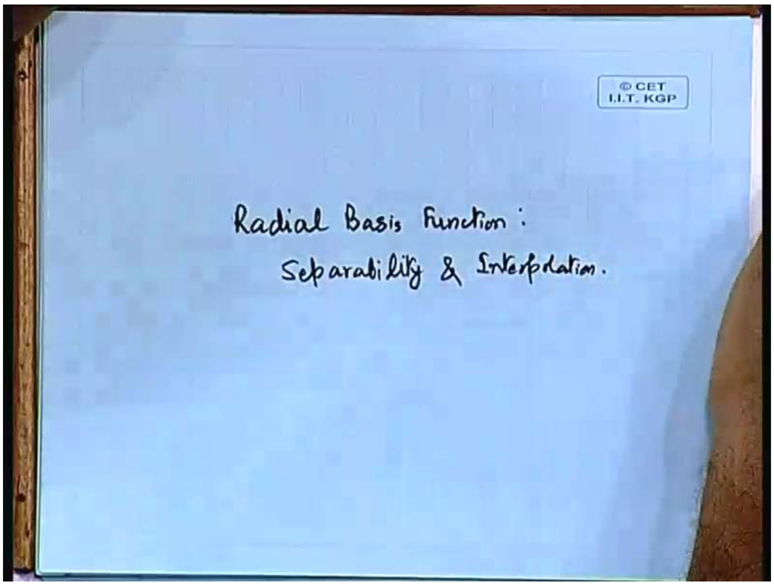
Neural Network and Applications
Prof. S. Sengupta
Department of Electronics & Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 25
Radial Basis Function Networks: Separability & Interpolation

Radial Basis Function.

(Refer Slide Time: 00:52)

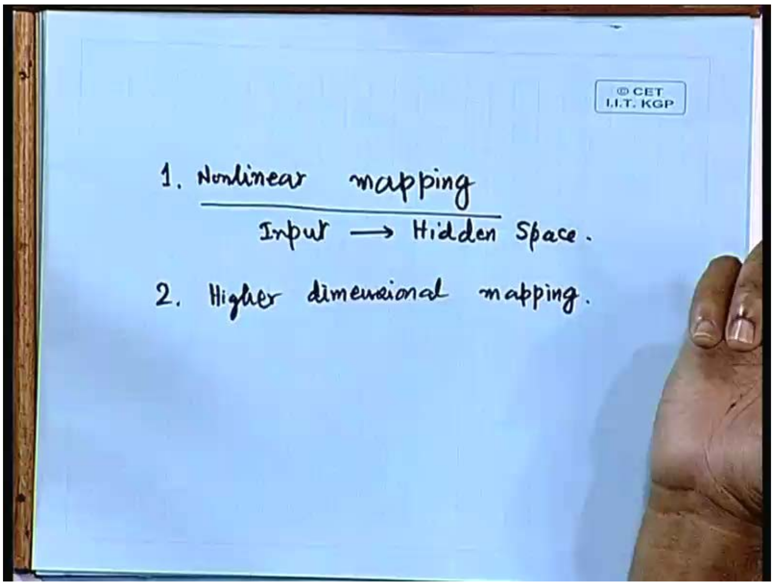


And in that we are going to cover, the aspects of separability and interpolation. Now, before we going to this aspect. We need to quickly summarise, what we did in our last lecture. Now, in our last lecture we had introduced the radial basis function. And where few of the points which you should noticed that. Firstly that it is a multilayer perceptron, but here the problem is formulated in an interpolation way not in a stochastic approximation way, that is number 1.

And then, we had seen a very important theorem which is Cover's theorem which states that if you are mapping from an input space to higher dimensional hidden space, then it is separability increases. So, we had presented the mathematical expression which shows us that, that means to say that where we consider that the patterns are separable. In the input space the patterns are separable, but not in the linear, they do not have the linear separability.

But, when we map it into the phi space or the hidden space, then it is becoming linearly separable. So, in other words, with respect to the input space we are having kind of a non-linear separability, which we had shown one or two examples, that it could be a spherical hyper sphere separability, or any quadric separability all these things can exist. Now, the two basic important points that came out of Cover's theorem is number 1 that the mapping that we are doing from the input space to the hidden space, that is essentially non-linear. So, one of the basic fundamental things that we have observed is a non-linear mapping.

(Refer Slide Time: 03:14)



It is a non-linear mapping from input to the hidden space that, is number 1 and number 2 is a higher dimensional mapping. In other words, we must ensure that when we are mapping from the input space to the hidden space. Then preferably phi should be having a higher dimension. In fact, with increase of dimension in the phi space, the possibility of having a linear separability in phi space increases.

In fact, we can very easily guess that what would be the structure of such a kind of an RBF network, because what happens that we are first taking the set of inputs. Those inputs we are mapping into the hidden space. And that mapping is being done by a set of basis functions. Now, all these hidden neurons that we are taking, they are working based on those functions.

And then, so then these mapped values which is in the m 1 dimensional of real space. If we are taking m 1 to be the number of hidden neurons, then from the hidden neurons to the output, there we can keep the usual linear activation unit neurons. Because, once we have separated out the patterns in the, once we have performed the linear separability after mapping into the phi space. There is no other difficulty for us to remap it into the outputs space using the neurons.

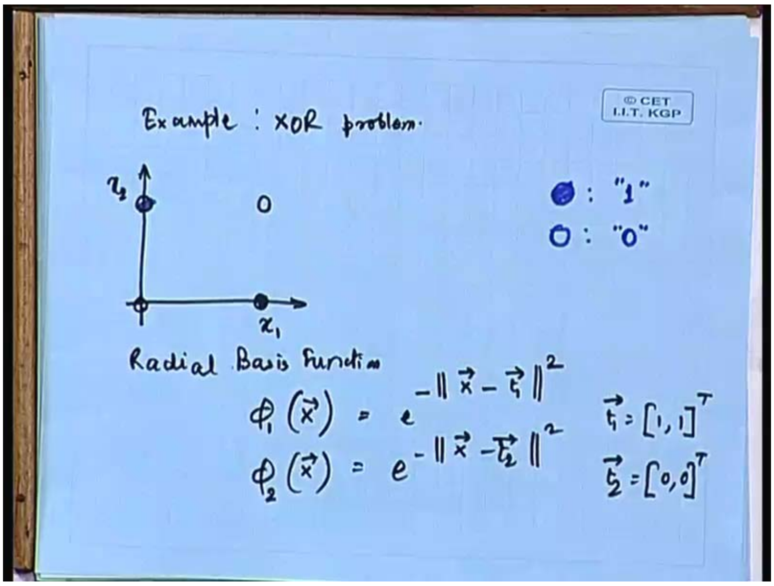
So, that is there and now these two points that we have got out of Cover's theorem, that is non-linear mapping and higher dimensional mapping. In this the essential that is there is the non-linear mapping. Higher dimensional mapping is indeed preferable, but of course, that increases the cost. Because, if you are keeping too many hidden neurons as used, as the radial basis function then, you are increasing the cost of the system.

Now, even if you do not decide to increase the number of hidden neurons, if supposing you try to keep the dimensionality of the input space and the dimensionality of the hidden space to be the same. Then the essential point that, one must have is that there should be a non-linear mapping. A non-linear mapping is very essential, even if you keep the dimensions same. If you are performing a proper non-linear mapping, in that case it is possible to have a separation.

Now, what we are going to, now in yesterday's class what we were actually talking about was, that when we are mapping from the input space to the hidden space, then we have to use some basis function. And now we are going to show that with one simple example. And again the simplest of the examples that one can always think of regarding the linear separability is our very popular exclusive OR problem.

So, we are going to take up an exclusive OR problem and a very typical example of the radial basis function which we are going to take from the book by Haykin. So, Haykin's neural network book itself gives this same example and which I am going to tell you for your better understanding.

(Refer Slide Time: 07:35)



Now, let us take, this as an example of XOR problem. And what we have here is that let us first draw the input space, so we draw the input space. So, the input space is the x vector space and the simplest of exclusive OR is a two input exclusive or where we are having two inputs and one output. So, the x vectors components are x 1 and x 2 which we are showing in the form of these two axes.

Now, that means, to say that the patterns that one can feed to the system are 0, 0 0, 1 supposing this is corresponding to x 1 equal to 1 and then, 1, 0 and 1, 1 these 4. Now, out of these actually two patterns they belong, out of these two patterns belong to the class 0. Let us say that 0, 0 means that it is, this belongs to class 1 and the open circle that we have drawn that belongs to class 0.

So, now what we are going to have is that we are going to, so it is definitely not linearly separable in the input space. So, what we do is that, we use a pair of Gaussian hidden functions as the radial basis function. So, as radial basis function we take a Gaussian function or rather two Gaussian functions are used as basis one is phi 1. So, where the definition of this basis function is phi 1, whose argument is the x vector is equal to e to the power minus norm of x vector minus another vector which is a t 1 vector.

Having of course, the same dimension as that of the x vector, that means to say that here x vectors dimensionality in this example is 2, because there are two inputs over here. And likewise the t 1 dimensionality will also be 2, and this is and this norm should be taken in the square. So, that means, to say that this is our very popular Gaussian function which you can immediately identify.

Now, this t 1 basically is what, t 1 is nothing but, the center of the function, so the Gaussian function that we are considering is definitely centered around t 1. So, this t 1 we can take some of the, so t 1 we take centered around two of the patterns that is giving us a class equal to 0. Let us say, so we take this and this to be the centers of the Gaussian. So, we take t one vector to be equal to 1,1 and we likewise take another radial basis function phi 2 x, which is equal to e to the power minus x minus t 2 this square.

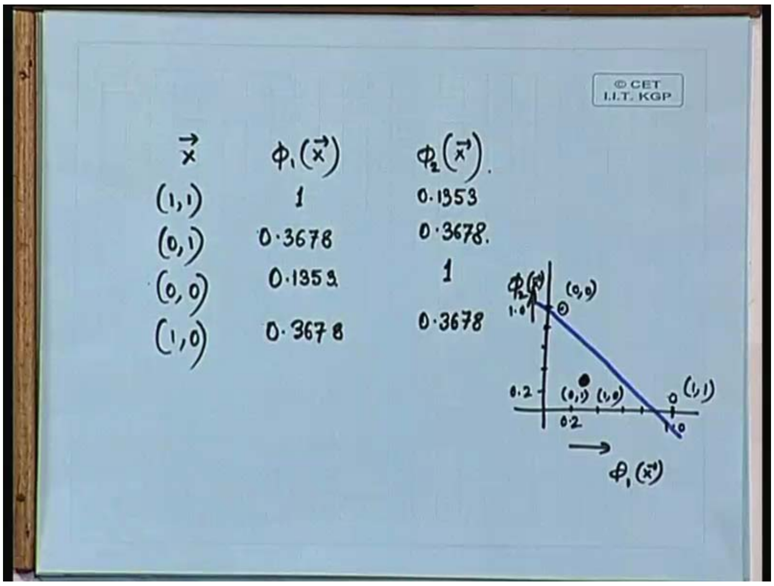
And t 2 in this case is defined to be the 0,0, so these are the two centers around which we have defined the radial basis function. So, how many radial basis function we consider only two. Our dimensionality is also two, that means to say that when we are remapping the function from the input space to the hidden space or the phi space. We are not changing the dimensionality of it. The dimensionality still remains as two, now what we have to compute is that.

Now, phi 1 x and phi 2 x, there are two radial basis functions and we have to compute their values, the values of this functions, because ultimately this functions are going to give us values in the real space. And if we consider the values of phi 1 x and phi 2 x together, then that gets mapped into a two dimensional real space or square space, if case may have to.

So, then what we have to do is that we have to, so let us now start mapping it; that means, to say that for x, we have to substitute the various values like 0,0 could be the 1. Let us take phi 1 x. So, once we will compute phi 1 x with 0, 0 pattern, then we will be computing phi 1 x with 0, 1 pattern, 1, 0 pattern, 1, 1 pattern. And like that phi 2 x also we will be computing for 0 0, 0 1, 1 1 and 1 0, now you can see that the values will definitely be different.

Like say for example, we can very clearly see that when x vector is fade as 1 1, then it becomes e to the power 0 over here. So, which means that the mapped values, mapped real value for the phi 1 function becomes equal to 1. So, let us see that what values do we get out of this, now the values that come out after the computation of this. So, can be written as follows, so for x vector, so we make it in a table or form.

(Refer Slide Time: 13:23)



The results of the phi 1 vector and the results of the phi 2 vector, and when x vector is equal to 1 1, then phi 1 vector gives what, a value of no e to the power 0, e to the power 0 means the phi 1 value is equal to 1. So, this is equal to 1 and phi 2 will be a value of course, other than 0, other than 1 it will be because x vector, then becomes equal to 1 1, now x vector becomes equal to 1 1 and t 2 vector is 0 0.

So, there will be some real value that is computed and in fact if one takes down a calculator, it is possible to calculate that phi 2 x is equal to 0.1353. Then we take 0 1, now 0 1 you see 0 1 is not any of these radial basis function, this Gaussian functions centers. So, that means, to say that both phi 1 and phi 2 will give some non zero and non one real values, in fact the values that work out to be, in fact they are becoming equal values.

So, the values that are computed are 0.3678 for phi 1 and 0.3678 for phi 2 and then, for 0 0 what is it that one can expect. 1 for phi 2 definitely because phi 2 is center itself is 0 0, so for phi 2 we will be getting a value equal to 1. And for phi 1 we will be getting the value that we had got out here, because after all it is a squared norm that we are computing. So, the value of this will be 0.1353, and what will be the value of one zero any guess.

The same value that we had got for 0 1 case, so that is 0.3678 and 0.3678 and now we plot now we plot in what space not in x 1, x 2 space anymore. Now, we have to plot it in the phi 1 phi 2 space, because that is our mapped space. So, now we have phi 1 in this axis and we have phi 2 in this axis, so this is phi 1 x and this is phi 2 x and we plot these values. You see that now if we are considering, let us take this to be 0.2, this to be 1, so that all these divisions that I have marked are in steps of 0.2 and likewise here this is equal to 1.

So, here all these are in steps of 0.2, so we can see that when it is 0 1 or 1 0, the values that are coming are 0.36, 0.36 again for this one also 0.36, 0.36. That means to say that we will be having a point somewhere here, we will be having a point somewhere here where the phi 1 value and phi 2 value both are 0.36. And they will be corresponding to the patterns 0 1 as well as 1 0, both the patterns will be mapped into the same position.

Now, these are what these, these should be open circles or a closed circles according to our definitions, they should be closed circles. So, this is the 0 1 and 1 0 pattern and then, for 1 1 pattern where will it be, it will be with phi 1 equal to a value of 1 and phi 2equal to a value of 0.13, so somewhere over here there will be an open circle. And likewise for phi 2 equal to 1 means here and phi 1 equal to 0.13, we will be having another value with open circle.

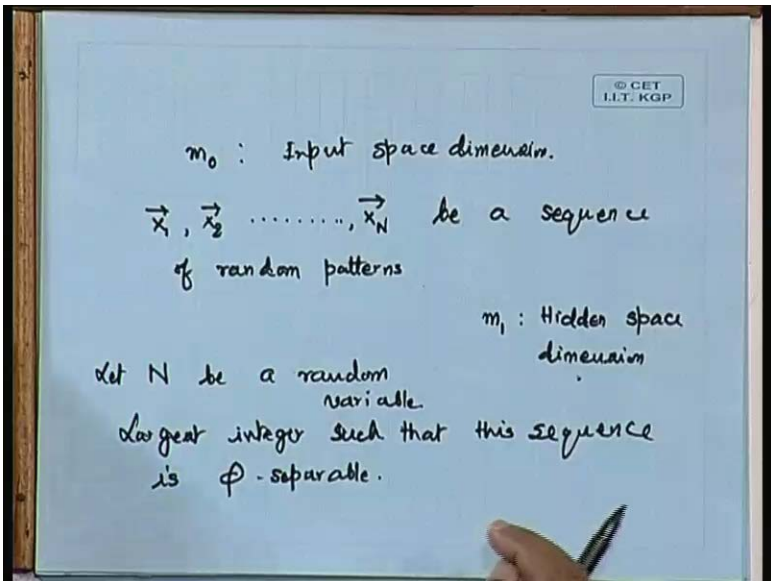
So, this will corresponding to phi 1 equal to 1 and phi 2 equal to 0.1353 would correspond to the pattern 1 1 and am I correct. And this one will become phi 1 equal to 0.13 and the phi 2 equal to 1 which will make it actually 0 0. So, now we will see that this is the space and now I can consider a separability by simply imagining a lying like this. So, the pattern which was inseparable in the original x 1, x 2 space, inseparable means linearly inseparable of course, but they are now becoming separable in the phi space.

So, this is the example that shows us that, it is simply by non-linear mapping what we have done is nothing but, mapping from the x space to the phi x space. So, by simply doing this non-linear mapping using Gaussian as a function, we could map it into a space where we could find a linear separability. So, that means, to say that the function that we have designed is basically a, so in this problem we find that there is a phi separability, we were talking about a phi separability in the last class.

So, phi separability is demonstrated by this example and now any questions pertaining to this. Now, we go over to the discussion on the separating capacity of a surface. Now, you see after all, what happens is that finding line in this case it becomes a simple line, a linear separability for three dimension it becomes a plane, for multidimensional it becomes a hyper plane separability in the phi space. Now, the thing is that it essentially transforms to a non-linear separability as far as the input space is concerned.

Now, the question is that if you keep on increasing the number of patterns let us say, you are feeding in the patterns as x 1, x 2, x 3, x 4 all these in the form of vectors mind you, so we have got an m 0 dimensional input space let us say, so what we have is m 0 dimensional input space.

(Refer Slide Time: 20:45)



So, this is the input space dimension and we take a sequence of random patterns and those, sequence of patterns are x 1, x 2 etcetera, up to x N, so this is a sequence of random patterns. Now, we would like to know that if we keep on increasing this N, then what is going to happen with the probability of separability, you remember that in the last class we were considering the probability of separability.

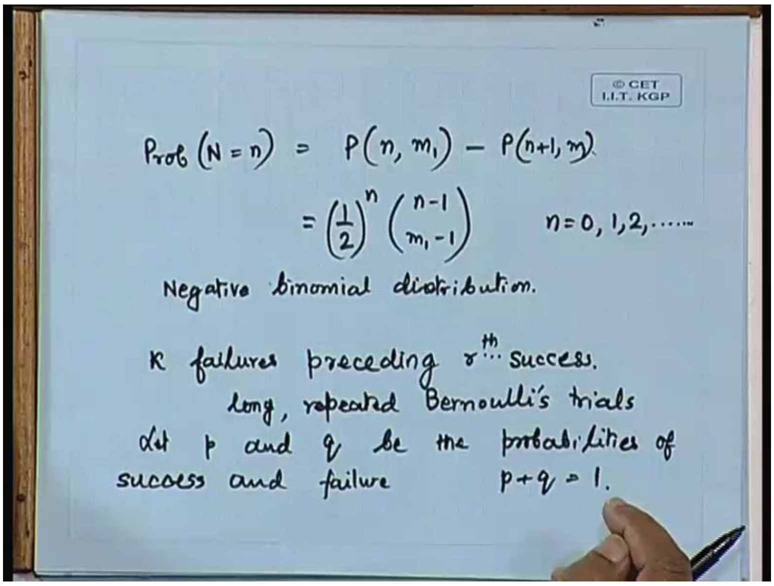
And probability of separability in the last class, we studied from a different standpoint there we were keeping the m 1 that is to say, the dimensionality of the hidden space, so m 1 is taken as the hidden space dimension. So, we there considered the dimensionality of while proving Cover’s theorem we were taking the dimension, the hidden space dimensionality to be one of the variables.

And then showed that as, m 1 increases the probability of separation improves, but the thing is that if let us say that, we have decided our structure that we are going to have freeze it, that we are going to have m 1 as the hidden space dimension with us. Then, the question is that from the m 0 space, so then the question is that how many such patterns can we have, if we increase n are we going to increase the are we going to improve the separability or are we going to make the separability what.

So, what we are going to do is that we are going to take now, N itself to be a random variable, so now we define N as a, let N be a random variable and how is N defined this is defined to be the largest integer, so this is defined as the largest integer such that, this sequence is phi separable. So, how we are going to formulate the problem, we are going to take the probability that this capital N is equal to n, we just write down and then you will be able to know.

Now, in the last class we had shown you the probability expression. So, you can have you can keep that in a in your reference.

(Refer Slide Time: 24:10)



And then, one can work out that the probability that N is equal to n will be given by the probability n comma m 1. What is n comma m 1, n comma m 1 means that there are n patterns, which are getting mapped into m 1 dimensional hidden space. This minus p of n plus 1 comma m obviously, which is going to be higher p n m 1 is going to be higher why, because if you change from small n to small n plus 1, if you are increasing the patterns then the separability probability is decreasing in the n.

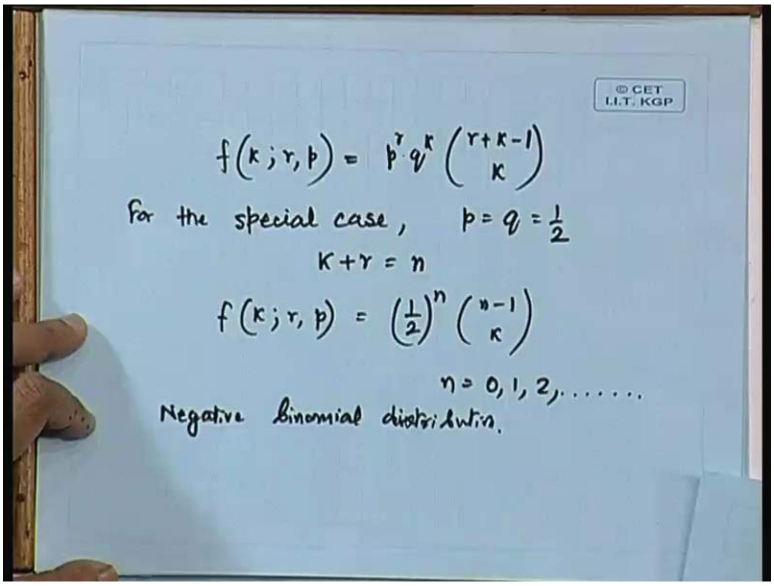
So, this in fact if you are substituting expressions that we had shown in the last class, for p for p n m 1 and for p n plus 1 m 1, then we can write down that this becomes equal to half to the power n and the combination n 1 c m 1 minus 1. And this we have to this will be defined for n equal 0, 1, 2 etcetera, so this is the probability that we are getting and this expression that we have got has got a negative binomial distribution.

This distribution is a negative binomial distribution, in fact the negative binomial distribution is the one, that we are normally getting from the from repeated long repeated Bernoulli trials. So, what we have to consider is that, you can consider any binary event experiment let us say tossing of a coin that is what you take and you take the probability that k failures in a coin tossing experiment. Let us say, you have k failures preceding r th success preceding r th success and what is so that means, to say that it is a long repeated binary trials.

So, we are keep on repeating this trial, Now there is a first success then again some failures, then again the second success, again failures then again, the third success like that it goes on, so it is coming from a long repeated Bernoulli trial. Infact those who are interested should refer to any good book on the probability and statistical theories, where this will be definitely a presented, so what we have there is that, let us say that we define two probabilities p and q.

So, let p and q be the probabilities of success and failure, so necessarily what we are going to have is p plus q should be equal to 1. And in this trial of experiments the binomial distribution, the negative binomial distribution will be defined as follows, so this will be defined as...

(Refer Slide Time: 28:14)



F of k r comma p, now this is k failures, r th success and p is the probability of the success. So, this distribution is given by, p to the power r, q to the power k and a combination of r plus k minus 1 k. So now, if we take a very special case, so for the special case like this is nothing unrealistic a fair coin toss, so for a fair coin toss we are going to have p is equal to q is equal to half.

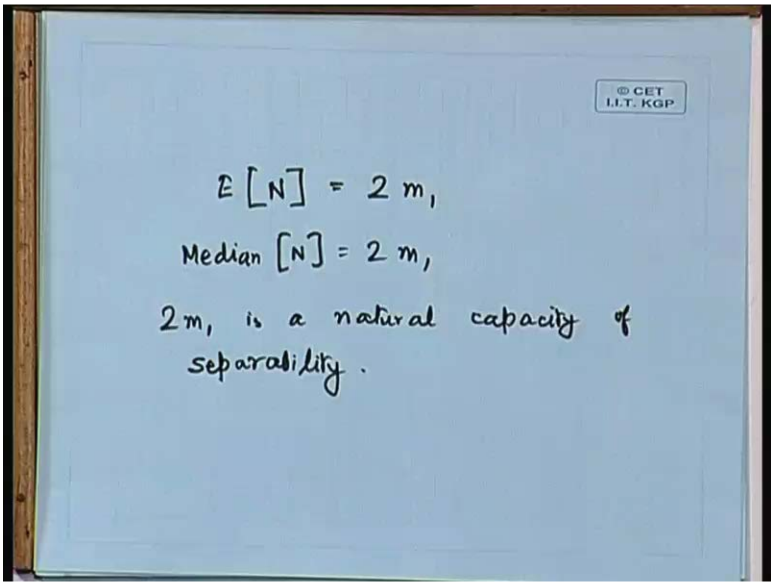
So, when we have p is equal to q is equal to half and as a special case, we take that k plus r is equal to n, we take k plus r equal to n. Then what it comes to is f k r comma p is equal to half to the power n you simply get that, because you have p and q both to be equal to half, so it is half to the power r plus k and r plus k is equal to n, so, we get half to the power n. And here, because r plus k is equal to n, we are getting a combination n minus 1 c k and this we have to do for n equal to 0, 1, 2 etcetera, so this expression is nothing but, a negative binomial distribution.

So, what we have got, as the probability of N equal to n, n is a negative binomial distribution. So, what essentially it means that we are varying this n, we are keeping this n as a random variable and we are getting a very interesting distribution, we compute n equal to 0, 1, 2, 3 like that. And we get some kind of a distributed, some kind of a distribution pattern and n is considered to be a random variable. So, definitely by from that distribution that we get it is possible for us to find out that ,what is the expectation of n, because that is very important.

Expectation of n, if we can determine from this set of statistical experiments, then from that expectation of n we can decide that, if that expectation is expressible, in terms of m 1, where m 1 is the dimensionality of the hidden neuron. Then, it is possible for us to say that, the maximum number of n that you should have is that the maximum number of pattern, that you should feed should be equal to 1.5 times say, or 2 point say, or 2.5 times.

And whatever, it works out form the expectation, in fact, the expectation and the median of the random variable n works out to be to times m 1, If we go through this mathematics then, we are getting the expectation of N out of this distribution, out of this negative binomial distribution.

(Refer Slide Time: 31:45)

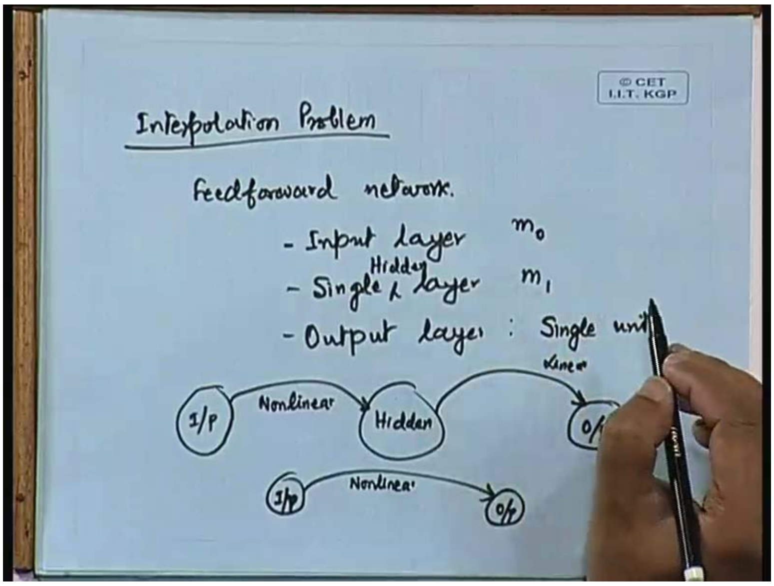


We are getting the expectation of n to be equal to 2 times m 1 and even the, median of N is going to be 2 times m 1, so that means, to say that we can definitely say that, 2 times m 1 is a natural capacity of separability. So, in Cover's theorem, we were considering separability from the standpoint of variation of m 1, the very fact that we wanted to show that the higher dimensional space results in a better separability.

Here, we are calculating that fixing of m 1, we are calculating that what is the capacity of separability in terms of it is number of patterns, where 2 m 1, 2 times m 1 becomes good number. Now, we having done that having know about the separability aspect, now we are in a position to begin the basic fundamental approach to the radial basis function. And again, I repeat what I have been saying, over the last 1 and 2 lectures is that essentially for radial basis functions, we are looking at it exclusively from an interpolation point of view, the whole problem is an interpolation problem.

And let us in order to formulate that interpolation problem in a nice mathematical way, we define a feed forward network.

(Refer Slide Time: 33:54)



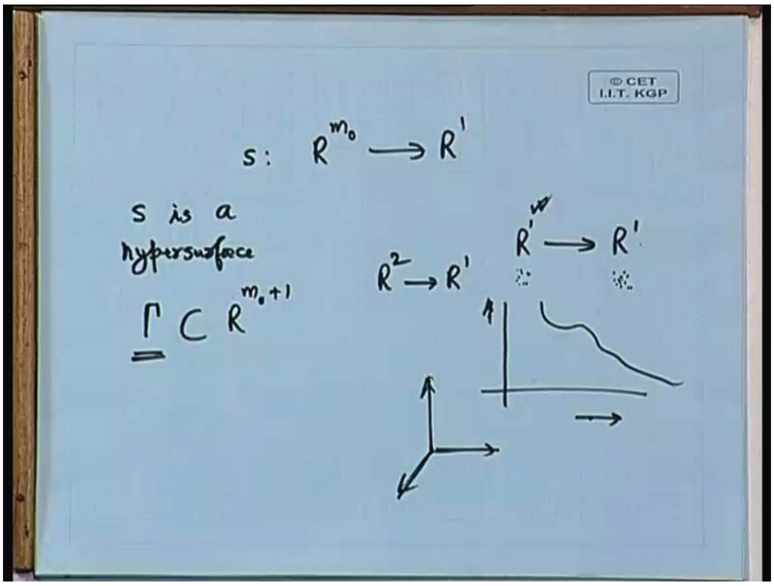
So, we talk about the interpolation problem now, and there we take a feedforward network. So, we consider a feed forward network for, which we are having input layer a single hidden layer and an output layer and just for our is of analysis and simplicity and without any loss of generality. We take that there is only, 1 neuron in that output unit, so there is, so as the output layer, we are having only a single unit that is, just for our simplicity that we are going to consider.

So, essentially what this network is going to do is that from, the input space let say this is the input space it is getting mapped nonlinearly into the hidden space. So, this is a non-linear mapping from the input space to the hidden space and then, from the hidden space to the output space there is going to be a linear mapping. So, in output we will be just combining this outputs of phi's, phi 1, phi 2, phi 3, phi 4, whatever we get.

Those, real values we will be weighting up w 1 phi 1, w 2 phi 2 like that and, we will add it up, so we will be generating a linear output. So, it is essentially a mapping from input to the hidden, hidden to the output, combinedly looking at it is definitely a mapping from the input space to the output space. And combinedly it is of course, a non-linear mapping is not it, so it results in a non-linear mapping from input to the output effectively.

Now, the inputs dimensionality we are considering to be m 0, the single layer I should have written single hidden layer, so please note that it is a single hidden layer. And this hidden layer is having m 1 number of neurons, so it is dimensionality is a 1. So, what essentially means is that from the m 0 dimensionality input space, we are mapping into the output which is a single unit. That means, to say that what is the output dimension output dimension is simply one.

(Refer Slide Time: 36:54)



So, essentially the mapping that we are considering for this network is, if we say that the mapping function is S. Then S is a mapping that is defined from m 0 dimensional real space to one dimensional real space, this is the input and this is the output. So, what is this S, what is the interpretation of this S, this S is a surface, this S is a surface of what dimension, can anybody tell me.

Surface of what dimension, m 1 no, there is no m 1, I have considered the entire mapping problem in an integrated way input to output. I am not considering the in-between mapping that we are doing to the hidden layer, hidden layer has performed it is job. And we are now looking at the whole radial basis function network from a black box point of view, there is input there is output.

So, there is no m 1 that is coming in, it is the simple m 0 dimensional to one dimensional mapping. In order for you to answer the problem in a better way, supposing we have got not here supposing, we have got some mapping function, which maps from one dimensional to another one dimensional space. So, let us consider a one dimensional to one dimensional mapping, you have got some real number and you are mapping it into another real number that is it.

So, there is a mapping function that is available, for those real numbers you are taking several real numbers. Several real numbers belonging to this space getting mapped into a set of real numbers in this space. So, definitely there is the function that is defined. And how are you expressing that function, that is also a surface fitting problem, but in this case it is a curve fitting problem. One dimension to one dimensional mapping is a essentially curve fitting problem, why.

Let us take that this is the domain of this, so this is the domain of this R, this function and then, this is the domain where it is getting mapped. So, this is our input and this is our output space. Now, this input to output is a mapping which is done by the definition of some curves. So, we are going to have curve may be a curve like this which is there, and this curve is defined in which dimensional space, two dimensional space.

So, one dimensional to one dimensional mapping results in a two dimensional surface, two dimensional surface of course, be generous to a curve. So, it is a two dimensional surface space where this S belongs to. Now, just complicate the problem a bit, you define that you have some inputs in the R 2 space, two dimensional space which you are mapping to another space which is R 1 space. So, it is a two dimension at the input, one dimension at the output.

So, how are you going to represent that problem. You are going to draw three axis like this, where the original input space you will be showing with respect to this. And the mapped values that you are getting, that we will be showing in this axis, in the z axis and for the input space you will be showing in x axis and y axis. So, that the fitting problem that you have got there, the surface that you have obtained there is a three dimensional surface.

So, R 1 to R 1 mapping in a two dimensional surface R 2 to R 1 mapping in three dimensional surface, make it out three, it will be four dimensional surface, so make it m 0, it becomes m 0 plus one dimensional surface. So, that map is now to be imagine, so S is a hyper surface. And what is that hyper surface, it is the hyper surface gamma. We take the hyper surface gamma which belongs to the m 0 plus one dimensional real space. So, what happens to us is that, this surface that we are having is unknown to us.

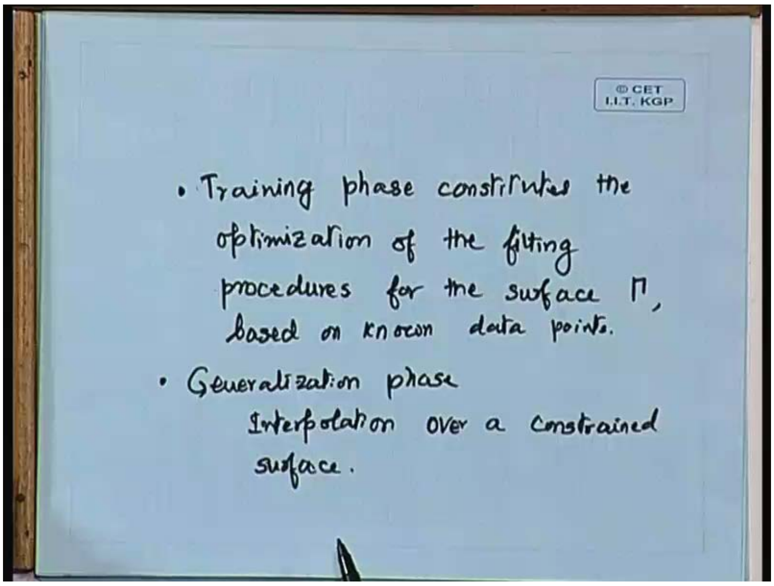
And how are we going to know it to the training. We are going to feed the training patterns, where essentially what we do is that where we definitely know, feeding the training patterns means that where we know definitely the input output mapping. So, the mapping function is initially completely unknown to us and we are going to feed the training pattern, where we know that we are telling them that.

This is the m 0 dimensional point corresponding to which the output that you are getting is this, we are specifying that. So that means, to say that the surface which was unknown for that, we are fitting some of the specific points. But, those specific discrete set of points does not constitute a surface, does it look like a surface. Like say for example, if you are considering that on this piece of plane surface, you just prick some needles needles of different heights.

Let us say that, here you prick a needle of long height here you prick a needle a lower height lower height. Again here you prick a needle of a higher height, like that you prick needles all over this of different heights. Does it constitute a surface, no if it has to look like a surface what do we do, we interpolate that surface. So, the job of the training phase is to interpolate that surface, why.

Because, if your surface is interpolated, in that case, only when we are feeding a test pattern, we come anywhere to this bed. And we want to determine that, what is the corresponding real value output, what is the corresponding output value. Then we simply look at the interpolated surface that we have got and determine it is value. So, now this interpolation, so this gamma, so about the gamma we can say two things That the training phase constitutes the optimization of the fitting procedure.

(Refer Slide Time: 44:02)



So, there is a training phase which constitutes the optimization of the fitting procedure. Now, the optimization is very important, because while feeding the training data I think that I have discussed this few times earlier also. That in our training set it is a, our training data its there could be some noisy points. So, during the fitting if there is some abnormal noisy points.

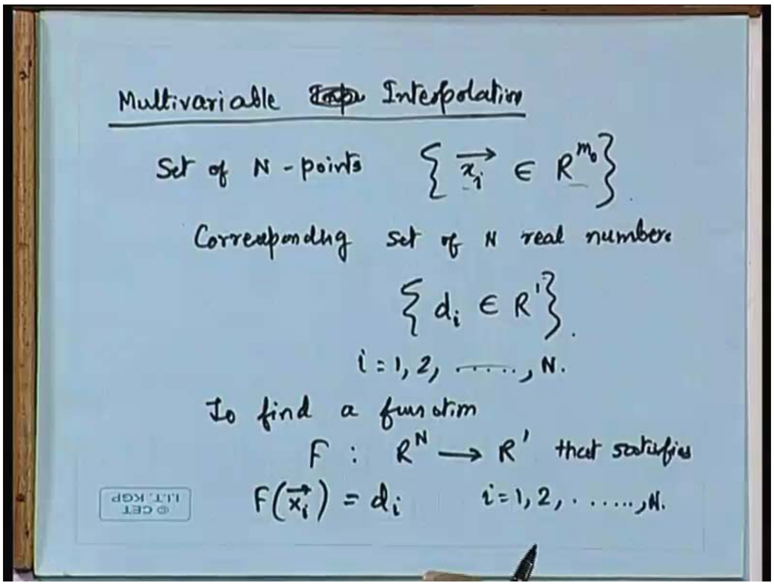
Like say for example, we have got fixed set of surface like this, but suddenly we find that, there is some point which is lying very high. Then, that is simply a point in error, so when we draw the surface we should not include that point. So, there is going to be some optimization of the fitting procedure, so only through optimization we will be able to get a surface. So, procedures for the surface gamma is the one that we are going to fit.

And this fitting is obviously, based on known data points. And then, the second thing is that in the generalization phase. The generalization phase can be looked upon as an interpolation between the data points performed over a constrained surface. Why are we calling that as a constrained surface, because we are fitting the known data points. Now, fitting the know data points means that, your surface has to always pass through these known data points, so it is becoming a constrained surface.

So, in the generalization phase it is the interpolation between the data points interpolation over a constrained surface. So, do not look at it from the usual characteristics way, like drawing the neurons, fitting the synaptic links and then, computing this summation of w 1, x 1.

Now, just a rethinking, RBF means that it is a rethinking where you are now thinking in terms of the whole problem to be an m 0 dimensional space to one dimensional space of mapping problem. And where the given points are constituting the constraints to the surface that you are going to fit. Now, that basically results in a multidimensional interpolation, multivariable interpolations.

(Refer Slide Time: 47:14)



So, we are considering multivariable interpolation, so we have got a set of n different points, set of n points which constitutes this set. So, the set of x i which belongs to the m 0 dimensional space, because m 0 is the dimensionality of the inputs. And we also take a corresponding set of n real numbers, so what happens we take the corresponding set of d i's.

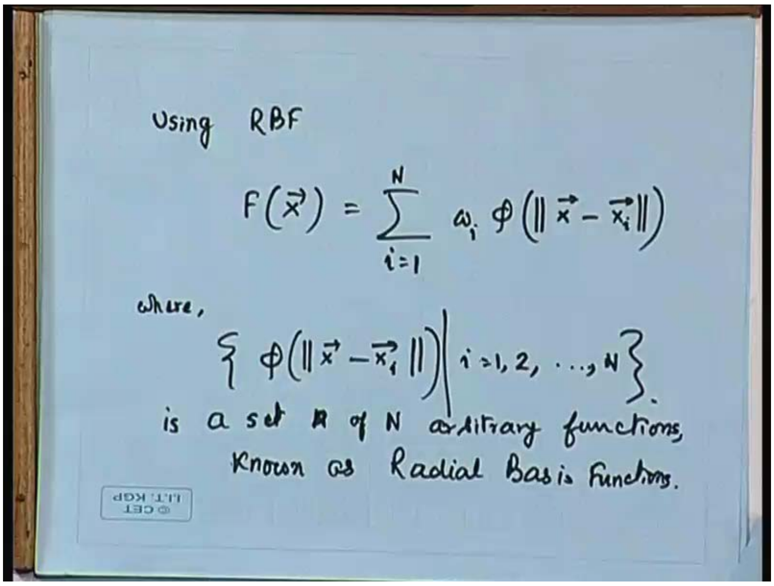
So, we are fitting x 1 vector, x i vector and correspondingly the output is d i which is a real number. In fact, because we are taking only a single neuron as the output, the output that we are getting is in one dimensional space. So, it is R 1, so d i belongs to the R 1 space and of course, here for both these things we have to know that our i could be 1, 2 etcetera up to N.

And the multivariable interpolation problem is that to find a function, so the problem is to find a function F. And what is the F, F should be a function that we are getting N such this things. So, N such patterns are being fit, so there are N number of points in this m 0 dimensional space. So, it is a mapping from this N dimensional real space to one dimensional real space, that satisfies which once all the training points, that must to say that all the constraint points should be satisfied.

So, we have to find a function like this such that, or fulfilling the constraint that at the point x i, we must be having what F x i is equal to d i, why hesitate to tell. Because, it is after all a mapping from the x i to d i. So, the function, so it is d i is equal to F x i simply and we do not know that what that function is. That function is nothing but, the interpolating function. So, it is to satisfy F x i is equal to d i for i is equal to 1, 2, etcetera up to N.

And now look at the radial basis function. Now, in radial basis function what we are doing we are taking the phi functions for mapping from the input to the phi space. And then, all the phi outputs we are linearly combining together, in order to get the final output. So, we can express from a radial basis function networks.

(Refer Slide Time: 50:50)



So, using RBF we can obtain, the same function F x as F x is equal to summation of w i times phi. And we can say that phi whose argument is x minus x i that means, to say that we are taking different phi functions, which are centered around the different x i's followed. So, we are having the inputs as x 1, x 2, x 3, x 4 like that. So, we are considering various such phi functions which are centered around all this.

And we are having in fact n such phi functions and we are going to add up this that all this outputs will be linearly combined. So, we have got n such synaptic weights will be there, synaptic weights will be from the hidden to the output. So, there will n such and we will be actually mapping from the hidden space of n dimension to the output, which is of one dimension.

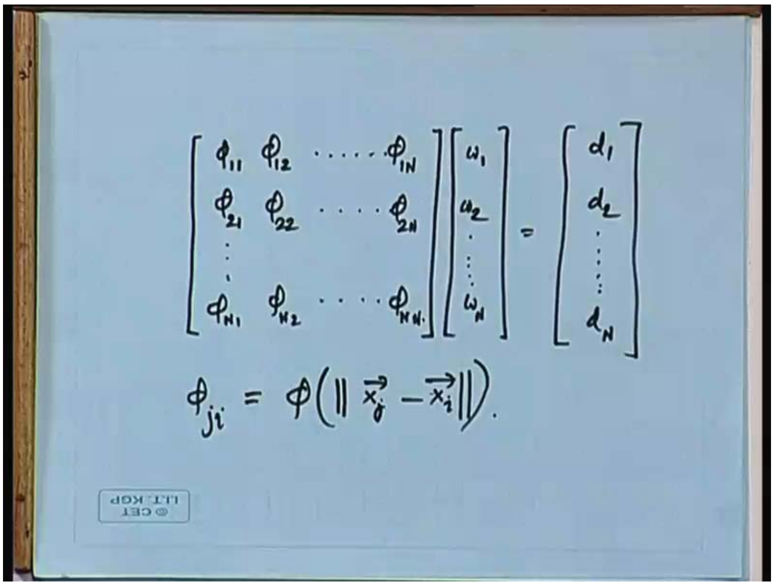
So, in this what is it that where the phi functions that we are having. So, the phi with argument x minus x i given that i is equal to 1, 2 up to N, is a set of N arbitrary functions. And what are these functions known as set of N arbitrary functions known as radial basis functions, so this is the way our function is defined.

Now, if we have to use the radial basis function for interpolation. Then what we have to import upon it, that this equation that we have got should be satisfied by this equation. Because, this equation is acting as a constraint to the surface interpolation that we are making, so this constraints should be put into this equation. Meaning what, if we are taking the first pattern let us say, if we are taking the x 1 as the input to it.

Then x if we have F x 1, F x 1 should be equal to d 1. So, d 1 should be equal to what summation of w i phi of x instead of x we will be making it as x 1, x 1 minus x i. And we have to compute all these n different outputs. So, for x 1 pattern we will be getting N different outputs, coming out of the N different radial basis functions that we have constituted. Now, what you do is that you take i is equal to 2.

So, that next time you are importing F x 2 is equal to d 2. Now, you substitute d 2 over here even d 2 also will be expressed as a weighted summation of all this N different radial basis functions. So, each point will be expressed as a summation of all this radial basis functions. And then, we can now write after putting this constraint, after applying this constraints it is possible for us to write these equations, reformulating this equation in a matrix form.

(Refer Slide Time: 55:17)



How, just look at it, it can be written as a matrix, a phi matrix which I am going to define shortly. Where let us write down first, Phi 1 1, phi 1 2, up to phi 1 N. Then phi 2 1, phi 2 2 up to phi 2 N and going all the way up to phi N 1 phi N 2 up to phi N N, this times w 1 w 2 up to w N. And this results in d 1 d 2 up to d N very simple to verify. What is d 1, d 1 is w 1 phi 1 1 plus w 2 phi 1 2 plus etcetera up to w N phi 1 N.

D 2 is w 1 phi 2 1 plus w 2 phi 2 2 up to w N phi 2 N, d N will be likewise. So, here the definition is that phi j i that we are considering is the phi whose argument is x j minus x i norm of this. So, here what is this x j, x j is the point for which you are calculating this function. So, I think because the time is ending today and long time back I just shown with that sign phi.

I have to stop, so here this is the formulation that we have got in terms of the... After putting the constraints of the radial basis function and then, we will be discussing more on this in the coming class.

Thank you.