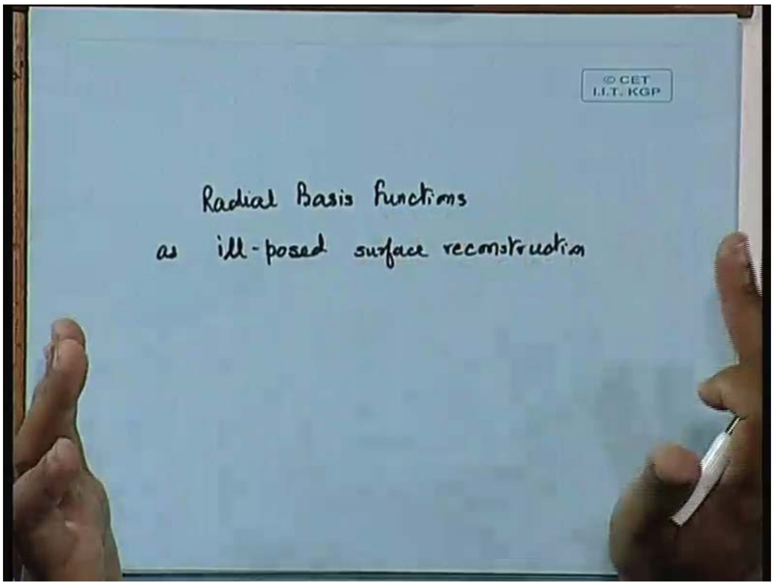
Neural Networks and Applications

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

Lecture - 26
Radial Basis Function as ill- Posed Surface Reconstruction

Today, we going to see another aspect of the Radial Basis Function. That is we would look upon the radial basis function as ill Posed Surface Reconstruction problem.

(Refer Slide Time: 01:10)

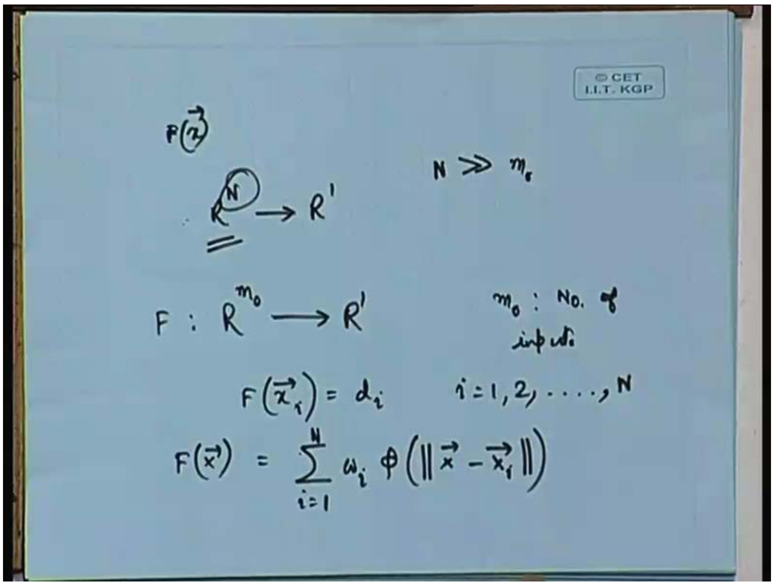


We will say why is it ill posed? But I think the aspect of surface reconstruction, we have already given some introduction to it. Because, the whole problem, we are having a new look at the training of a multilayer perceptron in the sense that, we are looking at it exclusively from the point of view of a surface fitting. That is to say, that when the input patterns are given and the outputs corresponding to that a specified, then essentially the mapping that is done within the network.

That one way are using for this, so that one we are using as a function which we obtain by training. And then, we are interpolating upon it, because the interpolation only would lead to the generalization. In the last class, we were talking about the interpolation problem as such and we were making some mathematical formulation of it of course, thinks where done in a bit of hurry. So, I thought that maybe it is good to spend a of bit of time about what we were doing in the last class.

So, you remember that, in fact what we had done in the last class is like this, that while talking about the interpolation problem. We said that we have been given these information write that, we will have to reconstruct a function f x.

(Refer Slide Time: 02:59)



In fact, here one correction that we have to make is that, we was saying that we have to find a function, that maps from R N space to R 1 space and this is wrong. So, I told something wrong it this N is in correct, because essentially what you are looking for is a mapping from R m 0 space to R 1. Because, m 0 is taken as the number of inputs or the dimensionality of the input vector, whichever way you refer to.

So, it is R m 0 space at the input and then we were considering the case of single neuron. So, the mapping that we are looking for is essentially, it is this mapping function F only that we have going to find out, which is essentially from R m 0 to R 1. So, please make corrections in your notes that it should not be R N, in fact immediately after the class one of the student, so where present here happen to point out this. So, I just remember to make a point to make this correction that it is a R m 0 to R 1 space mapping.

Now, what we are given with is all these points, that we are feeding the training patterns and training patterns means that essentially we have got a set of all this exercise as per inputs. And corresponding to this F x i, is means F x being the function to be realized to the mapping function to be realized. We will be having F x i is equal to d i, for i is equal to 1, 2 up to N assuming, that we are feeding N as the total number of patterns.

Now, this means to say that all this F x i is equal to d i is going to be fed as the constraints. And as we had written down that we have to, if we have to modulating terms of the radial basis function network. Then we are going to write it as F x is equal to summation of w i phi and the phi is argument is the norm of x minus x i. And here we are adding it F for i is equal to 1 to N. So, this is acting as a linear summation, weighted summation just like the way we are doing in any neural network.

So, that means, to say that after the input space x is mapped in to the corresponding phi space, then what we are doing is that after the x to phi space mapping, we are just linearly combining all the phi space outputs. And then, having that as our final output, that is exactly what we are doing. Now, if we look at this equation, then you can see that, it means to say that since we have got x minus x i over here and what is this x i in this expression, in this expression x i acts as a center of the basis function.

You see, we were taking the examples of the Gaussian basis function in the last class, so if we are taking the Gaussian basis function, then giving an x i like this would mean that you are shifting the origin of the Gaussian function. That is you are making a shift by adding the x i to it, so this x i is acting as that, so that means to say because we have N such patterns. We have got N such different x i's, which means to say that as in to say that we have got N different radial basis functions.

We have got N different radial basis functions, which we are obtaining as x minus x 1, x minus x 2, x minus x 3 up to x minus x n. So, N patterns that is what we are having and we are assuming that we have got n such radial basis functions. Not that we always required that, in fact you remember that the last example which we are taking was that of an exclusive OR. And in that exclusive OR example we were taking four pattern, we all together have four patterns.

The set of 0 0, 0 1, 1 0 and 1 1, but we were considering in the last example, only two of those radial basis functions where we made the radial basis function I think center around 0 0 and another be to which is center around 1 1. We did not take the radial basis functions center around 0 1 or 1 0, we avoid it taking it, in fact we could have taken that also. If we have take, we did not take it because we could solve the problem.

We could make it linearly separable in the phi space, but there would not have been any harm, if we had taken four such radial basis function also, because there are four patterns. Four would not have made it any harm, in fact if we had transform that in to a four dimensional space instead of a two dimensional space. So, from the two dimensional inputs space to four dimensional phi space, if we had mapped and that to nonlinearly.

Because, the basis function itself non-linear, then our separability would have been even better than that because then, we would have got both the advantages of non-linear mapping as well as the higher dimensionality space mapping. So, here indeed we are writing this equation or writing this expression with the thinking that, normally our N that is the number of patterns that we are feeding should be sufficiently greater than that of m 0.

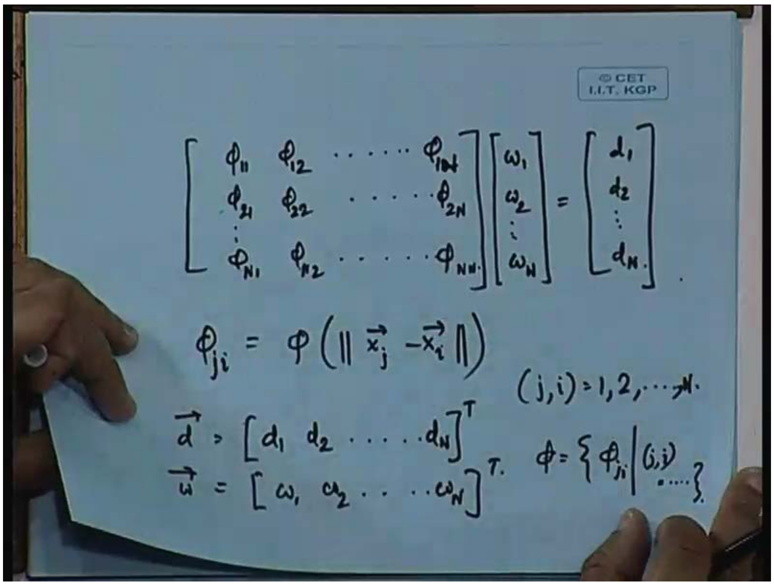
So, which means to say if we are having in such basis functions in such phi functions, each of these are center around all this x i's. Then naturally we are mapping from the m 0 space to an n dimensional space, which is a mapping in to a higher dimensional space. So, in other words, I may let us consider the there are n such basis function and we have obtained the F x like this.

Now, this we can in fact, after putting the interpolation conditions, because this F x i is equal to d i, this is a constraint or the restriction that we have to impose upon this equation. Because, at all these points we know, for when x is equal to any of this x i's, any of this given x i's we know that what it is output is going to be d i. So, simply what we will be doing is that we pick up the first, let us say that out of this x i we pick up the x 1 first.

So, we will put a F x 1 in this equation F x 1 is nothing but, d 1, so we will be having an equation like d 1 which will be realized as the summation of w i phi x minus x 1. We will be using, no will be using x minus x i only, all these N basis functions, N radial basis functions will be used in order to determine this. So, we will be having all these weighted summations added up, all these weighted summation of N different radial basis functions will be added up to form F x 1.

And like that we will be having F x 2, F x 3 like that or rather in so many words computing the d 1, d 2 etcetera. So, if we do that then the resulting matrix equation, I think we have written down in the last class.

(Refer Slide Time: 10:48)



But, no harm in repeating that just for our ease of reference we can write down here, that we are as if to say that having it as phi 1 1 up to phi 1 N, Phi 2 1 up to phi 2 N, phi 2 2 phi 2 N and then, lastly phi N 1 phi N 2 to phi N N, does not mean that we have got now N square number of basis functions. So, let us and not make mistake, we are not having N square number of basis functions.

We are having N number of basic functions only, but the notations, but now we are using the double subscripts, so the meaning of this double subscript should be clear to you which I am going to write, let me complete this. So, this w 1, w 2 up to w N is easily understood by you, so they are the N number of free parameters, because all these N outputs of the radial basis function space. They have to be linearly added, so this becomes equal to d 1, d 2 up to d N.

Now, here the meaning of these double superscript, double subscript is that in this we define phi j i as the phi function of x j minus x i, norm of this and where j i both of them vary from 1 to N. So, you can see that our basis functions still remain as N such basis functions only, but you see that when we pickup x 1, then what we do. Supposing this x i we have fix to something, let say that x i I have fix to we are taking the very first basis function.

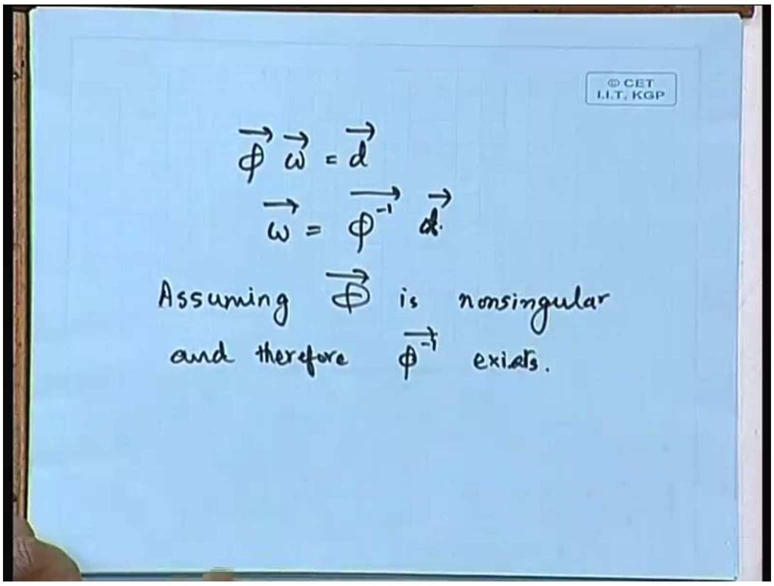
So, x i in this cases equal to x 1 and supposing we take the h first of these, so then supposing we take this F x 1 the first one, so F x 1 means that we will have to add the responses of this from all these. So, that means, to say that we have to add up that x 1 minus x 1, x 1 minus x 2, x 1 minus 3 like that all these N outputs will be added but, because we are having x 1 minus x i.

And similarly, we will be having while computing F x to we will be having x 2, will be having x 2 minus x 1, x 2 minus x 2, x 2 minus x 3 like that. So, in a sense we are making the computation of this N number of times, so in effect we are forming and N by N dimensional matrix. So, please do not mistreat that is it is as have to say that I am realizing N square number basis functions.

The basic functions still remain as N, now what I will do is that be define, we just want to represent this equation in a compact mathematical form. So, what we do is that we define d as the vector d 1, d 2 up to d n, the transpose of this and the w vector will be defined as w 1, w 2 up to w N again the transpose of this and phi we have already defined. Phi will be what, phi will be the set of phi j i's given that j i is this, 1 to N I am not writing the whole thing again.

So, this is the definition of phi, so this is the d, this is w, so simply this equation which I have a return in the matrix form is now to be rewritten as phi, where phi is a matrix in fact, phi should be a matrix, so I am putting this notation.

(Refer Slide Time: 14:41)



So, where phi multiplied by w, basically rewriting this whole equation is equal to what this equal to the d vector, so phi times w vector is equal to d vector. Now, what we are going solve for the network, we are going to solve for the what w, we are going to solve for w.

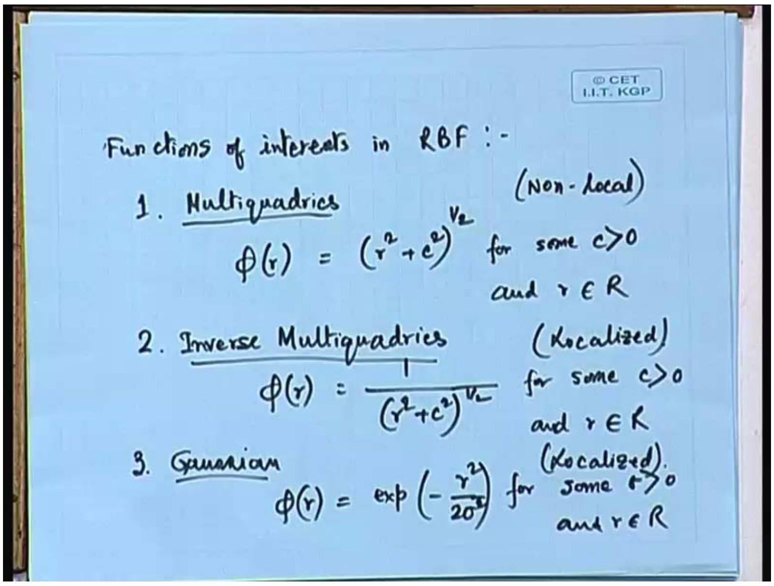
So, how are we going to solve it we can solve for w as this, the w should be equal to phi again you see that here phi is becoming a square matrix, N by N matrix. And assuming that the inverse of this exists, that means to say that assuming that the phi matrix is nonsingular, we can write that w is equal to phi inverse times d. So, what do we need very simple, we have to determine this w's, an in order to compute this w's what we need is this d's, which will be given to us all the training patterns will be given to us.

And then, we have to compute this phi inverse phi inverse means, basically we have to have the knowledge of phi's. And knowledge of phi's is not difficult to obtain, because we have got N different basis functions already defined for us, N different basis function with each centered around the given patterns. And then, we compute the h matrix of that, the phi matrix and then, we are taking the inverse of this.

So, this is possible assuming that phi is nonsingular and therefore, phi inverse exists. Now, this is a very important question that how do we assure that phi is nonsingular, in fact in recent years, good amount of work has been done related to the radial basis function. And there are some typical functions which have been found out to be nonsingular, if we are taking those functions as our basis, then the phi matrix that we will be getting.

Then those phi matrix is that we obtain is going to be nonsingular. And let us have a list of some of those functions which are of interest to us that means, to say the functions which can be used as a radial basis function. And some of these very popular functions are.

(Refer Slide Time: 18:01)



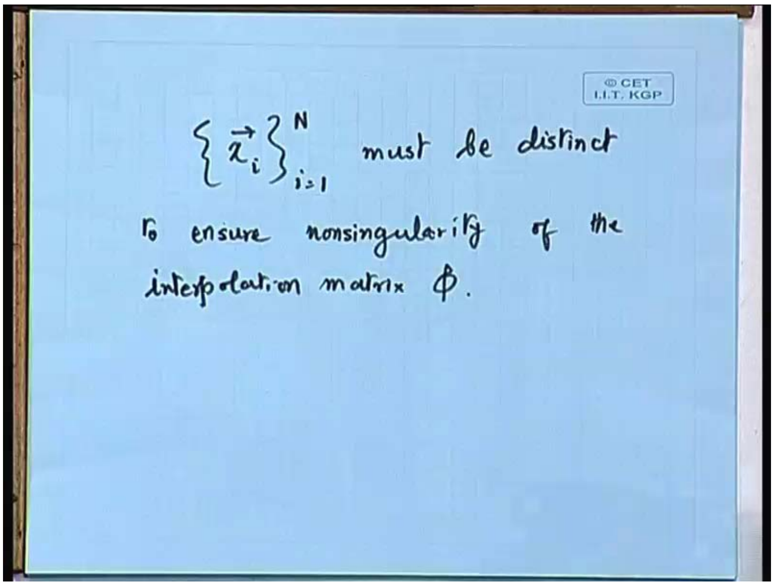
So, we can write down functions of interest in RBF and those functions are multi quadrics. Now, multi quadrics expression is given as phi r is equal to r square plus c square to the power half. Where c is some constant, let say for some c which is greater than 0 and r is number that exist in the real space. We can say as just the r space one dimensional real space, but of course the same definition could be extend for the multi dimensional case also.

They are what we have to replace is that, there r will be replaced by the vector that we will be feeding as inputs, so this is multi quadrics. Then another function which is used as a radial basis function is the inverse multi quadrics. And the inverse multi quadrics is given as phi r is equal to 1 by r square plus c square to the power half. For some c greater than 0 and r is less than and r belongs to this r space, the real space and then, the third is of course, the very popular Gaussian basis function.

So, for Gaussian basis function in the r space if we write, then it is phi r is equal to exponential to the power minus r by 2 sigma square minus r square, as what a realized that some mistake, so for some sigma greater than 0 and r belonging to r. So, these three functions they are used in the RBF and the reason why we can use them in RBF is, because if you are composing the phi matrices out of this.

Out of any of these functions, then phi inverse is also existing is seen to exist. And because of that, these are this can be used for the radial basis function applications is, the only restriction and I think it is quite logical also is that the set of x i's that your choosing, the set x i's that your choosing they must all be distinct.

(Refer Slide Time: 21:18)



So, I can say that if we take x i, for i is equal 1 to N we are saying that all these things must be distinct. To ensure nonsingularity of the phi matrix and phi matrix is given and name, what name would you think should be more suitable to the phi matrix, this name that phi matrix. It is use for what, phi matrix any good name for phi matrix that we can give.

Student: Radial

Radial basis function matrix, after all we are using it for interpolation, so we can term it as an interpolation matrix. So, we can say that this must be distinct in to ensure nonsingularity of the interpolation matrix phi. Now, we can have a quick look at this functions, that we have already defined multi quadrics, inverse multi quadrics and Gaussian. Some of the properties which becomes very obvious out of these three functions is that look at the multi quadrics.

If you increase r, as r tending to infinity, what is happening phi r also tends to infinity, so that means, to say that this is not a local function it is increasing. So, we can say that, this is a non-local function, in the sense that it becomes unbounded as we keep on increasing this r. Whereas, inverse multi quadrics and Gaussian there share a common property, as you make r tending towards infinity, you can see that the phi r function tends to be 0.

In fact, here this c square is there, but as r tends to 0, it is 1 by c, we have a very small quantity, normally c is not choose in to very high. So, you can then have the phi r a tending to 0, so this is more a localized function and even Gaussian also is localized function. Now, it is seen that for this localized functions, which are realized out of the inverse multi quadrics and Gaussian; the phi matrices which are realized out of them, out of these to functions they happen to be positive definite.

And so obviously, they are phi inverse exists, now this one multi quadrics although it is not positive definite. Because of this property that as r tends to infinity, we find that it is the function is also tending to infinity. So, it is not positive definite, but even though it is not positive definite still the phi inverse exists, and one can use these radial basis functions, for the interpolation problem of a here.

So, this is all that I want to talk about interpolation, but let us now see that whether such kind of interpolation is always going to be good for us. You see now what we are trying to say is that, you have to train a network with m 0 number of inputs. You have to generate single output that is a example with which, that is condition with which be proceeded. You can generate multidimensional outputs also there is no harm, it is just a simple extensional that idea.

But, you have to do this kind of think by feeding N different number of patterns, where you are assuming that N will be sufficiently largess compared to m 0. And you will be using for N corresponding to N different patterns, you will be using N different radial basis functions. Now, mind you one very important think to note here is that, we have taken N patterns and we have taken N radial basis function.

So, the realize matrix that we had got is an n by n dimensional matrix that we got. And the nonsingularity condition etcetera, etcetera are coming for that n by n matrix, if you have chosen less than N, then such kind of a nonsingularity is not always guaranteed. We may be luckily finding out the reconstruction problem, we may be having somehow generate a linear separability in the phi space and feel happy about it, but all the time it is not possible.

So, think is that in that case what is to be ensure, that if there are N patterns unit N radial basis functions and I think the naturally thinking is that in order to have a good surface fitting, you should have as large a value as for N as possible. More number of training points, if you are having, in order to learn better what we do in real life, we read more and more example. But, very often we also come across the situation that learning too many examples is also not good.

Supposing, somebody is trying to learn mathematics, now he just sees some book where there are some worked out examples he studies that and then, he goes in for another book. They approach of that could be different may be one or two problems are solved wrongly. Now, if we sees too many examples, then may be is that times we will feel confused and ultimately when the student is asked to solve an unknown problem. That is student would find it difficult to answered to that problem.

Because, he has got lot of confusions in mind by taking to too many examples, why because he has picked up 5 or 6 books in mathematics and may be that 1 or 2 books are not really good books. Their approach might be different, they are examples are not good, so such 1 or 2 books are, we can say noisy books, so likewise if you are having too many number of patterns taken from the real life.

There will be some pattern which are noisy, but in this interpolation problem which we have discuss so far, what are we trying to tell. We are trying to tell that interpolation means that inevitably all the points that unknown to you the surface must pass to that. Now, there lies the problem, if we constrain always that through every training pattern the fitted surface must be passing. And we will be having the interpolation in between, then it not only learns the good patterns.

It also tries to learn the idiosyncrasies that are present in the pattern set, all the noise that is present in the patterns set will also be there. So, strictly looking at it from interpolation point of view is not the best think that we are looking for. We should also try to make it noise free and how do we make anything noise free, we try to is move then. Whatever surface we fit if after fitting that surface we try to smoothen it out, why are we smoothening, because we are not totally relying on the data set that we have got.

We assume that inherently there is some amount of unreliability in the data that we have got, so that why after fitting the surface we try to smoothen it out. That means, to say that during the surface fitting itself, we should try to evaluate that whether we have not only minimize the errors. See minimization of a error is one of the considerations, now definitely if you are fitting a surface, through a set of given points.

Naturally you have to have a situation that the error between the fitted surface. And the corresponding actual points if you take those errors square it up and add it up, make it as the some squared errors. Those some squared errors you can minimize that is what you have been doing for so long. But, we are saying that if you have to look at it from the point of view of avoiding such kind of results or our surface interpolation problem, getting affected by the idiosyncrasies of the data or the noise present in the data.

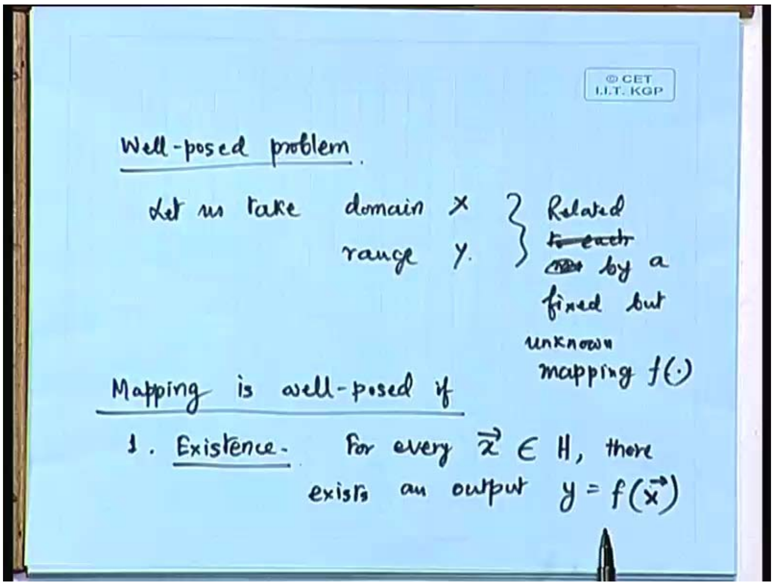
Then we should do something more, then we should in addition to this minimization of the error, we should also try to see that there is some aspect of smoothening that is to be done. Now, mind you because we have such kind of noisiness which are present in the data set itself. The problem that we are going solve, the surface interpolation problem that we are going to solve is not a well posed one, it becomes what is called as ill posed one.

Now, some people may not be very familiar with these two terminologies of what is mean by well posed and what is mean meant by ill posed. So, I would like to spend of bit of time, in order to make it clear in your mind that what is a well pose problem and what is an ill pose problem. And in fact, we are going to see that the surface interpolation is and ill posed problem, it can be reliably solve.

And in order to solve it reliably, we have to make and ill posed problem in to a well posed one by making some kind of transformation, which we will be discussing shortly. So, essentially this we are now looking at the hyper surface reconstruction problem, as an ill pose problem, but again digressing a bit that what is meant by ill posed. And in order to know what is ill posed, the definition that I would like to simply says that, whatever is not well posed is ill posed.

If somebody is not well his stated to be ill, so if some problem is not well pose that is an ill posed, so we must now that what is meant by well posed.

(Refer Slide Time: 33:41)



So, well posed problems definition, we can just put forward over about here. So, well pose means that let us take a, in order to defined it let us take a domain X and correspondingly we take a range Y in the matrix spaces. And they are related this domain on the range are related to a each other by a mapping function, so the domain a range related to domain. They are related by a mapping function, we should say by a fixed, but unknown mapping function.

We do not know that what exactly that mapping function is, we are determine. Now, having defined this, that it is just a mapping from the domain to the range, using N mapping function, now when are we going to say that this mapping is a well posed problem. The mapping is say to be well posed, if it for fulfill three conditions. So, mapping is well posed if three conditions are fulfilled and what are those conditions.

Number 1 is existence, by existence we mean that for every x vector which are belonging to a set H that means, to say that from the domain we pickup all these x's. And for every x belonging to H, if they are exists an output y and how is y because we are considering the mapping function to be f. So, we are going to write y as f of x vector, so yes.

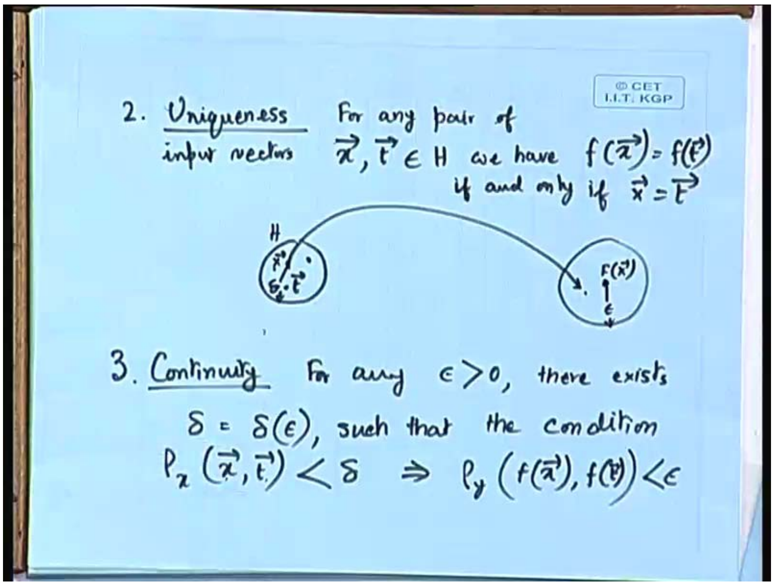
Student: ((Refer Time: 36:23))

x

Student: ((Refer Time: 36:25))

H, H is the space, H is the set from which we are choosing this x domain, set taken from the domain space that is. And is the output of that, so output will be expressed as f of x, so there is an existence of direct mapping simple. That you pickup any point, supposing we can the pick it using small illustrative diagram.

(Refer Slide Time: 37:05)



Let us say that this our input set, the range set, the domain set let say, the input set and supposing this is the output set, so we pickup any x. So, what the existent condition is saying is that, you pickup any x which is defined within the set and for every x there is a corresponding value in the y space that is so in the range space. So, if every x present it this set, this is the H set we obtain and output, then we can say that the existence condition is fulfill.

If for any one of the points like say for example, if for any of this points we cannot compute, given the mapping function we cannot compute any corresponding value of y. If the corresponding value of y becomes undefined, in that case we can say that the existence condition will be violated. Because then, the mapping from this space to this space for this particular point does not exist. So, the existence must be fulfill by all the points in this set, that is what this definition say.

H being the this set, that is the domain set and the second condition is written as the uniqueness condition, that is say that there is unique mapping and what is that translated in analogical terms. That for any pair input vectors, let us take any pair of input, we take one of the input vectors as let us say x vector. And another input vector we are taking as said t vector and all this input vector x, as well as the vector t they are belonging to the set H.

So, if for any pair of this input vectors we have f x is equal to f t. Now, f x is equal to f t means what f x is equal to f t means that as if, that x will be map to this point and t will be map to the same point. So, two of the inputs we take and they are getting mapped in to the same point, if that happens then it is a not unique mapping. So, for uniqueness condition to be fulfilled you take any pairing x and t, x and t any pairing you form out of all the patterns that have present in the H set.

If for all the mapping is distinctly different, then this is unique, so we have f x is equal to f t, under what condition f x is equal to f t will be there taking any pair only when x is equal to t. So, we have f x is equal to f t if and only if x is equal to t, so that is the uniqueness condition. And then, comes the third condition which is called as the continuity condition and what is continuity condition meaning, that if for any epsilon greater than 0.

Let us take it like this, let us say that we have got here point that is a which is f x, f x vector, we can write it as f x, because it is already a mapped point mapped from this x space this. So, f x a point we are taking and let us consider a region surrounding this point, this point in the multidimensional space of course, so this region is say having a radius equal to epsilon. Now, if for any such epsilon greater than 0, they are exists a mapping function delta is equal to delta of epsilon.

Such that, I am go to explain it after write it down such that, the condition rho x x comma t is less than delta implies rho y f x comma f t less than epsilon. So, this means to say what that, we are taking again a pair x and t from the H domain and we are mapping it in to the f x domain. And in f x domain we are having that us f x and f t, and their we are saying that here the bounding is that it is less than epsilon in this space. And if correspondingly we have here a bounding of taking x and t to be like this, so if this is x and if we take h region surrounding delta over here.

So, delta is the region that we are taking from the input space, then correspondingly it gets mapped in to the region center around this and with epsilon, so delta is going mapped in to epsilon. So, if from this to this such a kind of mapping exists, then it is continuity taking in a very, if the mathematical definition seems to be little confusing. In fact, it is not, but one can also physically image in also that means, to say that you take a small region.

And if that region entirely can be mapped in to a bounded region, you take the things within a bounded region in the input space, in the H space. And if it gets mapped into another bounded region in the output space, then the mapping is continuous. If that is not then, there is a problem, then it is lacking continuity, so if all these three condition that is to say existence, uniqueness and continuity. If all these three conditions are fulfilled, then only we are going to say that the problem is well posed.

If any of these three conditions is violated any, then the problem is not well posed and it is refer to as ill-posed problem. And let us see that the problem which is at our hand now that is to say the surface reconstruction problem, using the radial basis function, if that is a well posed one or an ill posed one. Now, the thing is that, first of all that let us see that when we are deriving the training set from the examples, that is a well posed.

Because, we are taking the examples from the physical process itself and that is a well posed problem, but when we are doing a surface fitting, then is it a well posed one and not, because the data may be continuing lot of noise. The data may be contaminated with lot of noise which are present in the example itself, but noises present in the physical process itself and that is why we got the noise in the examples.

So, physical process to example that mapping is well posed, even though noise is present in the physical process, noise is also present in this. But, the thing is that because of such kind noisy patterns, in fact it can be seen that in typical cases any of these three condition could be violated. You may be having some input points for which the output becomes undefined, if that is the case, then you are violating the existence condition.

Uniqueness definitely means that you cannot have a function to be multi valued, but because a here you can see that, you cannot have f x is equal to f t unless x is equal to t. So, uniqueness condition also can be violated, in the sense that it mean, so happen that, because of the noise x gets mapped to a point t also gets mapped to the same point. Although x and t are different, but x and t may be getting mapped in to the same point in the output space.

And noise can also drive you out of the continuity, so because of this all these three conditions, can I times we violated and that is reason, why the surface interpolation problem is always taken as an ill-posed problem.

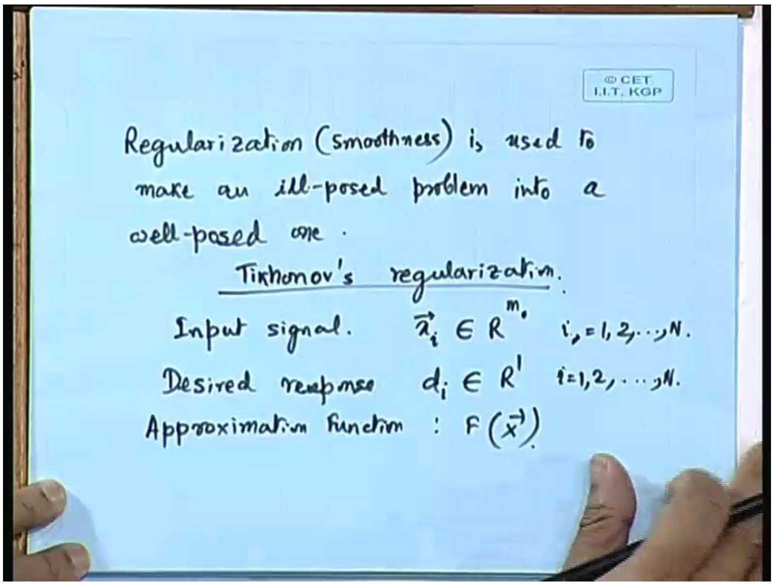
Student: ((Refer Time: 47:44))

That is the same thing yes, absolutely, whether you take delta here and epsilon here, it is to say that one bounded. Let us look at the this way that it is a mapping from one bounded space to another bounded space that is all, the mapping does not get unbounded. Now, the thing is that yes, because it is ill-posed, but we want to solve it and how are we going to solve it, so in order to solve it we should not look at it simply from an interpolation point of view.

Making a strict restriction that, it has to pass through all the points that we have defined in the training set. Let the surface be little more flexible, you have defined, let us say you have specified a bed of nails and you want to put some surface over it. Does not matter that this surface has to lie on the bed of nails, you can play with your surface, you can make it smooth. So, that the interpolation capability of the generalization capability of that is better and it does not get affected by noise.

So, that means, to say that we have to incorporate some kind of a smoothness condition also into it and that is what is called as the regularization.

(Refer Slide Time: 49:26)

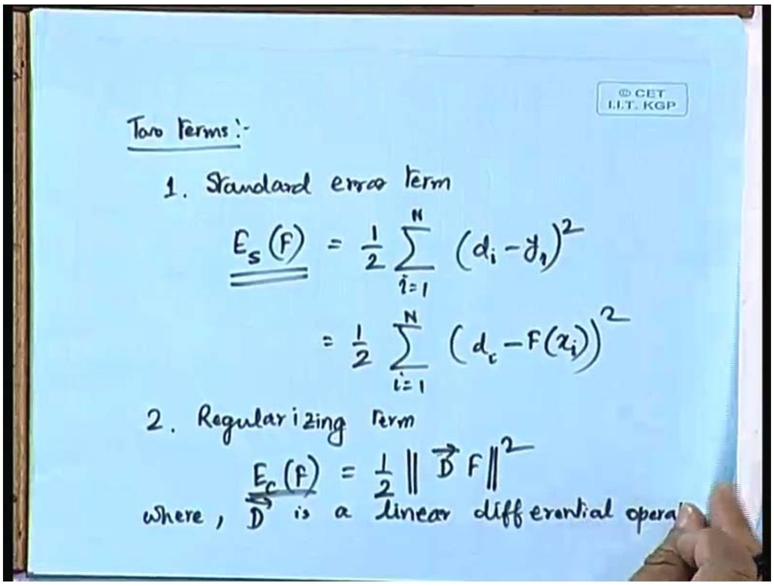


So, I can say that it is the process of regularization or regularization may appear to be little more sophisticated English, more common English smoothness, we want to make the surface smooth that is all. So, regularization of smoothness is used to make an ill-posed problem in to a well posed one. And in this case we are going to make use of the Tikhonov's regularization procedure, so we are going to make our mathematical formulation based on that.

So, how are we going to do it, so let us first take the input signal, so the input signal for us is the set of x i, where x i will be belonging to what R m 0 space. The input space is m 0, so it is 1, i is equal to 1, 2 to N and the desired response that we have is the set of d i is. So, d i and in this example we are having d i getting mapped in to just one, we are assuming only one output to be there, so d i getting mapped in to R 1 space.

Again i is equal to 1, 2 to N assuming that there are N patterns and we consider the approximating function or the interpolation function whatever you say. This approximating function is considered to be F of x, now Tikhonov's regularization procedure that involves two terms and what are those two terms.

(Refer Slide Time: 51:51)



Two terms of taken of regularization are number 1 is the standard error term, however, we are going to defined the standard error in our usual way, what we had already done earlier. So, that is the error term E s of F, we can write down as what half of summation di minus y i whole square summed up form i is equal to 1 to N. D i is the actual response and y i is the, no y i is the actual response and d i is the desired response. So, that means, to say that we are allowing the surface to deviate from the desired points.

So, only if it is so that all the y i are equal to d i's, then the standard error term would have been 0, but we may not want that, we may not wanted to pass through all the given points, we may allow it to deviate. So, that is why d i and y i are taken to be different assuming that the surface may not pass through the given set of point, so d i minus y i whole square adding it of ovaries is equal to 1 to N with would give us a standard error of fitment.

In fact, instead of y i one can always express it as F x i yes, that is correct, so it is d minus F x i, because F x we are taking as the approximating function. So, this is di minus F x i whole square this is the standard error term. And then, in addition to this where we are actually making it regular i's or smoothened is done by the second term which is called as regularizing terms. And the regularizing term is E c of F is equal to can you tell me that how would you regularize a function.

Student: ((Refer Time: 54:19))

Can I just have the answer from one of you

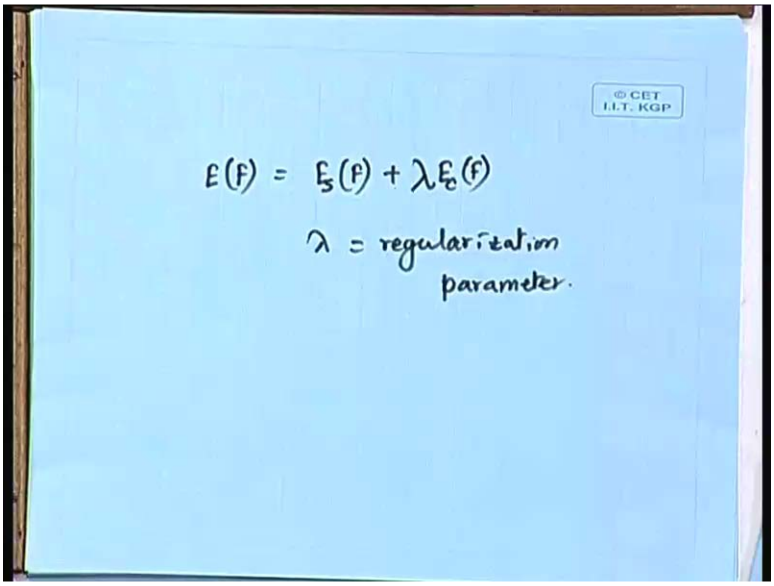
Student: ((Refer Time: 54:25))

Differentiate yes, and what that differential should lead to, the differentiation result should be minimum, because if the derivative is small in magnitude that means, to say that fitted surface is smooth. And this differentiation we have to do in how many dimension we have to do it in N 0 dimension. So, definitely we would like to do a multidimensional differentiation and just like the way the error term should be less, even the differentiation or rather multidimensional derivative also should be less

So, the problem is now to be reformulated in the sense that instead of like we are doing earlier, earlier we were only saying that minima the error that is all. Now, we are saying that not only minimize the error, but also have a smoothening on the surface. So, in order to import the smoothening in to consideration what we are doing is that, we are introducing this regularization term which is defined as half, again this half is to be used for a consistency.

Because, the measure of this regularization is the derivative operator D on the function F and D is very rightly written in the vector notation, because this differentiation of the process of taking derivative would have to be done in the m 0 dimensional space. And takes the squared norm of this, this becomes your regularization terms, so here we can explicitly write down that where D is a linear differential operator. And given these two terms E s and E c, one can combinedly put forward a measure like this.

(Refer Slide Time: 56:49)



One can make a combined measure E F, which is equal to E s F plus, shall I put E c F, but mind you one thing which is to be noted is that if I put E s F and E c F. That means, to say that as if to say, that I am giving to equal importance to the standard error minimization and to the regularization. Now, whether I should give equal importance to not is a debatable question.

So, to make it general on to place safe that one can play with it, let us about a term of E c F multiplied by lambda over here, where lambda is a regularization parameter we can say. So, we want to play with it, we want to play with this regularization parameter and how to fix of this lambda, that we will discuss in the coming class, because again I am exhausting my time limit.

Thank you very much.