Pattern Recognition and Application

Prof. B. K. Biswas
Department of Electrical and Electronic Communication Engineering
Indian Institute of Technology, Kharagpur

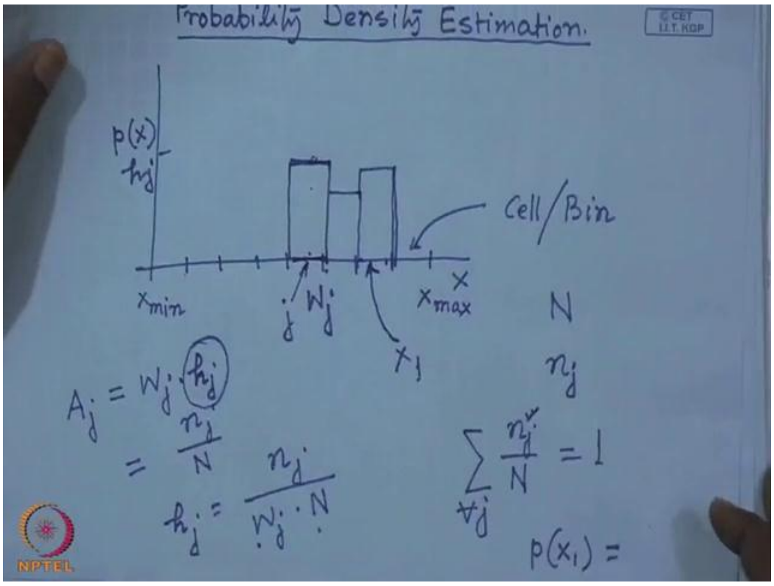
Lecture - 12
Probability Density Estimation (Contd.)

Hello. So, welcome to this lectures series on pattern reorganisation and applications. In our previous class, we have started discussion on probability density estimation. The reason why we have to go for this topic of probability density estimation is that, though we can have analytical expressions of probability density function in certain cases, Where the analytical expression will use some number of parameters and depending upon the kind of density that we that we have we have different types of different number of parameters. So, if I wants to have analytical expression of the probability density functions I have to know beforehand, that what is the parametric form the probability density function will take.

If I know what is the parametric form or what are the parameters, then I can go for maximum like, hood estimate for the values of those parameters. I can make use of some other technique, to obtain the based prosper values of those parameters from the trainings set of samples, which had given, but if the number of samples are very limited. In that case the estimated value of the parameters are not very reliable or not very suitable. In some other case we may have the problem that the parametric form is not really known.

We do not know that how to describe the probability density function as the presented by the set of samples which are provided. So, from those set of samples we cannot really estimate what are the parameters or which parametric formed the probability density functions has to take. So, unless we know the parameters, we cannot obviously estimate the values of those parameters. So, we have gone for the probability density estimation from the given set of training samples without assuming any parametric form of the probability density function. So, obviously in this case I cannot have any analytical expression, but the probability density estimate that we can that we will get that we will solve a proper. So, for as pack and reorganisation applications is concerned, so in order to do that. What we have said in the last lecture is something like this.

(Refer Slide Time: 03:08)



If I take a single variable say x, for which I want to estimate the p x or probability density function of x. Then first what I have to do is I have to know that what is the range of all those of x, that is what is the minimum value of x and what is maximum value of x. So, if I know the range of values of x, then that range is divided into a number of intervals. We have said that every interval is called either a cell or a bin.

So, what we have done is, suppose this is the minimum value of x or let me put it as x min. This is the maximum value of x that I can have say x max. So, this range of x is divided into a number of intervals. Each of this interval is called a cell or a bin. Now, suppose to estimate this probability density function you have been given certain number of samples of x. So, suppose that total number of samples I have is capital N and I take any bin say j th bin and I will tell to find out of this total number of samples sampled values of x, how many such samples like in this j th bin.

So, if the number of samples that lie in the j th bin is given by in n j then I assume that this ratio of n j by N, x may gives estimate of the density of the probability density. So, for as this particular bin j bin j th bin is concerned. So, this has to represent. So, this represents an estimate of the probability density function and this j th bin. I also assume that in a particular bin, the probability density uniform. So, what I will have is a probability estimate something like this.

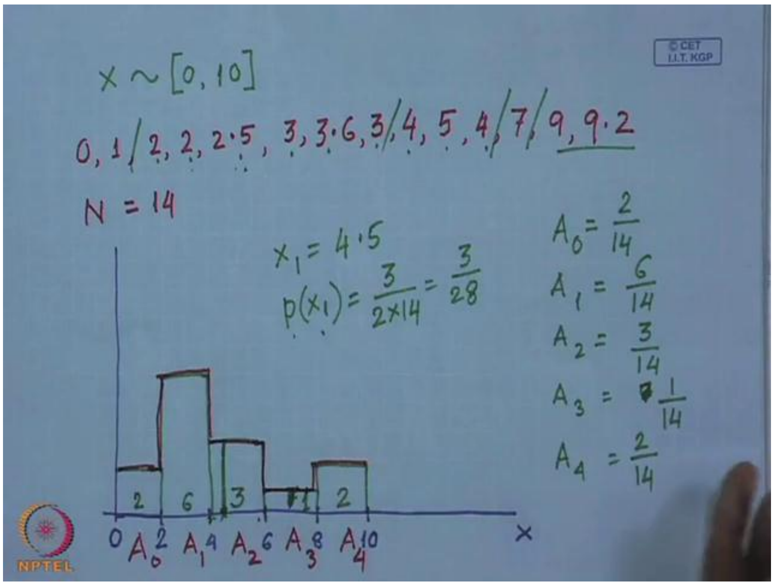
So, with in this j th bin, the probability density function is uniform. I can have different values of this probability density function in different things. I also have another constant that if this probability density function that aim estimating. It has to approximate a continuous probability density function, then the area under the probability density function has to be equal to 1. That means this sum of the areas, of all these bins has to be equal to 1. So, that clearly says that area of this particular area, under this probability density function has to be equal to n j by N because some of n j by N, if I take the summation over all values of j that is equal to 1.

So, given this if the width of these j th bin is equal to say w j, I have to find out what is the height or the value of the probability density function for this j th bin. As I have said that the area has to be equal to n j by capital N. Here n j is the number of samples which are falling in this j th bin and capital N is the total number of samples that I have. So, the area of these j th bin, if I call it as a j that has to be equal to w j, which is the width of the j bin and each j which is the estimated probability density function.

So, for us this j th width is concerned. So, this area a j has to be equal to w j times a h j and reaches nothing but n j by capital N. So, from this can have what is the value of h j which is nothing but estimate of the probability density. So, for us j th bin is concerned. So, I get the value of h j which is nothing but n j upon w j times capital N. At this w j is the width of the j th bin, capital N is the total number of samples that I have and n j is the number of samples, out of this total number of samples which falls under the j th bin. So, this is how we can estimate the probability density function. Once I estimate the value of each j of all values of each j for all values of j, that means for all the different bins I can have estimated values something like this.

Now, for an unknown x say x1, if I find that this x1 falls in this bin the probability p of x1. Will simply be given by this p of x1 is nothing but the height of this particular bin if x1 falls within this bin because that is our exemption that within a bin or with in a cell the probability density is constant. So, this is one of the ways in which we can estimate the probability density function form a set of given examples. Now, to explain this further we will take an example for simplicity, initially we will take an example for single variable, then will extent this concept to vector vectors spaces, where I have multi variant probability density function. So, let us take an example.

(Refer Slide Time: 10:05)



Suppose, I have a variable x at this x of the range of x is between say 0 and 10. So, the maximum value that x can take is 10 and the minimum value which x can assume is 0. So, I am assuming that I have the variable x, can assume any value within the range 0 to 10 were both 0 and 10 are in inclusive. Now, if I take a number of samples like this. So, suppose this is a set of samples which had given to estimate the value of p x, the probability density function for the variably x.

So, find that here I have a number of samples, the sampled values of x as 0 1 2 2. So, 2 appears twice. So, because I can have 2 samples having the same values then 2 .5 then 3 3.6 3 again 4 5 4 7 9 9.2. So, the number of samples that I have is 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14. So, I have capital N, that is number of samples is equal to 14. Now, I want to estimate the probability density function. So, for that as we have said earlier, that I will fought this x and let us assume that as this x or the range of x between 0 to 10 will be divided into a number of cells. Let us assume that every cell will be of equal with and we start with a cell width will be equal to 2.

So, this x it will have a minimum value of 10 a minimum value of 0 and maximum value of 10. I am having cells of width 2. So, the cell boundaries will be 0 2 4 6 8 and 10. So, these are the cell boundaries that I have. Now, I have to find out that out of these samples each sample falls in each of this bins. So, you find that 0, which is on the cell bounded and we have assumed that if a sample falls on the cell boundary, then by convention we assume that cell that particular sample will belong to the cell which is to the right hand side.

So, this sample which is 0 will actually fall in this 0 th bin whose boundaries are 0 and 2. Similarly, this will also fall 0 th bin, the sample value 1 that will also fall in this right bin. If you look go through this layers do find that there is no other sample which is falling with in this 0 th bin. So, the number of samples which falls in the 0 th bin is equal to 2, come to the next one 2, 2, 2.5, 3, 3.6, 3 again all this samples 2, 2, 2.5, 3, 3.6 and 3 all these samples they fall under these bin having the bin boundaries 2 and 4. So, the number of samples which falls within these bin is 1 2 3 4 5 and 6.

So, there are 6 samples which are falling within this. The next 1 within the next bin having the bin boundaries at 4 at 4 and 6 have samples 4 5 and 4 again. So, there are 3 samples which are falling within this bin. Next, 7 this particular sample the sample 7 falls within the bin having the bin boundaries is 6 and 8. The next 2 samples 9 and 9.2 these 2 samples fall in the bin having the bin boundaries 8 and 10.

So, I have 2 samples for falling in this bin. Now, if I compute the areas you find that I call this to be a 0, this to be area 1, this to be area 2, this to be area 3 and one having area 4. So, value of a 0 will be equal to the number of samples in this 0 is equal to 2. I have total number of samples which is equal to 14. So, these area of this under the 0 th bin will be equal to 2 by 14. Similarly, a 1 that will be equal to 6 by 14, a 2 will be 3 by 14, a 3 will be 7 by 14, here I had on the one sample not 7 sample.

So, it will be 1 by 14 and a 4 will be 2 by 14. From this I can compute what is the probability density in different bins because that is nothing but earlier divided by width. I have to assume that every bin has the same width is equal to 2. So, if I divide a 0 by 2, which is equal to 2 I get what is the probability density estimate in the a 0 these divided by 2 gives me what is the probability density estimate in the fast bin and so on.

So, if I compute this what I will have is say if 0 will have this is my each 0 which is nothing but 2 by 14 into 2 that is 1 by 14. Similarly, this will be 6 by 14, 14 into 2 that will be the probability estimate of this. So, I will get it how value something of this form over here it is 3. So, I will get the value of this form, here it is 1. So, I get a value of this form and here it is 2 again. So, I get a value of this form.

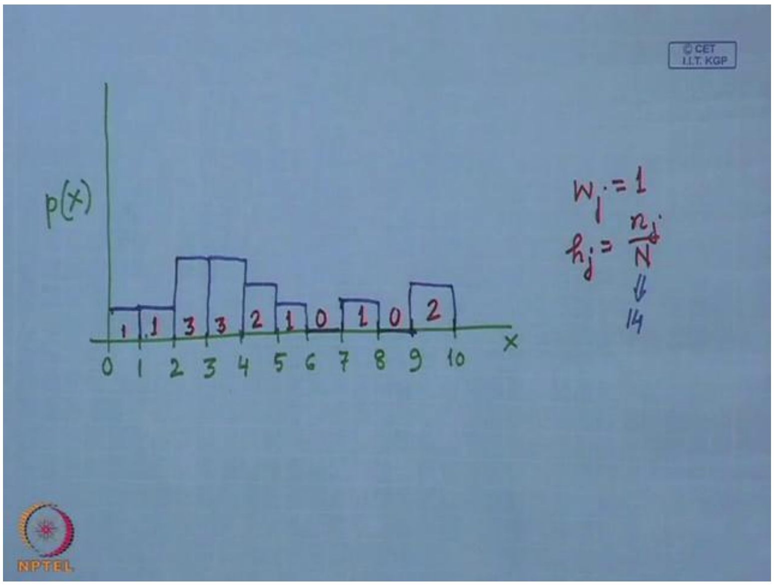
So, you find that given this set of samples and assuming that I have bin size equal to 2 here every bin of same size, the probability density estimate is given by a piece voice constant approximation which is this one. So, this my probability density curve. Now, if you are given a value of x and unknown x. So, I take any value say x 1 is equal to say 4.5 and I have to find out what is p of x 1.

So, here find that x 1 was value of 4.5 in which bin in this x 1 falls. If you look at this diagram this bar graph you find that 4.5 actually falls on this bin. So, 4.5 will be some over here. So, the probability of this x 1 is given by the height of this. So, which is nothing but 3 by 2 into 14. This area is 3 by 14, I have to divide this by 2 because width is 2. So, that gives me this height which is nothing but the probability density.

So, for as these particular bin is constant and x 1 falls within this bin, so the probability estimate p of x 1 has to be equal 3 by 2 into 14 which is nothing but 3 by 28. So, this is how I can estimate the probability density function, for a variable from a given set of samples for that particular variable. Now, having down this the other problem that one has to face is that what should be the width of the bins. Depending upon the width of the bins the number of bins will be different because I have a fixed range of the variable x and in this example what you said is the minimum value of x is 0 and the maximum value of x is 10.

So, if I reduce the bin width, instead of taking the bin width to be equal to 2, if I reduce the bin width to 1, I will have more number of bins. On the other hand from 2, if increase the bin width to say 5, I will have only 2 bins or say bin width say 4, I will have lesser number of bins. So, let us now try to see that, if I vary the bin width, if I vary the number of bins in which I want to divide in the range of x, what effect it is going to have on the probability density estimate. So, I will take the same example and try it with different bin width.

(Refer Slide Time: 21:39)



So, first letter see that what will be the situation if I assume the bin width is equal to 1. So, this is x this is where on the vertical x is I want to put the estimated probability density which is p of x. This is divided into bins of width 1. So, 1 2 3 4 5 6 7 8 9 and 10. So, here it is 0 8 is 10, 1 2 3 4 5 6 7 8 9 10. Now, given this bin let us see and I take the same set of samples which I have taken for the earlier example, that is 0, 1, 2, 2, 2.5, 3, 3.6, 3 again 4, 5, 4, 7, 9 and 9.2. So, give him a set of samples, you find that this sample 0 falls within the first bin and this is the only sample which is which falls in this bin.

So, the number of samples which are falling in this bin is equal to 1. Similarly, the next sample 1, that also falls in this bin. So, the number of samples falling in this bin is equal to 1, then 2 which is on the boundary. So, I will assume that falls in the next bin. So, the next bin contains 1 2 2 and 2.5. So, these are the samples which are contain in the next bin. So, next bin contains 3 samples. Then comes 3 3.6 and 3 again. So, these are the 3 samples which are falling in the bin having the bin boundaries 3 and 4. So, I will have 3 samples over here again.

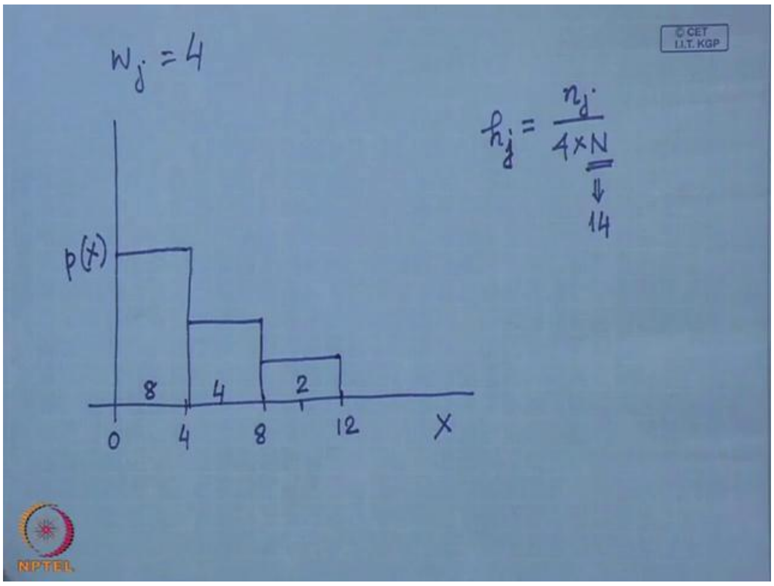
Then 4 and 4 these of the 2 samples which are falling in this bin 5 goes to the next bin. So, a number of samples that I have in this bin, have been bin boundaries 4 and 5 is equal to 2. Whereas, the bin having been boundaries contains these sample value of 5. So, this contains only 1 sample, these bin having bin boundary of 6 and 7, you find that no sample is falling in that particular bin.

So, number of samples falling in these bin is equal to 0. This sample 7 it falls in the next bin having bin bounded is 7 and 8. So, the number of samples that I have over here is equal to 1. Again 8 and 9 this does not contain any sample. The next one 9 and 10 this contains 2 samples, one is 9 other one is 9.2. So, the number of samples that I have over here is equal to 2. So, naturally given this sought of situation. You find that w j that is the width of the bin is equal to 1.

So, what I will have is h j that is height of the probability density function of the probability density estimate will be simply is equal to n j by N. Where N is the total number of samples as we said before and n j is the number of samples falling in that bin. So, if you follow this you find that this one within this 0 at bin I have a probability estimate which is nothing but 1 by 14 because one sample is falling in this bin. The total number of samples are capital N, that I have this is equal to 14. So, this will be 1 by 14 this will be 1 by 14 again, this will be 3 by 14. Here it will be 3 by 14 again, here it will be 2 by 14, here it will be 1 by 14, here no sample is falling with in this bin. So, value of n j is going to 0.

So, this is 0 here again I have 1 bin only 1 sample here again it is 0 here I have 2 samples. So, this is the kind of probability estimate that I will get against this. So, if I reduce the bin width or if I increase the number bins, then the kind of situation that I have is I have shown here. Now, let us see the regards scales, if I increase the bin width or reduce the number of bins then what I will have.

(Refer Slide Time: 27:12)

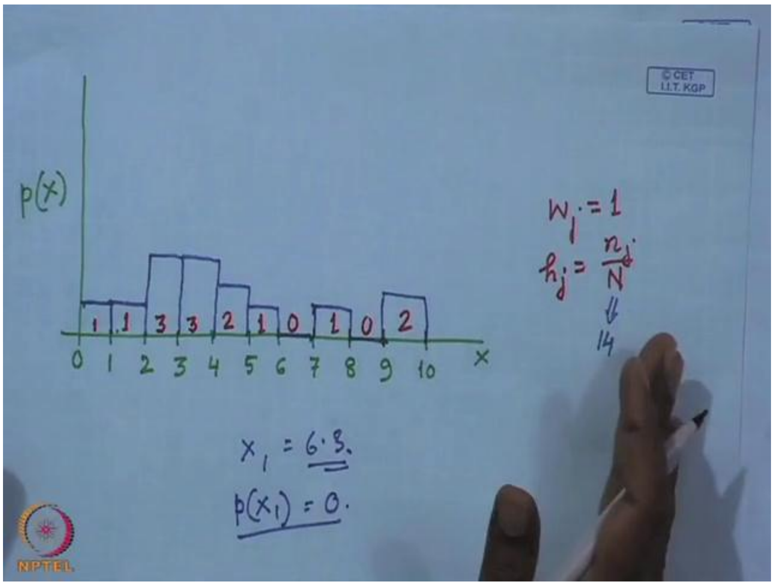


Now, let me assume that I have been with or w j which is equal to say 4. That is same for all g. So, the situation that I have is something like this 0 4 8. The next bin boundaries 12, so I have only 3 bins, but my actual danger phase is up to 10 from this axis I plot x and this axis I plot p of x or the probability of x. Again I use the same set of samples for estimating the probability stability functions. So, these are the samples I have. So, first I have to see, that out of this samples how many samples are falling in the bin 0 to 10. That means the values of x which are greater than or equal to 0, but less than 4.

You find that all these samples they have values greater than or equal to 0, but less than 4. So, I have 1 2 3 4 5 6 7 8, 8 samples which are falling within this bin. Next from 4 to 8. So, all the samples having values greater than or equal to 4, but less than 8 they will fall within this bin. So, here you find that I have these samples 4 5 4 and 7. So, 4 samples are falling within this bin. The remaining 2 that is 9 and 9.2, these 2 samples they are falling within this bin.

So, if I use this now you will find that here w j is equal to 4. So, height of every bin h j will simply be n j by 4 times N, but this N is equal to 14. I have to tell 14 samples. So, for the first 1 it will be 8 by 4 into 14. So, I will have something like this for the next one it will be 4 by 4 into 14. So, the probability density estimation will be something like this and the next 1 will be 2 by 4 into 14. So, the probability density estimate will be something like this. If I use bin size of 2, this is the probability density estimate. If I use bin size of 1 then this is the probability density estimation. So, you find all these defined cases the probability density estimate that I get is different. Then the question comes what should be the proper value of bin width or what should be the proper numbers of bins, because apparently, it appears that this is what I should have from these given set of samples.

(Refer Slide Time: 31:01)

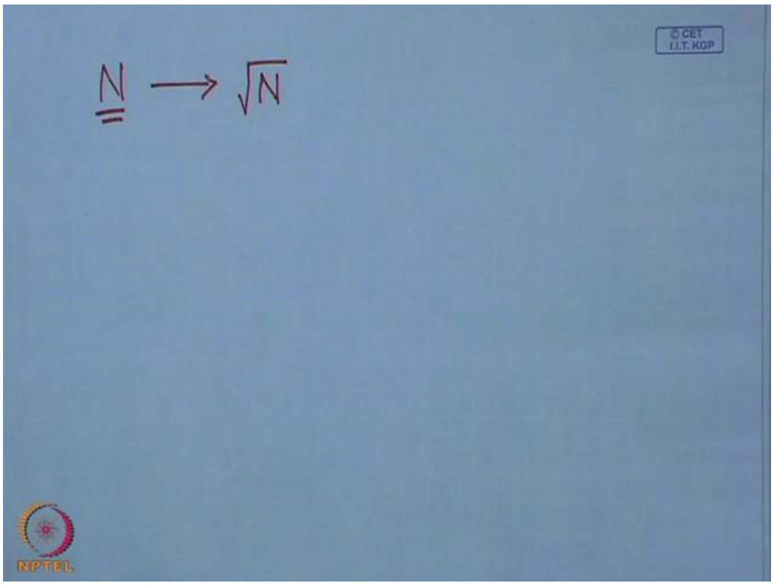


If I use this then this the y fit because if I have a unknown sample say x 1, which is equal to 6.3 and I compute this p of x 1 as this 6.3 falls within this bin for which the probability density estimate was 0. So, p of x 1 will always be equal to 0. So, this is an over fit given the set of samples. Whereas, if I use the larger bin size, that is the situation of this form, then what I am going for is, I am going for more and more uniform density estimation. So, the bin size is larger it will be more and more fat. So, as a result the certain details in the probability density estimate as seen over here will be lost.

So, the choice of the number of bins or the size of every bin should be such, that the bin width is not too large. If I use too large bin width will lose the detail information. On the other hand if the bin width is too small or I have large number of bins, then it is quite possible that many of the bins will be empty. That is none of the samples the given samples will fall in many other bins. As a result the probability density estimate for those bins will be equal to 0.

So, this bin width or the number of bins has to be properly chosen, but unfortunately there is no analytical method of choosing the number of bins or bin width. So, there is a thumb rule which says that, if I have capital N number of samples.

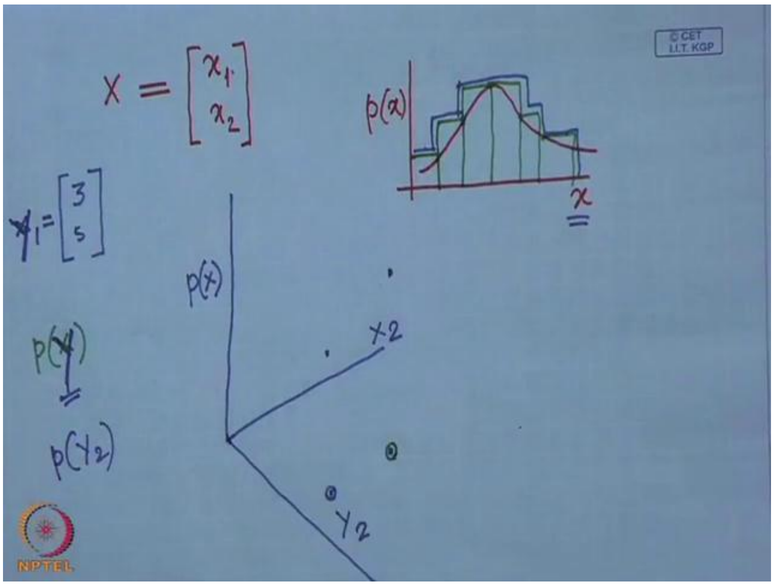
(Refer Slide Time: 33:25)



If I have N numbers of samples then the number of bins should be near about square root of N. This is just a thumb rule, but there is no analytical tools to decide what should be the number of bins for proper probability density estimation. So, this is the case that we have done in case of one dimension. That is assuming that our we have a single variable x and we want to estimate the probability density for that single variant case, but when we go for pattern recognition.

Earlier we have said that in case of pattern recognition, we normally do not deal with a single variable, but we deal with vectors which are called features vectors, where every element in the feature vectors gives some property of the pattern or some information of the pattern, when I take all the information given by all the elements in the feature vectors together. So, all those information together gives me a representation of the pattern. So, in our case or similarly, many other applications the probability density that we need to consider is not a single variable probability density, but it is a multi variant probability density. So, what we have to use is a multi variant probability density estimation. Let us see that how this probability density estimation technique for a single variable can be extended to multi variant. If I have a vector space having multiple number of vectors have been multiple number of elements, how this technique can be extended to multi variant case or if I have a vectors spaces.

(Refer Slide Time: 35:27)



So, I will start with a simpler case, I assume that I have two dimensional feature vectors. That means now this x instead of been a single variable, it is a vector which I represent by capital X. This vector has got two components, one I say x 1, the other one is x 2. So, this is the vector and when I take different samples for estimating the probability density what I will have is different sampled values of x that is this vector. Every sample will have a corresponding sampled values of x 1 and corresponding sample values of x 2. So, I will have different instances of these vectors x 1 x 2.

Now, find that unlike in case of one dimension, where this probability density function was a curve. So, this was my x this was my p x, if I go for continuous probability density I get a curves something like this, if I go for this approximation I have piece wise constant approximation of this probability density estimate which is something like this. So, where effectively my probability density curve is this one.

So, find that in one dimensional case, the probability density function is represented by curve. So, in case of one dimension, this probability density function is represented by curve. So, find that if I extent this to two dimension where I have to two different axis. One axis is representing this component variable x 1 and the other axis is representing component variable x 2.

So, I have a two dimensional space or a plain. So, if I represent, that by co ordinate system I have this axis is representing component x 1 and I have this axis is representing the component x 2. Any vectors having a component value x 1 and then component value x 2. Say for example, if I take vectors a 3 5, this is a vector say x 1, it is equal to 3 five. So, this 3 5 this particular vector is nothing but a point in my two dimensional space or two dimensional plain define by x 1 and x 2. So, this 3 5 is a vector or is a point some over here. So, this is my 0.35. So, every vector every two dimensional vector is nothing but a point in this plane.

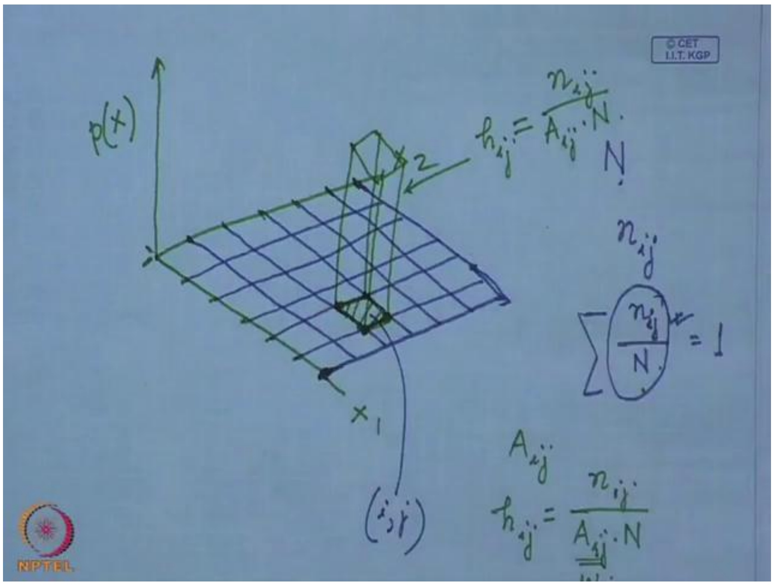
Which is defined by x 1 x 2. So, if I want to now find out what is p of x, that is the probability of vector x. So, thus p of x will be represented on an axis, which is perpendicular to both x 1 and x 2. So, what I get is a three dimensional space x 1 x 2 representing the plain in which all the vectors will lie. I have another dimension another direction which is orthogonal to both x 1 and x 2 and that represents my p of x or x is a vector.

So, p of x for these particular case will be a point in three dimension somewhere over here. Similarly, if I take a vector over here which is say x 2 or instead of calling them x let me call them as y to avoid ambiguity. So, I am vector is y whose components are x 1 x 2. So, this is a vector y 1 instead of calling it x 1 let us call it as y 1. Similarly, if I have a vector y 2 over here that will p of y 2 will be another point in this three dimensional space. So, if I join all these points in three dimensional space to represent my probability density function. You find that all this points define a surface in the three dimensional space.

So, when I say the probability density function of a probability of vector this paralytic density function is nothing but a surface in a three dimensional space, whereas in case of one dimension for a single variable this probability density function is a curve in a two dimensional space. So, that is the difference from curve we are coming to the surface. So, for probability density estimate, as in one dimensional case we have divided the range of x into a number of bins of certain width. In the two dimensional case I have 2 different variables x 1 other 1 is x 2 these are the 2 components of my vector x.

So, both x 1 and x 2, they will also have their ranges. So, x 1 will have a minimum value of x 1 and it will have a maximum value of x 1. Similarly, x 2 will also have a minimum value of x 2 and it will have a maximum value of x 2, which defines what is the range of x 1 and what is range of x 2.

(Refer Slide Time: 42:16)



So, once I have that, then this x 1 x 2 plain, let us put it like this x 1, this x 2. So, this is let us assume that this is the maximum value of x 1 and this is the minimum value of x 1. So, it is 0 to maximum. Similarly, this is the minimum value of x 1 this somewhere over here we have the maximum value of x 2. So, which defines what is the range of x 2? So, if I divide this x 1 into a number of bins as we have done in case of one dimension. Similarly, x 2 is also divided in to a number of bins as we have done in one dimension.

So, find that effectively what we are doing is, this limited space bounded a space, which defines the range of x 1 x 2 because I cannot have any value of x 1 is greater than this, I cannot of any value of x 2 is greater than this, I cannot any value of x 1 which is less than this I cannot have value of x 2 which is less than this.

So, this defines a bounded space or bounded area and when I break this x 1 and x 2 into a number of bins, effectively what I am doing is these bounded area is divided into a number of rectangular size. So, the cells that I have will be rectangular in nature something of this bound. So, these are the cells that I have. You find that if I go to this one dimension, what the cells were linear in nature or it have 1. So, these were the cells in one dimension which are linear a line segments and the sales in two dimension there will be rectangular cells or rectangular bins something like this.

So, as you have done in one dimension here what I will do is I will take a cell say i j cell because it is in two dimension. So, I have to 2 in this is unlike a single index in one dimensional case. So, I pick up an i j cell. As before the total number of samples which have given for probability density estimation, if I assume is capital N and out of this total number of samples suppose n i j is the number of samples which have falling within this i j bin. Then the probability density estimate of these will be given by n i j upon capital N. You notice another point that as I said in case of this two dimensional vector the probability density function is given by a surface.

So, if this surface representing the probability density function is an approximation of continuous probability density. In that case, the volume under this surface here. So, you note the difference in case of one dimension I had a curves. So, I had configured the area under the curve. In this case in two dimensional case it is a surface. So, I have to consider the volume under these surfaces, suppose this is the surfaces and this is my x y plain.

So, I have to find out what is the volume within the surface and the 6 y p. So, if this probability density estimate is an approximation of a continues probability density function. Then the volume under this surface the surface which represents the probability density function has to be equal to 1. In other words what I must have is this some of this has to be equal to 1 and which is obviously true, n j is the number of samples out of this now which falls in the i j th bin. This n is the total number of samples. So, if I simply some them of overall I and j that has to be equal to 1.

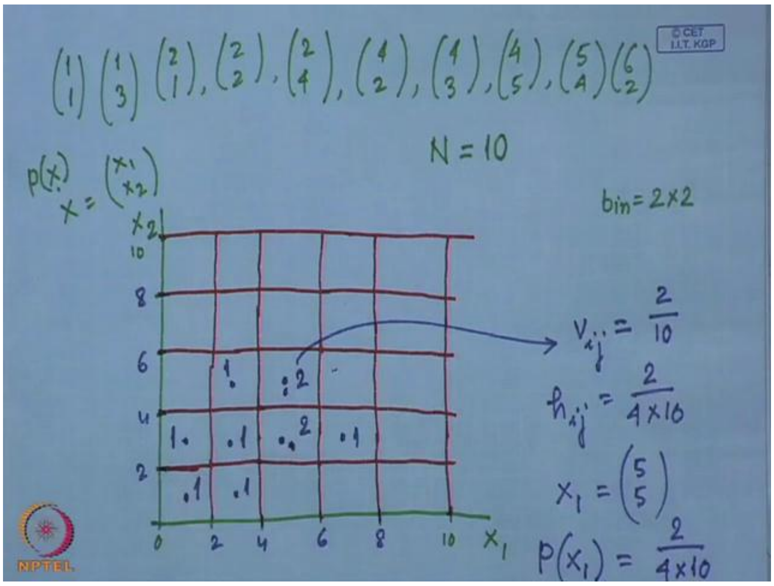
This represents the volume of the probability density function over this i j th bin. So, you find that in case of two dimension the kind of estimate that we had over here. So, these represents an area, here this represents the volume. So, the probability density estimate if this is my p of x, where over this i j th bin, the p of x will be given by a bar something like this. Quite the volume of this bar is given by thin n i j by N and if the area of the base of this bar is a i j, then the height of the bar h i j will be given by n i j upon a i j times capital N.

So, if u compare this with our one dimensional case instead of this area a i j what I had was the width of the bin w j. So, were here this is not the width, but it is the area of the bin i j this is the area of the i j bin which is a i j. Now, it is a very simple extension of the probability density estimation technique from one dimension to two dimension. So, you have been given a large number of samples or capital N number of samples. I have to check that out of these capital N number of samples, how many samples are falling in the i j th bin.

If I know or obviously I have to know that what is the area of i j th bin, then the probability estimate or h i j will be given by n i j that is the number of samples falling in the i j th bin divided by capital N that is total number of sample divided the area of the i j th bin which is a I j. So, over here the height of this which represents the probability density estimate within this bin, that is h i j is simply equal to n i j up on a i j times capital N.

So, this is a simple extension from or one dimensional case to two dimensional case. So, again to explain this further to clarify this concept again I will take an example to explain, how this probability density estimation in two dimension will walk. So, again let us take an example with a number of two dimensional feature vectors.

(Refer Slide Time: 51:14)



So, I have a set of features vectors let us assume that we have set of features vectors that is 1 1 all I in two dimension 1 3 2 1 say 2 2 2 4 4 2 4 3 4 5 5 4 and say 6 2. So, suppose these are the two dimensional features vectors that we have. Using these two dimensional feature vectors I have to estimate the probability density function or p of x had this x is a vector of the form x 1 x 2.

So, total number of sample vectors that I have is 1 2 3 4 5 6 7 8 9 10. So, my capital n or the total number of samples is equal to 10. So, what I will do is, I will consider a two dimensional features space, having x 1 x 2 has 2 different axis. Here if you analyse if you find out the values of x 1 and if you find out the values of x 2, here find that I can easily say that the range of x 1 is between 0 and 10 and the range of x 2 is also between 0 and 10 because I do not have any value of x 1 the first component which is greater than 10 or less than 0.

Similarly, for x 2 I do not have any component x 2 whose value is greater than 10 or less than 0. So, I can very easily assume that the range is between 0 and 10 here also the range is between 0 and 10. Now, let us say that we have the bins, every bin is say of size 2 by 2. So, x 1 will be divided into 5 bins each of length to x 2 is also be divided into 5 bins each of length 2. So, suppose these are the bins. So, once I do this effectively what I am doing is, I am dividing this space into 5 into 5 that is 25 rectangular vectors.

So, these are the bins in that two dimensional space I have. Now, let us try to see that out of these given samples which samples is falling in which bin. So, if you look at the first 1 this is 1111 is falling 11 is nothing, but because this of size two. So, this is 2 4 6 and 8 and 10 2 4 6 8 and 10. So, 1 1 is the point somewhere over here. So, it is falling with in this bin 1 3 1 3 is a point somewhere over here.

So, which is falling with in this bin. Similarly, 2 1 2 1 is a point somewhere over here. So, it is falling within the bin 2 two 2 two is this point, but by convention as we said that we assume that this falls within this 2 4 2 4 again that is this point, by convention we assume that this falls on this bin, 4 2 which is this, by convention we take that that it falls within this bin, 4 3 is somewhere over here by convention I assume that it falls over here. 4 5, is somewhere over here by convention I assume that it falls within this bin, 5 4 by convention I assume it falls within this bin.

Then 6 2 it is this point I assume it falls in this bin. So, find that here sample. So, the volume of the bar over this, if I take this particular 1 volume of the bar, v i j will be equal to 2 by I have total number of samples which is equal to 10. Height of this bar h i j which will be simply this divided by area of the bins and this bin of size 2 by 2 area of the bases is equal to 4. So, h i j will be equal to 2 by 4 into 10.

So, that is nothing, but probability density estimate with in this bin. Now, find that if I have x 1 which is equal to let us say 5 5. So, 5 5 falls within this particular bin. So, p of x 1 is nothing, but this h i j which is 2 by 4 into 10. So, find that we have simply extended the concept of probability density estimate from one dimension to two dimension. I will continuous with this further in lecture.

Thank you.