number of points in this two-dimensional features space while the points will be very close to each other.

Similarly, the patterns which are taken from this particular class omega 2 those will also be placed in points mapped to points in this two-dimensional features space while these two points will also be very close to each other. Whereas, these two points corresponding to the patterns from class omega 1 and the points corresponding to the patterns from class omega 2, they are likely to be wide apart. So, effectively what I will have for all the patterns which are taken from class omega 1 they will form a point cloud in this two-dimensional space.

Similarly, all the patterns which are taken from class omega 2 they will also be mapped to point clouds in this two dimensional feature space quite these two different clouds are likely to be wide apart if the patterns are wide apart. So, I will assume that all the points corresponding to the patterns in class omega 1 maybe they will be mapped to points like this. Similarly, all the patterns belonging to class omega 1 they may also be mapped to points or point clouds something like this. So, this is the set of points which corresponding to feature vectors taken from class omega 1 and this is the set of points representing feature vectors corresponding to the patterns taken from class omega 2.

Now, in this simple kind of situation this pattern recognition problem is nothing but something like this that I have to find out a decision boundary. So, the if the points lies on one side of the boundary I will say that point belong to one class if it lies on the other side of the boundary I will say that the pattern belongs to some other class. So, in this simple example you find that I can draw a straight line separating these two different sets of point clouds.

So, if I have an unknown pattern or a feature vector corresponding to an unknown pattern and suppose the feature vector falls on this point somewhere over here. Now, the since the feature vector is on the left side of this decision boundary also that this feature vector belong to class omega 1. If the feature vector happens to be somewhere over here it is on the other side of the decision boundary I will say the feature vector belongs to class omega 2.

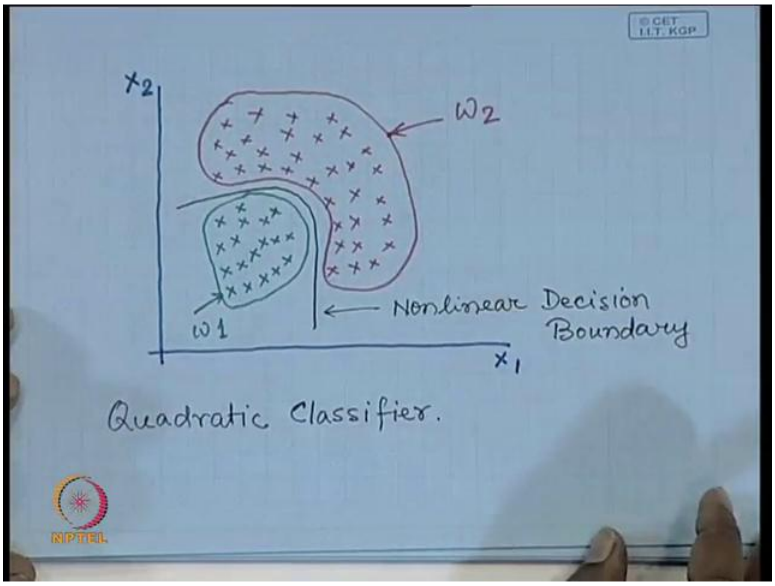
So, here designing of this classifier or training on the classifier means I have to find out the equation of this decision boundary from the set of samples that are given to us. So, what are those training samples this is the setup training samples taken from class omega 1. And this is the set of training samples marked in red taken from another class that is omega 2.

So, both of these two sets of training samples I already know what the last belongingness is? So, that is why this is called supervised learning, that means for designing this classifier or for training of this classifier I take a set of training samples for which the class belongingness is known. I take a number of training samples from plus omega 1, I also take a number of samples from class omega 2 and making use of these two training samples I have to find out this linear decision boundary.

And because in this case the two classes can be separated by a linear boundary this is also known as linearly separable classes. So, classes in this case are linearly separable. So, only when the classes are linearly separable I can find out a straight line separating the two classes in two dimensions. If it is in three dimensions that is I have three dimensional feature vectors in that case, I will have a plane separate separating the two classes. If the dimensional is more than three, I can neither have a line a straight nor a plane, but what I have is a hyperplane.

So, a hyperplane of dimension 4, hyperplane of dimension 5, in case of five dimensional vector feature hyperplane of dimension 6 in case of six dimensional feature vector hyperplane of dimension n in case of n dimensional feature vector, but in all these cases the equation of the decision boundary that will be linear. So, this is a very simple case why the classes are linearly separable. If the classes are not linearly separable in that case what we do?

(Refer Slide Time: 11:32)



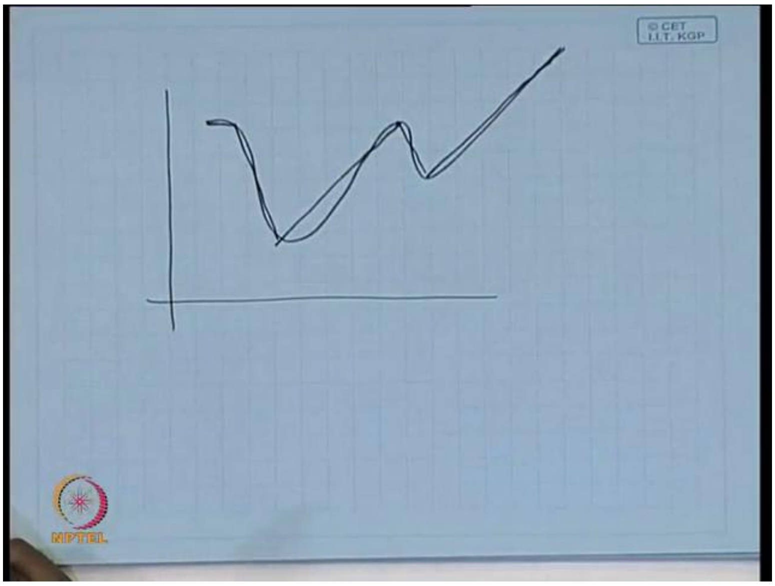
So, let us take another illustration say something like this again I have two-dimensional feature vectors having components X 1 and X 2. I have so these are the samples which belong to class omega 1 that is one of the two classes. So, still I have I am taking the concept of supervised learning that means obvious samples I already know to which class this sample belong to.

Similarly, I have another set of training samples taken from class omega 2 which are distributed like this. So I am putting this samples in red and so you find that I have this set of training samples which are taken from class omega 2, so when the samples belonging to class omega 1 and the samples belonging to class omega 2 they are distributed like this. Now, you find that I cannot separate this process in a straight line. Similarly, in three dimensions, if the samples are distributed like this, I cannot separate the classes by then or in higher dimension I cannot separate the classes.

So, in such cases I have to have a non-linear decision something like this. So, I have to have non-linear decision and we say that the classes are not linearly separated. So, among this non-linear decision boundary the most popular one and the most common one is a quadratic classifier, or at the most we can go for a cubic classifier. Classifiers of higher order more than quadratic or cubic are not very common because designing such classifier is not that simple. Now, if I have a decision boundary as simple as this possibly

I can separate the boundaries by quadratic classifier or cubic classifier, but if the decision boundary is much more complicated than this.

(Refer Slide Time: 15:09)

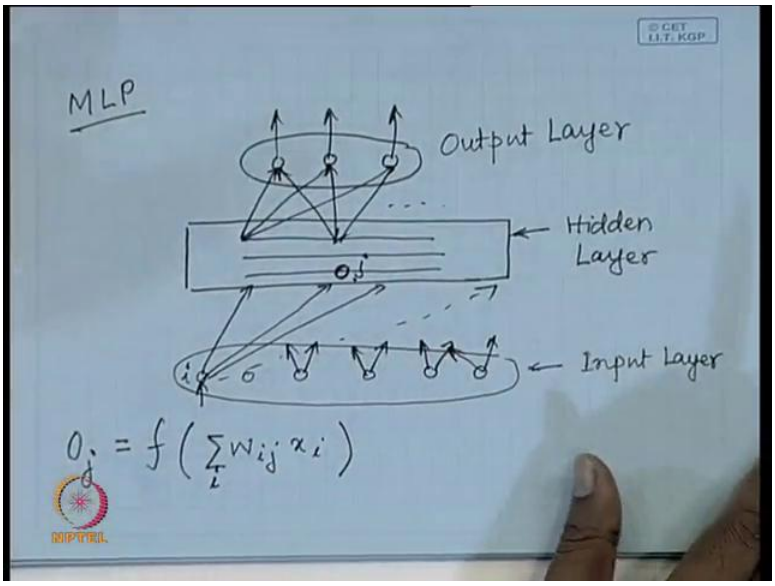


Suppose, I have decision boundary like this something like this, so you find that such boundaries decision boundary are so complicated. It is not very easy to design such decision boundaries analytically, or to have an analytical expression for such a decision boundary. So, this is the case with when I am considering only two class problem that is I have only two classes that is omega 1 and omega 2. The decision boundary becomes much more complicated if the number of classes is more than two. So, I can have 3 classes, I can have 4 classes, 5 classes, I can have even 10, 15 or 20 number of classes. So, as the number of classes goes on increasing and the boundaries are non-linear having analytical expression for such decision boundary is almost impossible.

So, the kind of approach that people take in such complicated cases is making use of neuron network classifiers. However the decision boundary the information of the decision boundary is actually encoded in the weight vectors, or the weight matrices of uneven network. How many of you have done neural network classifier only 1, 2, 3, 4, 5, 6, 7. So, let us see I mean we will come to neuron network classifiers later on, but what does neural network do.

So, this sort of decision boundary though it is a complicated non-linear decision boundary, but I can have a linear approximation of this decision boundary something like this. A neural network in the simplest form actually tries to form a collection of such straight line boundaries or ((Refer Slide Time: 17:12)) straight line boundaries and that collection of straight lines actually model a complicated decision boundary like this. So, what you have in neural network let us see.

(Refer Slide Time: 17:25)



A neural network has one input layer one output layer and non or more hidden layers. This is the output layer and this is input layer. In every layer you have a number of neurons. So here I can have one or more number of hidden layers right. Now, in absence of any hidden layer if i have just an input layer and outputs layer this what is single layer ((Refer Slide Time: 18:36)). If I have hidden layers within that I can have just one hidden layer I can have two hidden layers and three hidden layers and so on.

More and more hidden layers you incorporate more and more complicated decision boundary it can go. So, if I have one or more hidden layers it is called a multilayer perceptron or m l p. Right now what you have is on each of the notes of one layer you have connections to every note of the upper layer. So, from every note in the input layer I have connections to every note in the upper layer.

So, it continues like this and each of these connections have a connection weight. Similarly, from the i th layer to i plus first layer every layer has a connection. Similarly, over here from every node on this layer I have connections to every node in the output layer, so this way it continues. And each such connection has an associated weight with it right so when i take any note in let us say j.

So, let us call it a j th note and it gets connection from every node in the i th layer or every node in the previous layer. So, let us take a particular node say i node to the j th note the total sum of inputs coming from the layers in the i th node will be given by w ij. This is the connection weight from the i th node to the j th note. Times input to this, which let us say X i o k then submission of this over all i because this node is getting the inputs from every note in the previous layer.

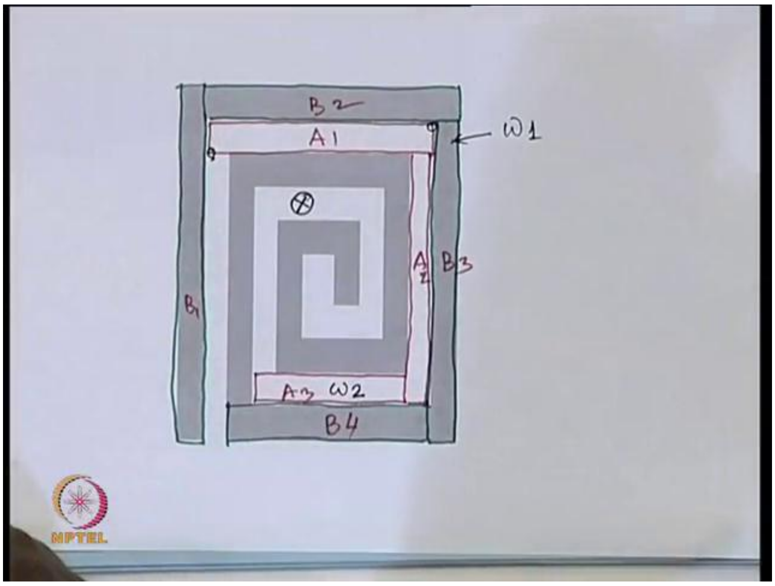
So, this i corresponds to i th node in the previous layer. So, this is the total input to this j th node right then you apply a non-linear function of this f and this gives you the output of the j th node which is o j. Right now if i forget about those nonlinearity what is output of the j th node o j that is nothing but w i j times x j submission over all i. And you find that it is nothing but a linear equation it is linear combination of all these inputs.

Similarly, for every j in this note I will get a linear equation. Right coming to the next layer there also feeding these inputs o j to the notes above it which also combines them by a linear equation and that continues till top up to this. So, finally what I get is output of every node on the output layer is nothing but a linear combination of the imports which are coming to the inputs notes. And when you take the linear combination the coefficients of this linear equation, they are different for different outputs, which is nothing but a combination of the connection weights coming from all these different layers.

So, effectively what I have is I have a number of linear equations which are actually encoded within this weight vectors or weight matrices. So, more and more number of connections or weights I have within this network more and more number of linear equations i can have. So, as a result such a kind of complicated boundary is actually modelled by a number of linear equations by such a simple kind of multilayer perceptron or m l p.

So, will come to details of this later on and how this is related to some sort of analytical linear classifier that we can design following a criteria, which is called a perceptron criteria. Now, you find that in all these cases the decision boundary is are such that either the classes are linearly separable or the classes are non-linearly separable.

(Refer Slide Time: 23:56)



Now, let us consider a situation something like this. So, you find that it is a spiral shaped object and if i assume that all the points which are belonging to this grey shades suppose this is the distribution of two-dimensional feature space right. So if i assume that all the points which belong to this grey shades they belong to class omega 1 and the points belonging to the white region they belong to class omega 2.

So, here you find that the points are not coming as cluster of points, but the points are distributed following a structure, but the structure is so complicated that the points belonging to one class is totally intermixed with the points belonging to other class. Is it possible to have decision boundaries in such cases so that the classification will still be successful.

So, you find that neither linear classifier nor quadratic classifier. Nor even this simple multilayer perceptron that you have discussed just now will be able to give me will be able to model a decision boundary which can classify the points belonging to two classes which is the simplest problem two class problem that we talked about in pattern recognition.

So, I cannot have a decision boundary so easily if the points are distributed like this, so one of the options that we can go in this particular case. Suppose, I define number of rectangles like this, so suppose this is one rectangle or boxes this is another box, this is another box, this is another box and so on. Similarly, four points belonging to the other class i can also have number of boxes like this so it continues like this. So, I actually mark these boxes as class omega 1,I mark these boxes as class omega 2.

Now, once I have a box in two-dimension or a hyper box in n dimension I can represent these boxes by something called mean point and max point or whichever way I represent this boxes. So, what i actually have is I have multiple number of such boxes or multiple number of such hyper boxes. Now, if somehow i can put collect this hyper boxes under the same umbrella. So, all these boxes B1 B2 B3 B4 are put under the same umbrella named as omega 1.

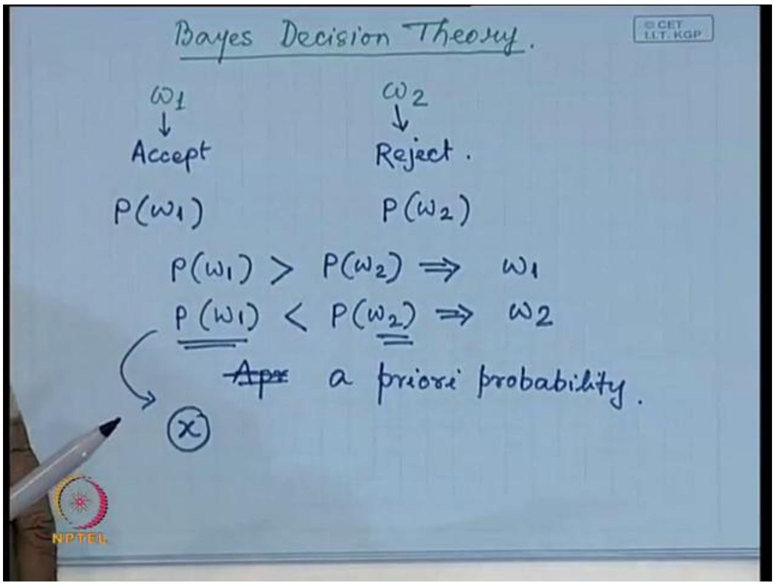
Similarly, these boxes let me call as A1 A2 A3 and so on these boxes I put under another umbrella named as omega 2. So, actually i have a collection of boxes put under two umbrellas one umbrella is omega 1 other umbrella is omega 2 right. Now, if get an unknown pattern suppose the unknown pattern is falling over here. If I have a representation of the box, I can easily find out in which of the boxes this unknown pattern is falling this unknown feature vector is falling. Then I look at the umbrella under which this box belongs and accordingly, I can say whether this unknown sample belonged to class omega 1 or class omega 2.

So, this is what is known as hyper box the classification based on hyper boxes. Such hyper boxes can also be represented by neural network, in which gives the neural network will represent this boxes in terms of two points. One is min point and one is max point. So, I can say this is the min point, this is the max point because this is having the minimum X 1 X2 feature values, this is having the maximum X 1 X2 feature values. So, once I have this mean point and max point it is something like left bottom corner and top right corner if i have these two then immediately I can draw a rectangle i do not need any other information provided.

So, if this boxes represented by this min point and this max point it is what is called min max classifier. On top of this if I incorporate some idea of faze, this is what faze min max hyper boxes. So, if time permits will come to all these details later on. So, far I have talked about all these different concepts just to tell you that what is the domain of pattern recognition problems?

And how complicated the pattern recognition problems can be starting from the similar, very, very simple linearly separable classes to non-linearly separable classes to classes, which can neither be separated by linear boundary nor by simple and non-linear boundary rather the classes will be represented as collection of junks of datasets, which have some class belongings. So, these are the different problem domains or dimension of the problem that we can deal with when we talk about pattern recognition. Now, let us come back to what I said that will be my topic of discussion today that is Bayes decision theory.

(Refer Slide Time: l30:57)



Bayes Decision Theory.

$\omega_1 \downarrow$ Accept $\omega_2 \downarrow$ Reject.

$P(\omega_1)$ $P(\omega_2)$

$$P(\omega_1) > P(\omega_2) \Rightarrow \omega_1$$

$$\underline{\underline{P(\omega_1)}} < P(\underline{\underline{\omega_2}}) \Rightarrow \omega_2$$

~~Apr~~ a priori probability.

$\textcircled{x}$

Now, to discuss about this Bayes decision theory let me take a very, very simple example. The example is suppose I have a manufacturing industry that manufactures machine parts. Now, you know that in any industry there is section called quality control or quality inspection. What is the job of that particular section, they go for inspection of the products produced in that firm. After inspection if they find that the products are acceptable there making in the norms and specifications set for that particular product. It will report under accepted category. If the product is not acceptable there is some defects it will be put under reject category.

So, I was talking about two class problem one class I said omega 1 and the other class I said omega 2. Here, let me assume home that this class omega 1 actually means the category accept and class omega 2 actually means category reject. So, when this quality control department then put an object under accept category or under reject category, they look at some of the features or something of that particular object to decide about this.

Now, out of this if I want to automate the process let me assume that I want to form a decision rule to decide whether the object will be rejected or the object will be accepted. So, for that suppose i want to go for the supervised learning note. So, I had to take the previous history that how many objects have been rejected and how many objects have been accepted by the quality control department. And based on that I generate two probabilities one is P omega, that is the probability that the object will be accepted or the probability that the object belongs to class omega 1. And the other probabilities P omega 2 that is the probability that the object will be rejected, or the probability that the object belongs to omega 2.

So, once I have these two probability I can form very simple decision rule. The decision rule can be something like this that if probability of omega 1 is greater than probability of omega 2 then you decide in favour of omega 1, or if probability of omega 1 is less than probability of omega 2 then you decide in favour of class omega 2. So, here you find that though here we have been able to form a decision rule, but this decision rule is not really logical.

The simple reason is if from the history I have found out that P of omega 1 is bigger than P of omega 2 then all of the new coming objects I will always decide in favour of omega 1 even if it should actually belong to class omega 2, or if P of omega 1 is greater than P of omega 2. I will always decide in favour of omega 1 is less then P of omega 2 I will always decide in favour of omega 2 even if the object may actually belong to class omega 1. That means an objects will always be rejected or it will always be accepted based on or a priori probability P omega 1 or P omega 2.

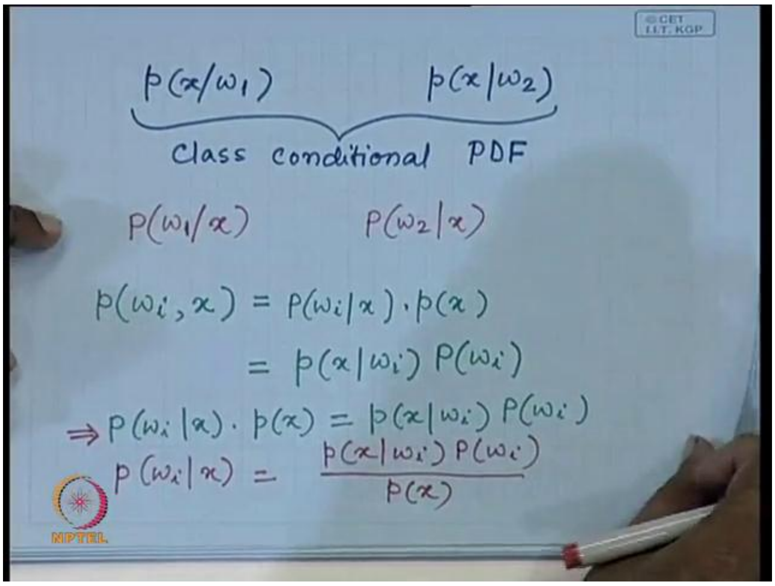
So, this is not at all a logical decision that we are taking. So, to make our decisions more logical what we have to do is, along with this a priori probability we have to combine some feature let us say feature x. And in the simplest form what this feature x can be. Suppose my decision whether the object will be accepted or whether the object will be rejected is based on the finishing of the polish given to that particular object.

So, it is the quality of the polish if I can quantise that if I can measure that then that quality of polish be represented by this variable x. It is good, very good, excellent, bad, and so on. So I can have different shops of measuring this. And suppose it is measurable then that becomes an observation, and this observation is represented by this feature x. By feature x can have various values.

So, what I would like to do is. along with this a priori probability I will also make use of this observation x or feature x to decide whether the object showed belong to class omega 1 or object showed belong to class omega 2 is that. So again I go for the supervised learning mode that means using some objects for which the decision has already been taken.

So, I take some objects from class omega 1 that is the objects which are accepted and I also take some samples of the objects from class omega 2 that is the objects which are rejected. And I measure this feature x for those objects which belong to class omega 1 and I also measure the same a x for the objects which belong to class omega 2. That means I can find out a probabilistic measure a probability density function of variable x for the objects which belong to plus omega 1. I can also find out the probability density function of the same observation x for the objects which belongs to class omega 2.

(Refer Slide Time: 38:24)



$$\underbrace{p(x/\omega_1) \qquad p(x|\omega_2)}_{\text{Class conditional PDF}}$$

$$P(\omega_1/x) \qquad P(\omega_2|x)$$

$$p(\omega_i, x) = P(\omega_i|x)\cdot p(x)$$

$$= p(x|\omega_i)\, P(\omega_i)$$

$$\Rightarrow P(\omega_i|x)\cdot p(x) = p(x|\omega_i)\, P(\omega_i)$$

$$p(\omega_i|x) = \frac{p(x|\omega_i)\, P(\omega_i)}{p(x)}$$

That is I can find out what is P of x given omega 1. So, this is nothing but the probability density function of x taking the objects from class omega 1. I can also find out P of a x by taking the objects from class omega 2. So, I can find out P x given omega 1, I can find out P x given omega 2. So, these are the probability density functions which are called class conditional probability density function. So, class conditional p d f. So, I have P of x omega 1, I have P of x omega 2.

Now, I pattern the recognition problem is the decision problem is for an unknown object i can measure x. And from this measurement x measurement of the future x I have to decided whether I should put this object in class omega 1 or i should put this object in class omega 2. That means what I am interested in is that is my decision should be based on P of omega 1 given x because I have this observation x. And based on that I have to take decision omega 1 or I have to take decision P of omega 2 x.

So, if I find that if P of omega 1 given x is greater than P of omega 2 given x then I will decide in favour of class omega 1. If P of 2 omega x is greater than p of omega 1 given x then I have to decide in favour of omega 1. And this decision appears to be more logical than our simplest decision if the a priori probability are more than, I will put the than an opening at the more logical will be if this probability density function that is p of omega 1 given x and P of omega 2 given x can be combined with a ((Refer slide Time: 40:47)). So, if I can combine these two then I will have a more logical decision rules.

So, let us see that how we can combine these two. Now from the preliminary probability theory you might you know that the joint probability distribution, the joint probability density function that is an object belongs to class to class say omega i. Let me generalise this instead of calling omega 1 and omega 2 let me call it omega i, I can have a value one or two.

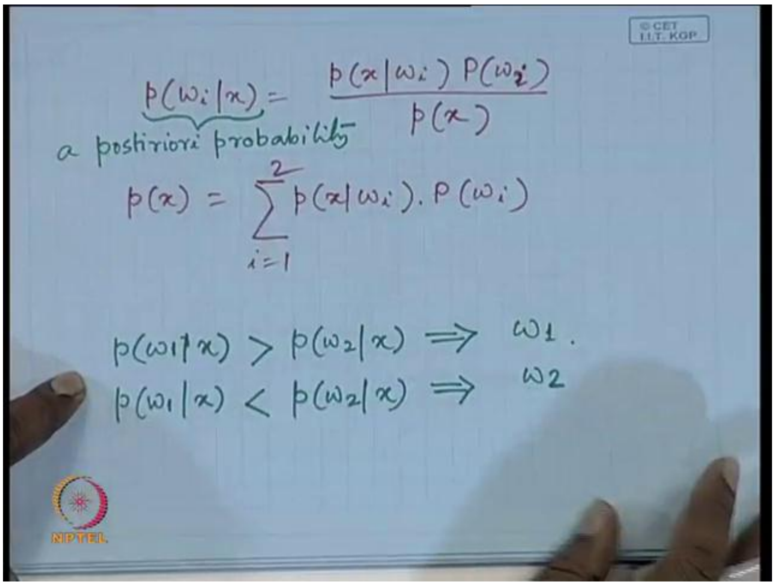
So, a joint probability that an object belongs to class omega i and at the same time has the feature x, this is not a class conditional probability this is joint probability. So, an object having is taken from class omega i and at the same time it will have the feature x. So, this joint probability density function is given by in terms of class condition probability. This is nothing but p of omega i given x into probability of omega x into probability of x P of x or this is same as P of x given omega i into P of omega i.

So, this joint probability can be written in terms of conditional probability that p of omega i x. That is the joint probability that an object is taken from class omega i and at the same time it has a feature x is nothing but P of omega i given x that is the conditional probability into P of x and which is nothing but P of x given omega i into P of omega i. This is again an conditional probability and this is the a priori probability that you have already studied. Now, from here you find that I get a very simple expression that P of omega i given x into P x is same as P of x given omega i into P of omega i.

So, i get this expression from this preliminary probability theory. Now, from here what I get is I already know what is the of omega i. That is a priori probability based on what is the history of classification in that particular form how many objects have been rejected how many objects have been accepted out of the total number of objects that has been produced in that form. I take objects belonging to different classes that mean those objects which have been rejected. I also take those objects which have been accepted. And based on that I find out the class conditional probability density function of x that is P of x given omega i.

So, this P of x given omega i in and P of omega i are unknown object. I measure for an object I measure the feature x and what I have to find out. I have to find out P of omega i given x because then only I can say whether or I will be able to say that whether this particular object having this feature x should be classified into omega 1 or it should be classified as omega 2. So, from here I get P of omega i given x is nothing but P of x given omega i into P of omega i upon P of x. So, from here I can have my decision rule.

(Refer Slide Time: 45:22)



$$\underbrace{p(\omega_i|x)}_{\text{a postiriori probability}} = \frac{p(x|\omega_i)\,P(\omega_i)}{p(x)}$$

$$p(x) = \sum_{i=1}^{2} p(x|\omega_i).\,P(\omega_i)$$

$$p(\omega_1|x) > p(\omega_2|x) \Rightarrow \omega_1.$$

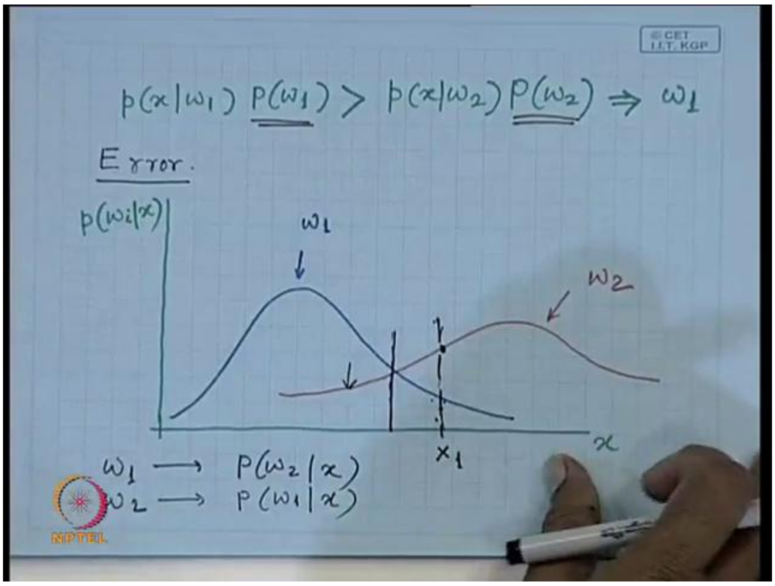
$$p(\omega_1|x) < p(\omega_2|x) \Rightarrow \omega_2$$

So, as I have P of let me repeat this expression omega i given x is nothing but P of x given omega i into P of omega i upon P of x, but what is this P of x P of x is nothing but P of x given omega i that is class conditional probability into P of omega i take the summation because I have got only two classes this is omega 1 this is omega 2. So, i is equal to 1 to 2 is that. So, what I have is I have class conditional probability density functions of the feature x, I have a priori probability that is P of omega i.

And using these two I am computing P of omega i given x which is called a posterior probability. Now, over here my simple decisions rule. So, this is what Bayes theory actually right and from Bayes theory I can have simple Bayes decision rule which will be that if P of omega 1 given x is greater than P of omega 2 given x then we decide in favour of class omega 1. Or if P of omega 1 given x is less than P of omega 2 given x then you decide in favour of class omega 2.

So, this is my simple decision rule, but when I perform this decision when I take this decision I make use of a postiriori probability. And this a postiriori probability actually combines the class conditional probability and the a priori probability. So, I can make use of both that is class conditional probability and a priori probability to take this sort of decision. Now, here you find that overhear my condition simply becomes because P of x will appear at the denominator for both the classes omega 1 and omega 2.

(Refer Slide Time: 48:30)



$$p(x|\omega_1)\underline{P(\omega_1)} > p(x|\omega_2)\underline{P(\omega_2)} \Rightarrow \omega_1$$

Error.

$\omega_1 \longrightarrow P(\omega_2|x)$

$\omega_2 \longrightarrow P(\omega_1|x)$

So, if I expand this expression my expression will be P of x given omega 1 into P of omega 1 this is nothing but P omega 1 given x I am not taking into consideration P of x because that appears in the denominator for both classes omega 1 and omega 2. So, this is my P of omega 1 given x. So, if this is greater than P of x given omega 2 into P of omega 2 then I take decision in favour class omega 1 that means I say this object belongs to class omega 1.

So, over here if P of omega 1 is same as P omega 2 that means, the particular form produces objects which are equally likely to be rejected or to be accepted. In that case my decision is based on the P of x given omega 1 and P of x given omega 2 that is class conditional probability.

For instance if this is same that means for a given x it belonging to class omega 1 or omega 2 they are equal then my decision is based on the a priori probability P of omega 1 and P of omega 2. So, if I cannot take a decision based on the observation I make use of a priori probability. If I cannot take a decision based on a priori probability I make use of observation. In other cases you consider both to take the decision. So, this particular Bayes decision rule combines both your a priori probability and class conditional probability to give you a decision rule which is more logical, and simple to the strong a priori probability.

Now, once we have this what is the error that will encounter. So, what is the probability of error or what is the total error that will have in such cases. Now, let us see suppose this is my x and along the vertical axis I plot the posteriori probability P of omega i given x. And suppose the posteriori probability is something like this. This is for say omega 1 that is P of omega 1 given x. Similarly, this is for omega 2 that is P of omega 2 given x is that. So over here my decision rule was whenever P of omega 1 given x is greater than P of omega 2 given x i decide in favour of class omega 1 and when i p of omega 2 given x is better than P of omega x i decide in favour of class omega 1.

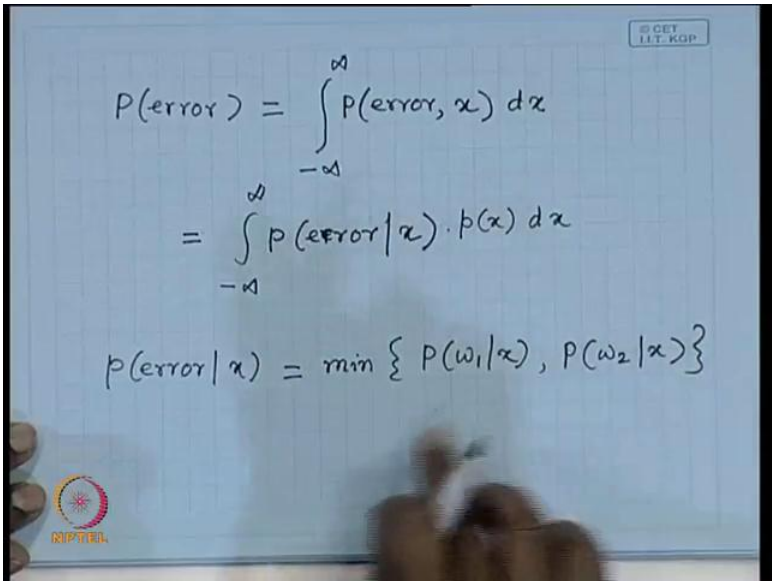
So, my decision boundaries actually the point however p of omega 1 given x is same as P of omega 2 given x. So, this is my decision boundary however P of omega 1 given x is greater than P of omega 2 given x, but still we find that for a given x if i decide in favour of omega 1 still there is a finite probability that the object of may belong class omega 2. Because here P of omega 2 given x is non zero. If it was 0 then I would have said that there is no error, but there is a nonzero probability that the object may belong to class omega 2.

So, there is a finite probability. And what is the probability of error if I decide in favour of class omega 1 the probability o f error is the probability that the object may belong to class omega 1 as if i decide in favour of class omega 2 then the probability of error is the probability to that the object may belong to class omega 2. So, it is very simple that if I decide in favour of omega 1 then the error probability is P of omega 2 given x.

However as if I decide in favour of omega 2, then the probability of error is P of omega 1 given x, but whatever is the probability of error will that is the minimum possible I can have because if I decide for an object for which the value of x is here say x 1 over here p of omega 1 given x is less than p of omega 2 given x. But if I decide the object to belong to class omega 1 my probability of error is P of omega 2 given x which is quite high.

Whereas, if I decide in favour of favour of omega 2 then the probability of error is P of omega 1 given x which is less than p of omega 2 given x. So, whatever decision that we take based on a particular observation x the Bayes decision rule to ensure that the probability of error is minimised is that. So, given a situation like this, what is the total error that we can have that is probability of error.

(Refer Slide Time: 55:04)



$$P(\text{error}) = \int_{-\infty}^{\infty} P(\text{error}, x)\, dx$$

$$= \int_{-\infty}^{\infty} p(\text{error}|x) \cdot p(x)\, dx$$

$$p(\text{error}|x) = \min\{P(\omega_1|x), P(\omega_2|x)\}$$

P error is nothing but the joint probability P error for a particular x into d x, here I have to take the integral from minus infinity to plus infinity, because over here you find that asymptotically the error value extends up to class infinity of the positive side extends up to minus infinity on the negative side.

So, the total error P error will be given by the joint probability P error given x take the integral from minus infinity to plus infinity which is nothing but P error conditional on x P error given x into p x d x take the integral from minus infinity to plus infinity. And what is this P error given x if I decide in favour of omega 1 this P of omega 2 given x if I decide in favour of P of omega 2 this P of omega 1 given x. So, this p of error given x is nothing but minimum or P of omega 1 given x and P of omega 2 given x. So, for a particular value of x whichever is minimum whether it is P 1 given x P omega 1 given x or P of omega 2 given x whichever is minimum P given x is that minimum because I am taking the decision in favour of the other, So, let us stop here today.