Pattern Recognition and Applications

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

Lecture - 15
Probability Density Estimation (Contd.)

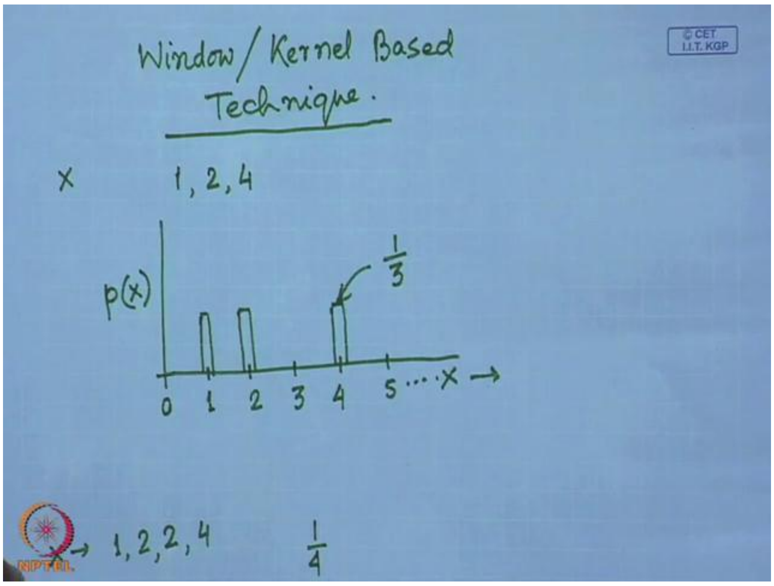
Hello, so in the last class we have discussed about the histogram best technique for probability density estimation. We have started our discussion on kernel based technique for probability density estimation and after discussing about the histogram technique when we started the kernel based technique for probability density estimation. You have noticed that in histogram based technique what we do is, we divide the feature space or the vector space into a number of sales or beams or every sale or beam in one dimension or two dimensions.

In case of one dimension, it has a certain length or certain width; in case of two dimensions it has got a certain area, in case of three dimensions it has got a certain volume and when we move above three dimensions say dimension four and above we call them as hyper volumes. Accordingly, we can define the volume of every cell or every beam, and then the density multiplied by length or the area or the volume, we have said that this should be equal to the fraction of the samples, the fraction of the feature vectors which are given for turning purpose.

So, this probability density multiplied by width of the beam in case of one dimension or the probability density multiplied by area of the beam. In case of two dimensions or it is the probability density multiplied by the volume of the beam, in case of three dimensions and more dimensions that should be equal to the fraction of the samples falling within the beam. Accordingly, we can find out that what is the probability density function in that particular beam? And our basic assumption was that within the beam, we assume that the probability density function is constant.

So, once we do that then we can find out, we can estimate what is the probability of a sample falling within a beam, because we know what is the probability density function, value of the probability density function within the beams. Then we have said that we have got other estimation technique, that is the window or kernel based technique.

(Refer Slide Time: 03:11)



So, what we are going to discuss now is the window or kernel based technique. So, in this window or kernel based technique what we have said is each sample, each feature vector is a representative of a probability density function. This probability density function has to be normalized that the total area or the total volume under the probability density function remains equals to 1. So, in the simplest case if we assume that every sample represents a probability density function or the probability density function is a delta function. So, we started with an example we said that suppose we have got, let us take a very small example.

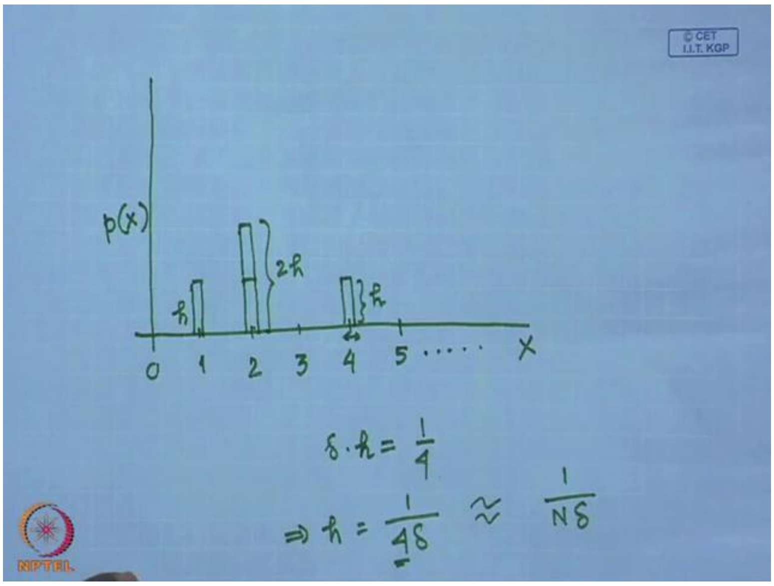
Let us say we have got samples, I assume that I have a single variable or a scalar variable x and the samples which are given for estimation of the probability density function those sample are say x equal to 1, x equal to 2 and x equal to 4. So, using just three samples, we want to estimate the probability density function and I assume that every sample represents a probability density function which is a delta function. So, a delta function has a small width and certain height and because we have got three samples at 1, 2 and 4 for the estimation of probability density function.

So, we have three such delta functions and the total area of the delta functions have to be equal to 1 which means that the area of a single delta functions. The area under single delta function has to be equal to one-third, so I have a situation something like this. I have this variable x and on this axis I put probability of x and I have x equals to 0 here, x equals to 1 here, x equal to 2, 3, 4 5 and so on. So, here you find that if a sample at location 1 gives a probability density which is a delta function, so this probability density estimate will be something like this. Similarly, at 2 it has a probability density estimate which is something like this; at 4 it also has a probability density estimate something like this.

The total area of these three delta functions have to be equal to 1, so that clearly means that area of each of these probability density functions have to be 1 upon 3 because we have three such delta functions. Now, I can have situations something like this that I can have more than one sample or more than one instances of a single value of x. So, I can have a situation something like this that I have x samples of x 1 at location 1 two samples or two values of x which are 2 and I have 4.

Now, I have four samples of x and using this four samples of x I have to estimate the probability density function, so because I have four samples of x. So, I will have four such delta functions and the total area under these delta functions have to be equal to 1, so the delta function or the area of the delta function are for a single sample obviously has to be 1 by 4.

(Refer Slide Time: 07:48)



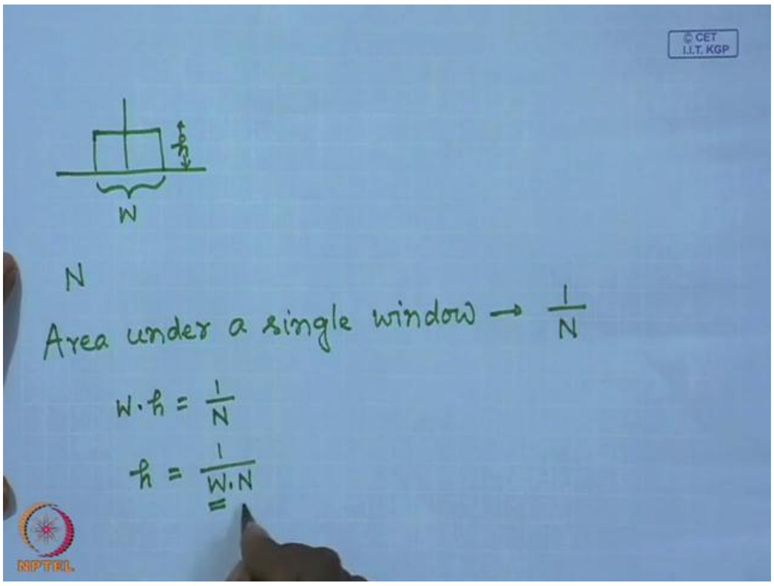
So, given this I will have a situation of this form I will have so on this side I will put probability of x, this is my x, this is my 0, 1, 2, 3, 4, 5 and so on. So, I have one delta function at location 1, I have two delta functions at location 2 because you see that I have two samples of x both of them have to be two, so I have two samples at location 2, so here I have two samples. So, the height of this delta function has to be doubled and I have one sample at x equal to 4, so this is the kind of situation I have and if the width of every delta function is say delta and area has to be 1 by 4.

So, if the height of every delta function, if it is 8, I have h is equal to 1 upon 4 delta, so here is 4 because I have four samples to estimate the probability density function, if I have N number of samples then this h will be 1 upon N times delta. So, the height of this delta function will be h the height over here will be twice h and height over here will be again h.

Now, I have such a kind of delta functions to be used for the estimation of probability density functions. You find that I do not get a smooth probability density estimate or in other sense that this is not really suitable for estimation of continuous probability density function. So, if I want to estimate continuous probability density function from the set of given samples, this choice of delta function is not a good choice.

So, instead of using this delta function what I can use is window function or a Kernel function and the window function or the Kernel function is windows can be of different type. I can have a rectangular window, I can have a triangular window or even the window function can have the shape of normal distribution, so depending upon the choice, I can use a rectangular window.

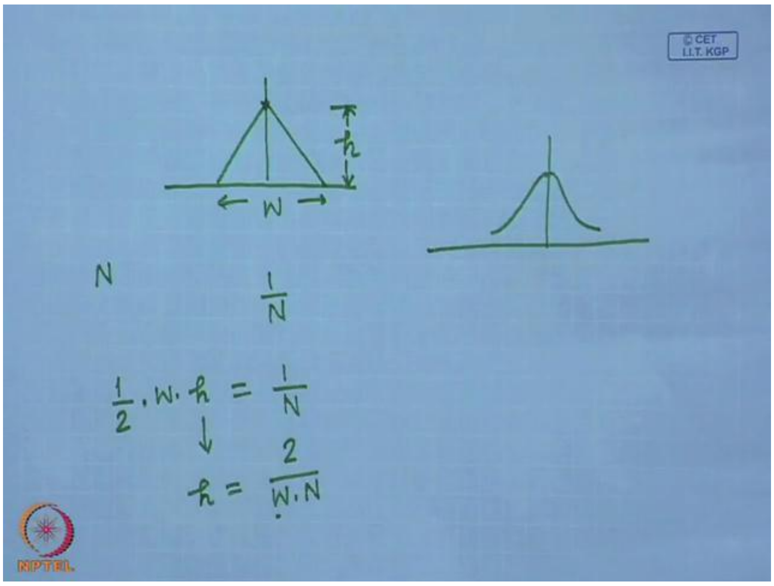
(Refer Slide Time: 10:53)



So, the rectangular window will be something like this of course, it will have certain width certain height. So, if I decide if I fix the width of the window then the height of the window will actually be determined by the total number of samples which are available for the probability density function. So, if I have a total of N number of samples, I will have N number of such windows and the total area of all these windows taken together has to be equal to 1. So, the area of a single window area under a single window, so I will write area under a single window will be 1 upon N. If N is the total number of samples which are which is given for the probability density estimation and from here you find as I have fixed the width of the window function that is W.

So, the area of under this window is W times h which will be equal to 1 upon N. So, naturally the height of the window function is h will be equal to 1 upon W times N, where W is the width of the window and N is the number of samples which are given for probability density estimation. Now, instead of this rectangular window if I use a triangular window, let us see what kind of window, if I use a triangular window.

(Refer Slide Time: 12:57)

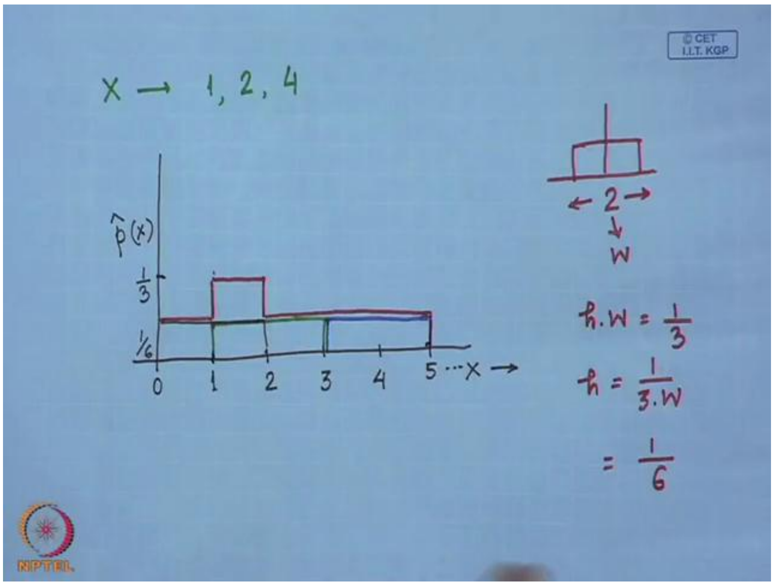


Triangular window will be something like this; again this triangle will have certain width and certain height and if I fix the width to be W then height h of this triangular window will be determined by again the number of samples. So, again if I assume that I have total N number of samples given for the probability density estimation, then the area under the single window. I have total N number of samples and I have total N number of such windows and the area of all these windows are taken together will be equal to 1. So, area under a single window will be 1 upon N, so over here if I compute the area of this triangular window area of the triangular window is nothing but half into W into h and which one has to be upon N.

So, from here I can compute the height of the triangular window h which has to be equal to 2 upon W times N, but this W is the width of the window or the length of the base of the window and h is the height of the window. So, every sample represents a probability density function the probability density function is given by such a triangular probability density having base width is equal to W and the height is equal to h. I can also have a normal density function to represent the density given by every individual sample.

There I can have other types of window functions as well this is not all that I can use however these are the simpler ones. We will explain our concept of window technique for probability density estimation using the rectangular window or a triangular window. So, let us take the same example of three samples.

(Refer Slide Time: 15:40)



I have this scalar variable x having sample at location 1, a sample at location 2 and a sample at location 4. So, what I will do is, I will first use this rectangular window and see how this probability density can be estimated using the rectangular window. So, I will plot estimated probability density p of x and because it is estimation instead of writing p x, I will write p hat x because I exactly do not know what is p of x and along the horizontal axis along the x axis, I will put this scalar variable x.

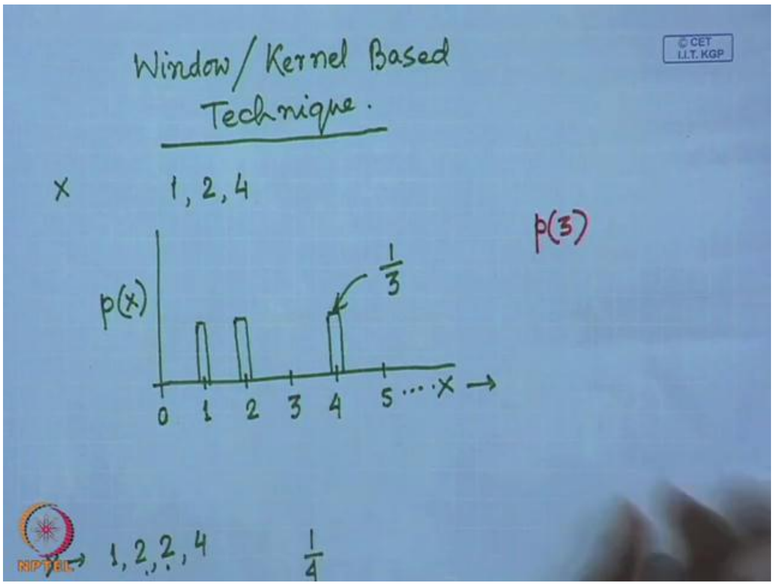
So, this is 0, this is 1, this is 2, this is 3, this is 4, this is 5 and so on and initially I will use a rectangular window. Let me assume that the width of the rectangular window is equal to say 2, so I use a rectangular window and width of the rectangular window I assumed to be 2. So, this is my width W and as I said before that I have got three such samples, so the area of individual window has to be one-third.

So, accordingly I have to compute what is the height of the rectangular window, so the height of the rectangular window h times W that has to be 1 by 3 one-third. So, height has to be 1 upon 3 times W and this W is equal to 2, so this will be 1 upon 6. Now, let us place such a window of width 2 and height 1 upon 6, one at each of the sample locations. So, I have one window of width 2 and height 1 upon 6, this is 1 upon 6 at location 1, at 2 I have another sample. So, at location 2 I have to put one such window again so that window let me use different colors will be this height remains the same 1 upon 6.

The window spans from 1 to 3 because the width of the window is 2 and the same window I also place at location 4, so I will have a window something like this it spans from 3 to 5. So, once I have this sum of all these areas gives me the sum of the probability density function. So, if I add them up you find that from location 0 to 1 in this range, I have got only one window so the height of this window will remain 1 upon 6. Then from location 1 to 2 in this region I have two windows which are super imposing one window which is placed at location 1 and the other window which was placed at location 2, so these two windows overlap.

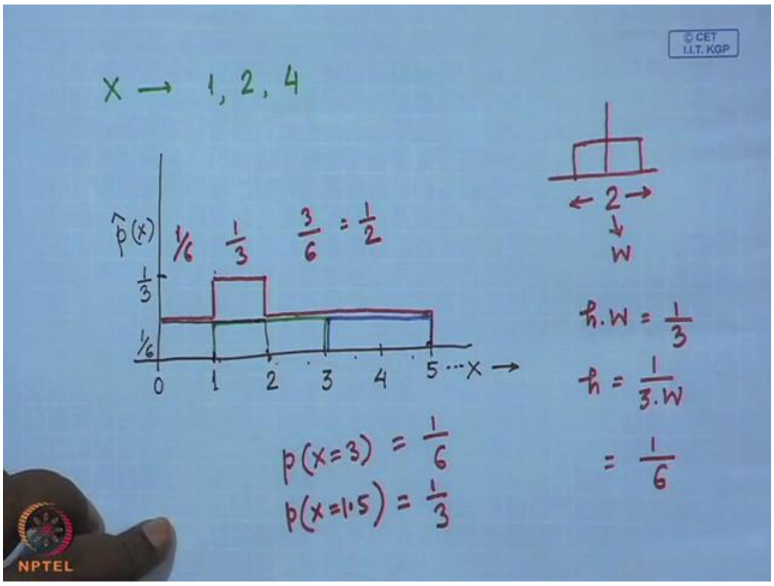
So, obviously here the probability density will be the sum of the heights of the individual windows, so here it was 1 by 6. So, at this location at this height will be 2 into 1 by 6 that is 1 by 3 and from 2 to 5 the height will be 1 by 6. So, this red line or this red curve gives me an estimate of the continuous density of the probability function as represented by these three samples at location 1, 2 and 4. Now, if you compare this we this one that we did earlier with delta function you find this probability density estimate is not a continuous probability density estimate, but it is a discrete probability density estimate.

(Refer Slide Time: 21:02)



Say for example, if I want to estimate what is p 3 here it will give p 3 is equal to 0 because the delta functions do not span up to x equal to 3.

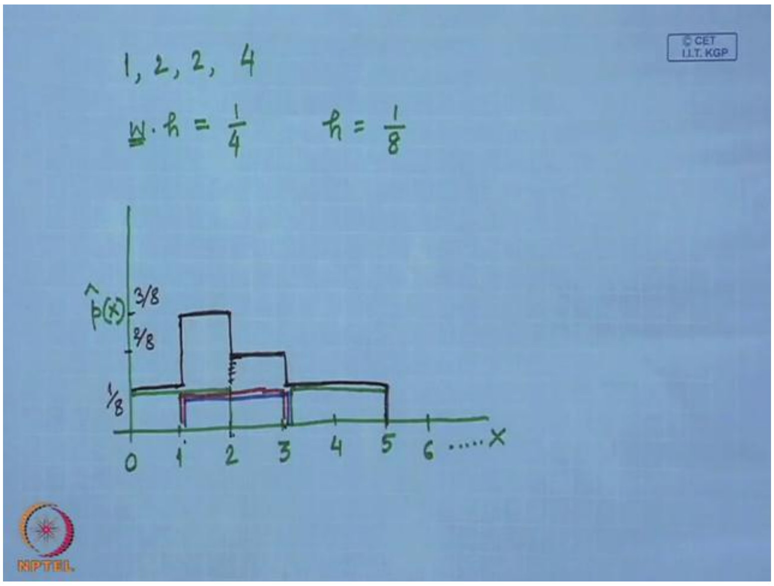
(Refer Slide Time: 21:18)



In this case, when I use a window function or a kernel function I can say that p x equal to 3 will be equal to which is height of this probability density curve. This is equal to 1 upon 6, whereas over here if I want to compute this p of x equal to 1.5 that probability that x is 1.5 which is given by 1 by 3 that is the height of the probability density curve in the region 1 to 2. So, I can use this window technique for such probability density estimate. Now, if I want to compute what is the total area under this probability density curve you find that from here to here, the area is 1 upon 6 because height is 1 upon 6 width is 1.

So, here it is 1 upon 6 from here to here, the area is 1 upon 3 and from here to here, the area is I have 1, 2, 3, three different segments of length one each is having an width of 1 upon 6. So, it will be 3 by 6 which is nothing but half, so if I add all these two, I will get the total area under this probability density curve. So, this was a very simple case when we have used a rectangular window for estimation of probability density. Now, let us see for the same example one more thing what will happen if I get multiple numbers of samples at a particular location.

(Refer Slide Time: 23:35)



So, before I go for use of other window functions, let us see that what will happen, If I have the samples as 1, 2, 2, 4 if I use this then obviously I have to place two windows at location 2 and the total number of samples I have is 4. So, area under every window will be 1 upon 4 and let us assume the width of the window I use the same which is equal to 2.

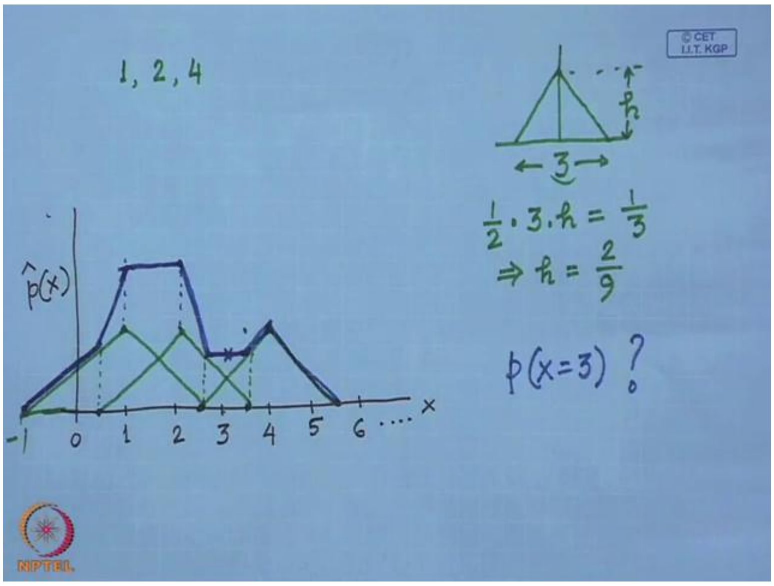
So, width into height has to be equal to 1 by 4, now this is width is 2, so height of every window will be 1 upon 8 instead of 1 upon 6 which we had over here which we had when we had only three samples. Now, we have four number of samples though there are two samples at x equal to 2, so when I have this, now let us try to estimate the probability density I have x I have the estimated probability density p hat of X.

So, again let us put 0 1, 2, 3, 4, 5, 6 and so on, so here what I will do is I will put an window of width 2 and height 1 upon 8 at location 1. So, this window will span from x equal to 0 to x equal to 2, I place two windows of same height and same width at location 2. So, I will have one window and this will span from x is equal to 1 and x equal to 3 because width is 2, so this is one window and this is another window and then I have a single window at location x equal to 4 again of height of width 2 and height 1 upon 8. So, this window will span from x equal to 3 to x equal to 5 because width equal to 2 and now if I combine all these windows together that is what is going to give me p the probability density estimate.

So, here you find that I had only one window this is 1 upon 8 here from 1 to 2 in this region; there are three windows which are superimposed. So, in these locations the height of the probability density that will be 3 by 8 from 2 to 3 I have two windows which are super imposed.

Over here, the height will be 2 by 8 and then from 3 to 5 I have a single window, so height will be this. So, this is what is my final probability estimated probability density function using a rectangular window of width 2 and I have four such samples with two samples at location x equal to 2. So, this is how I can estimate the probability density function, now let us see that instead of using rectangular window if I use q triangular window, then what will be the situation.

(Refer Slide Time: 28:00)



I use the set of same set of samples one sample at location 1, one sample at location 2 and one sample at location 4, so I have three samples and I assume that my rectangular window that I will use that will have a width of say 3. So, I have to find out what will be the height of this window this is the width, so as you know that area of such a triangle is nothing but half into width which is 3 into height. So, this is the area of the triangle and area of each of these triangles has to be equal to 1 upon 3 because I have three samples and the total area of all the windows have to be equal to 1.

So, the area of individual triangular window has to be equal to 1 upon 3, so I put this is equal to 1 upon 3 and once I have this from here I can obviously compare what is the height h which is nothing but 2 upon 9. So, this is quite fine, now let us see how if I place this windows use this windows rectangular windows, then how the overall estimated probability function is going to look like.

So, again I put it the same way this is x this is the estimated density which is p hat of x this is x equal to 0, 1, 2, 3, 4, 5, 6 and so on. So, I have to put this triangular window at location x equal to 1, one of this windows I have to put one window at location 1 equal to 2 I have also to put one window at location x equal to 4. So, let us place this window at location x equal to 1 I put a triangular window and the width of the triangular window is 3. So, that means that I will have 1.5 to the left side and 1.5 to the right side. So, it will be something like this says here if I have minus 1 I will have situation something like this this is the triangular window placed at location 1.

Similarly, the triangular window placed at location 2 it will span from 0.3 to 3.5, so I will have this triangular window, sorry it will come over here this triangular window will be like this. I place one window at location 4, because I have a sample at location 4 and this window will span from 2.5 to 5.5, so it will be like this, so my final probability density estimate will be the sum of all these triangular windows. So, if I compute this I will have a situation of this form say up to this point, I have only one window from here to here I have super imposition of two windows and again up to this point I have superimposition of windows.

So, the kind of probability density estimate that I am going to have will be something like this from here to here it is single window, then up to this point it will be like this from here to here, it will be constant again from here to here, it will be like this. Then again it will be constant, then it will follow single window like this, so my final probability density estimate it is given by the super imposed curve which is shown in this blue line this is the final probability density estimate. Now, given this obviously if I want to find out that what is the probability density at location x equals to 3. You find that probability density location x equals to 3 is nothing but the height the functional value of the final probability density which is given by this value.

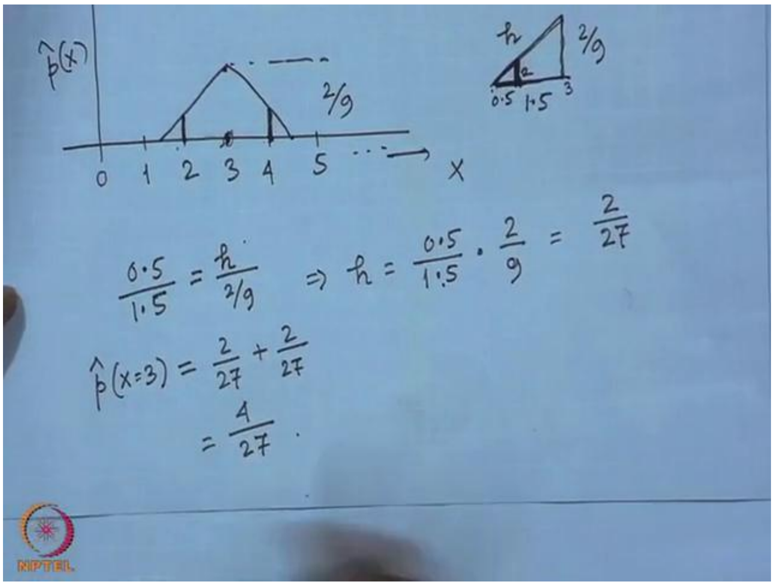
Now, how do you compute this value this is what conceptual, but finally I have to find out what is p x is equal to 3, this is what I have to compute. Now, if you notice carefully you find that this super imposition of different windows placed at different sample locations is nothing but a convolution operation. So, what I am doing is if I assume that every sample location at every sample location I have unit pulse that is a pulse of height equal to 1.

So, what I have is I have a set of pulse trains one unit pulse at location x equal to 1, one unit pulse at location x equals to 2 and one unit pulse at location x equal to 4. So, if these pulse trains is convolved with my triangular window the shape of the triangular window is defined over here. Then I will get this same probability estimated probability density curve. So, this is nothing but convolution operation and because it is a convolution operation, so what will happen if I have multiple number of samples at a particular location.

So, before if I have two samples at location x equal to 2 the height of that pulse I have to take is equal to 2, if I have a sample single sample at a particular location, the height of that pulse will be equal to 1. It will be a unit pulse and if I have say k number of samples at a particular location the height of the pulse at that location will be equal to k. So, I can have pulse train of different heights and once I have this pulse trains in what I will do is I will convolve this pulse train with the window function of the kernel function.

This convolution output will be my final estimated probability density function and that is what it is, so once I have this concept once I understand this is nothing but a convolution operation. I have other way of estimating the probability density at location x equal to 3, so what I will do because it is understood to be a convolution operation I can simply put it like this.

(Refer Slide Time: 36:55)



So, I will have this p hat x and this direction I will have x I have x equal to 0, 1, 2, 3, 4, 5 and so on what I will do is I will put a window at location 3 of height as we have said the height of this window is 2 by 9. So, I will put one such triangular window at location x equal to 3, so this triangular window will be like this. Then what I will do is I will find out that what the samples which are falling under this triangular window, and what is the value of this triangular window function at those sample locations wherever I have a sample.

These values if at any location within this window I have multiple numbers of samples say k number of samples that value has to be multiplied by k that is what is convolution and add all these products together all these products together. So, that is what gives me the convolution, so here at location 2 I had one sample at location 4 I also had one sample, but I did not have any sample at location 3, height of this window as we computed it was 2 by 9 our one sample over here.

So, I have to compute what is this height I had one sample over here I have to compute what is this height and sum of these two heights is the estimated probability at location x equal to 3. So, I can easily compute that following symmetrical triangles I have a triangle this is half of the base so this 1.5, the distance between three and two that is nothing but 1. So, this location 3 and this is location 2 this distance is 1, so this distance is 0.5 this height as we said it is 2 by 9, so I have to compute what is this height. Now, if you follow the properties of symmetric rectangles, you find that I get two triangles over here, one triangle is this and one triangle is this and these triangles are symmetrical.

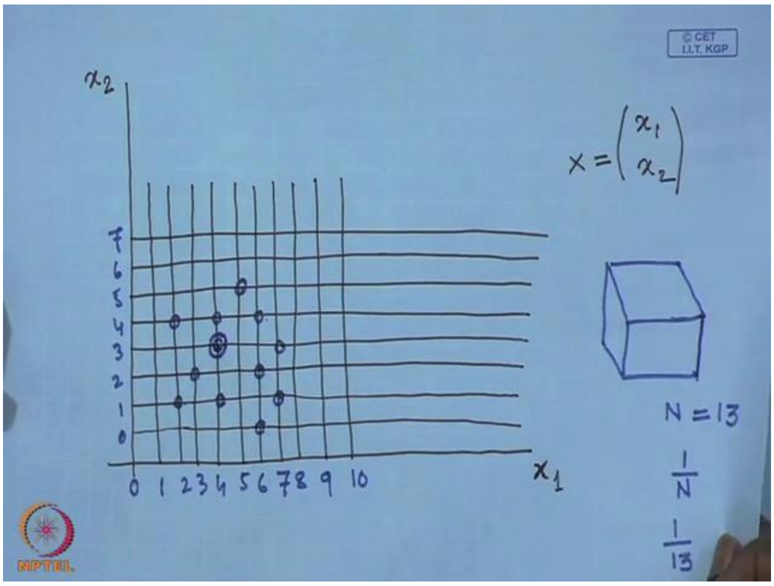
So, if this height is 8 h following the properties of symmetrical triangles I can immediately say that 0.5 upon that is this length upon 1.5 is equal to this h upon 2 by 9. So, I can immediately compute that this h will be equal to 0.5 upon 1.5 into 2 by 9 and this 0.5 upon 1.5 is nothing but one third, so this will be equal to 2 by 27. So, this height is 2 by 27 and by symmetricity this height is also 2 by 27 because this window is a symmetrical window.

So, final probability density at location 3 will be twice of this 2 by 27 because it is this height plus this height because these two are the samples which are in line within my rectangular window. So, the probability estimate p hat of x equal to 3 this will be simply 2 by 27 plus 2 by 27 which is nothing but 4 by 27, so this is what I have. So, the probability x equal to 3 estimated probability x equal to 3 is simply 4 by 27 which I can compute like this if I use this one. So, you find that the probability density estimated probability density at location x equal to 3 is this one which is 2 by 8 if I use a rectangular window assuming that I have two samples at location x equal to 2.

So, whether I use a rectangular window or I use a triangular window like this I can always estimate the probability density function at sample locations where the samples are not originally present and that are what is required for application. So, all these sample locations where we have the samples they are used for training purpose or for learning purpose. Once your system is learnt, then the probability function density is the learnt using that estimated density probability function then I can go for classification of an unknown sample.

So, this is what we have done in one dimension what will happen in multi dimension if the feature instead of being a scalar feature it is a feature vector. We have done using the histogram based technique, how I will do it in case with the kernel based or window based technique, so for simplicity I will assume that our window was are rectangular.

(Refer Slide Time: 42:27)



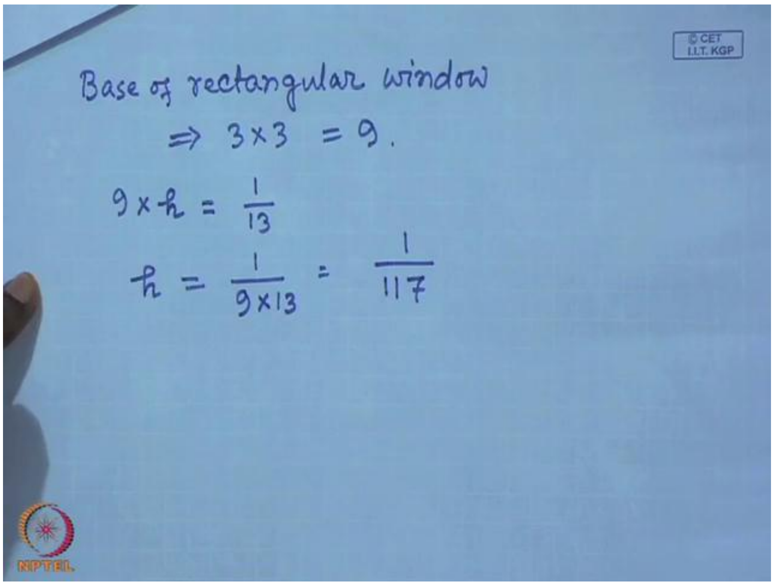
So, let us define a two dimensional space, so I have feature vectored x which is of the form x 1 x 2 where x 1 x 2 are the components of the feature vector x. So, I am using two dimensional feature vectors and I will put use this as x 1 axis x 1 dimension and the vertical axis as x 2 dimension let us put some red. So, these are the grades I have and suppose I have been given samples, let us take some arbitrary samples say one at location, so I will put at this as 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 and so on. So, suppose I have been given these samples of this say one sample over here one sample here one sample here one sample here.

Let us assume that we have got two samples here, then one sample here one sample here one sample here one sample here these samples are arbitrarily distributed. So, using these samples I have to compute my probability density function using the kernel based technique we have already seen how to do it using the histogram technique. In this histogram based technique the entire x 1 x 2 plane will be divided into a number of rectangular beams. For every rectangular beam depending upon the fractions of the samples falling under that beam I will compute what will be the height of the bar in that beam and the height of the bar gives you the probability density estimate.

Here I will use a Kernel based technique or window based technique, so let us assume that we use rectangular windows obviously in because this is 2 dimensional case. So, this rectangular window will actually be a box, so it will be something like this and this box will have certain volume under the probability density surface. As we have used and as we have done in case of one dimension I have to find out what will be the volume of every such box volume under every such box. The volume will be 1 upon the total number of samples which are given for probability density estimate. So, if I have total number of samples which is equal to N, then volume of the individual box will be obviously be one upon one by N.

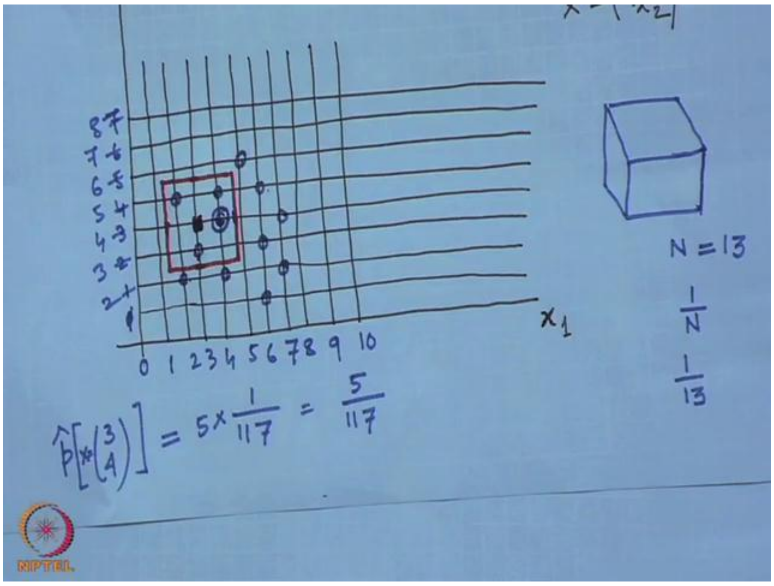
So, here you see that I have 1, 2, 3, 4, 5, 6, 7 because here there are two samples, 7, 8, 9, 10, 11, 12, 13, so there are total 13 samples, so I have N is equal to 13 and as I have thirteen samples volume of every box will be 1 upon 13. So, from this volume I have to find out what will be the height of this box or the height of the rectangular window function. So, to determine the height I have to know that what the base area of this rectangular window is, so again if I use the base area the rectangular window let me assume here that this rectangular window is of base given by let us say 3 by 3.

(Refer Slide Time: 48:33)



So, the base of the rectangular window is 3 by 3 that is area is equal to 9, so I have 9 into height of the box that has to be equal to the volume of the box and this volume we have computed that this has to be 1 upon 13 because we had total 13 samples. So, I have this nine into h is equal to 1 upon 13, so obviously the height of the box has to be 1 upon 9 into 13 which is equal to 1 upon 100 and 17. So, I have decided about what my Kernel function is or what my window function.

(Refer Slide Time: 49:50)



Now, from this suppose I have to estimate that what is the probability density at let us assume this location here I do not have any sample what is the probability density at this location. So, you find this location is three sorry here I have made something wrong this has to be one this has to be 2, 3, 4, 5, 6, 7 and eight, so this location this is feature vector which is 2, 4. So, I have to find out what is the probability that x is equal to 3, 4, so what I do is at this location I place this box whose base is of size 3 by 3. So, 3 by 3 means it will be 1.5 on this side and 1.5 on this side, so this box will span from here to here and here to here.

So, this is the box that I have and now I count that how many training samples are actually falling in this box. So, you find that the total number of samples which are falling in this box is given by 1, 2, 3 and there two samples, so total five samples are falling in this box, so as total five number of samples are falling in this box. So, what I have to do is this probability estimate p of x equal to 3, 4 this will be 5 into height of the box as you have computed which is 1 upon 100 and 17.

So, this is 1 upon 117 which is equal to 5 upon 117, so this is the probability density this is the estimated probability at location x equal to 3, 4. So, this is in the vector space, so how we have been able to do it simply by placing the box and counting the number of samples within the box it is equivalent. If I place a box over here place a box over here place a box over here plus two boxes over here then find out how many of these boxes are actually superimposing at location 3, 4 as I said this operation is same as convolution operation. Alternatively I can place a box at location 3 four and then I can find out that how many samples are falling within the box at location 2, 3, 4 and those samples multiplied by the height of the individual box.

That gives me the probability estimate at that particular location and that is what exactly I have done, now obviously as we have done before this can be done for classification purpose. So, if I have assuming that these are the samples which are coming from say class omega one and I will have another set of samples coming from class omega 2 and I want to classify an unknown data which is may be same location at location 3, 4. So, by considering only samples from class omega one if I compute what is p hat x equal to 3 four, and now because now I am considering the samples from class omega 1, so that will be the class conditional probability density function.

So, it will be p hat x equal to 3, 4 given omega 1 similarly by considering the samples from class omega 2 I can also estimate what is p hat x equal to 3, 4 given class omega two and if the apiary probabilities of classes omega 1 and omega 2 is same. Then out of this p of x given omega 1 and x given omega 2 whichever is higher x will be classified to that corresponding class if the apiary probabilities are not same. They are different, then p of x given omega 1 will be weighted by apiary probability p of omega 1, so this is p of x within omega 1 into p of omega 1.

Similarly, p of x given omega 2 into p of omega 2 which is the apiary probability and out of these products whichever is more x the vector unknown feature vector x will be classified to the corresponding class. So, this we have done in case of one dimensional feature in case of two dimensional features, the same concept can also be extended in case of three dimensional features in case of four dimensional features. In case of N dimensional features by simply by extending this concept of window function to three dimensional window functions to four dimensional windows function to N dimensional window function and so on.

I cannot visualize this I cannot draw that on a piece of paper over here I will not discuss this, but the concept we have already explained when we talked about the histogram based technique for probability density estimation. So, it is the same concept which will be extended which can be used for this window based technique probability density estimation also, with this we come to an end to our discussion on probability density estimation. You find that when we have estimated this probability density function we have not assumed any parametric form of the probability density function.

So, unlike in case of normal distribution where you have mean and standard deviation which are the parameters of the probability density function, I am not used any such parameters. So, the probability density that you estimate is the non parametric density and in many cases because we are not aware of what will be the parameters or what form the probability density function will take. This non parametric estimation of the probability density function is extremely useful, so with this I come to the end of our discussion of the probability density estimation next class we will start some other topic.

Thank you.