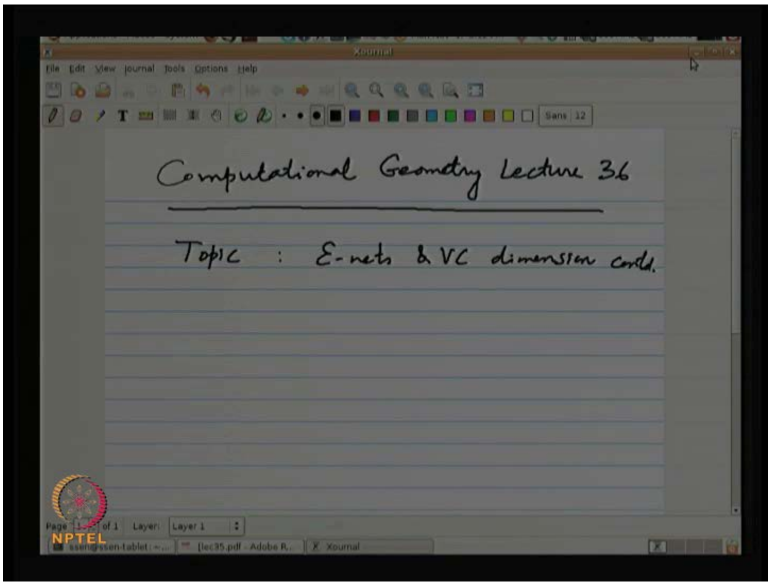
Computational Geometry

Prof. Sandeep Sen
Department of Computer Science and Engineering
Indian Institute Of Technology, Delhi

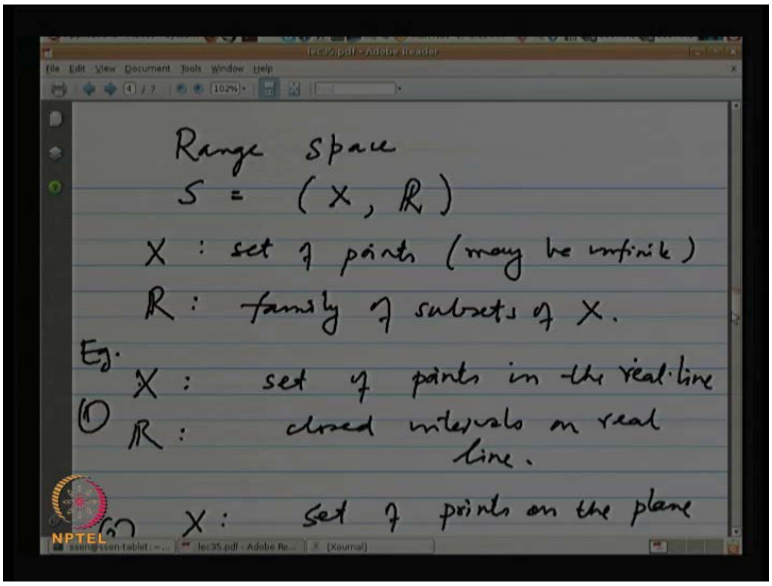
Module No. # 13
E-nets, VC Dimension and Applications
Lecture No. # 02
E-nets and VC Dimensions (Contd.)

We continue with our discussion on VC Dimension as related to construction of epsilon nets.

(Refer Slide Time: 00:34)

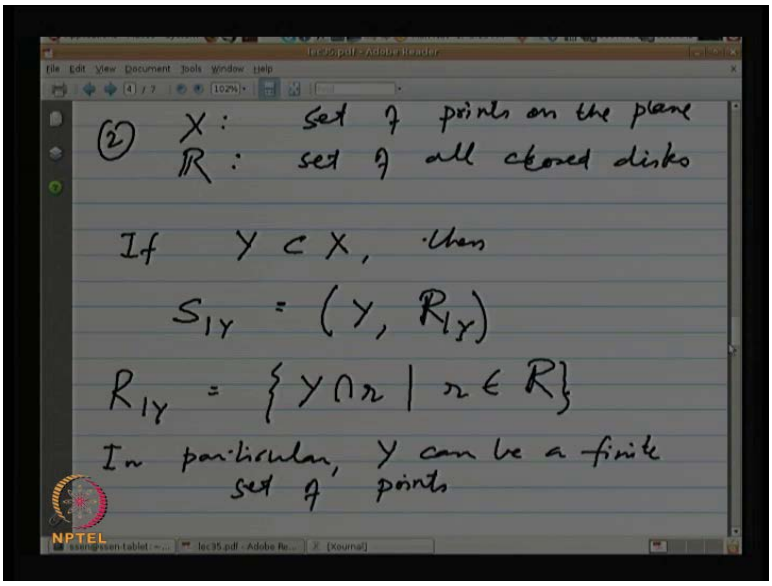


(Refer Slide Time: 00:53)



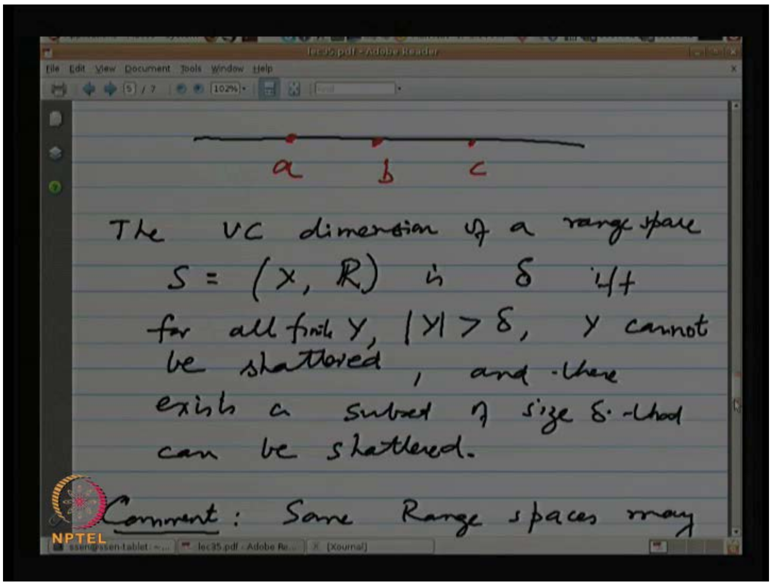
Just to recap, what we did yesterday well yesterday in last lecture; so, we went through some formal definitions of VC dimensions and range spaces, where we started with the started with the range space X, R, where X is a set of points and R is the set of family of subsets.

(Refer Slide Time: 01:09)



And the notation that we used that, when Y is the subset of X then the range space restricted to Y was defined as Y, ranges restricted to Y which is nothing but, the subset of subset of Y in that is and the intersection of Y with with the ranges of original range ranges right. So, this helps us take care of the situation, where Y is finite as in the original definition of range space X may not be finite. So, if Y is finite then everything becomes finite right, then you have the set of points to be finite, and also you have a finite number of ranges right.

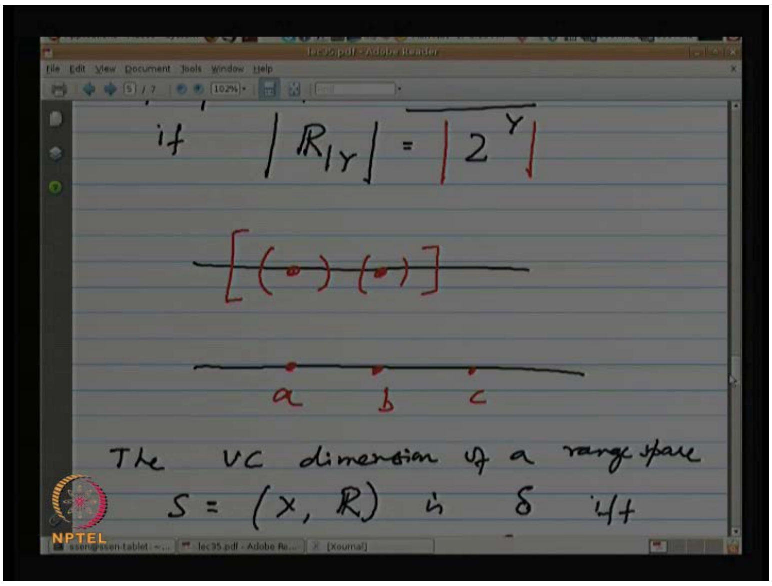
(Refer Slide Time: 01:59)



And then we had seen some examples natural examples of range spaces. And in particular, we went through this definition of VC dimension which was defined on the basis of what is called shattered. A finite set Y f cannot be shattered if well the other way around. So, a finite set of points can be shattered, if you have a subsets or the the number of ranges range spaces equal to the power set of the of the power set of points right.

So, every possible subset of the given set Y has a corresponding range. So, that is when we says is set of points is shattered. The maximum number of points that can be shattered is called the V C dimension of the range spaces.

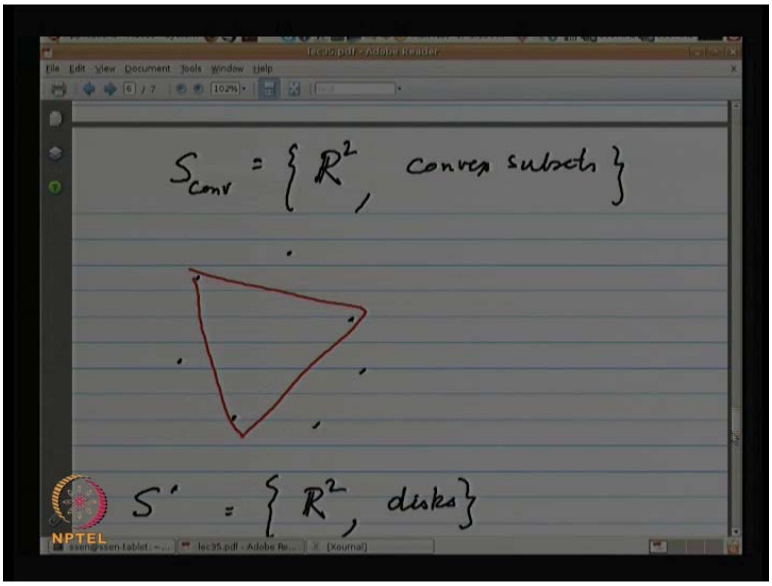
(Refer Slide Time: 03:06)



And we actually saw situations, where V C dimension well a simple example like you know intervals and points, the line is when we have you cannot shatter a instead of 3 points, because you can never include the 2 boundary points and not include the internal points; you cannot shatter a set of 3 point, a line using intervals, you need something more than that.

So, but range spaces are always to find the V C dimensions define with respect to the the kind of ranges that that goes along with the with the range space. So, for intervals we cannot shatter even 3 points which means that according to the definition the an of course, we can shatter the pair of points on the line. So, therefore, the V C dimension of this is two.

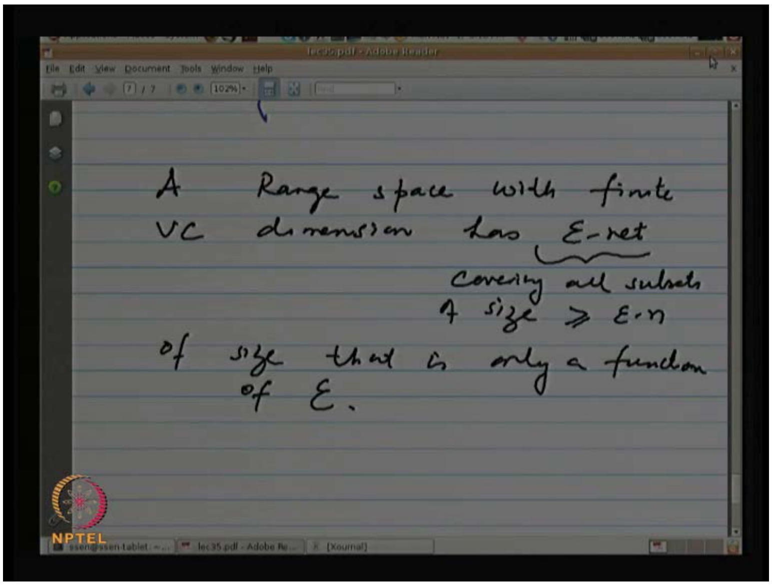
(Refer Slide Time: 03:49)



And we saw some other examples, where this you know range spaces were defined by this set of points in the plane and all the convex subsets. Now, this is a classical example, where the V C dimension is not even bounded. Because, I can take a set of points on the either points are basically on the convex hull. So, all the points are in the convex hull, then I can capture any subset of the points using convex subset, because subset of a convex polygon is also convex right; any any subset of convex polygon and the points define upon convex polygon is also convex. So, this is an example, where you do not even have finite V C dimension.

But, you know our our interest is in those range spaces where V C dimension is finite. So, that we can you know define certain other common property with respect to V C dimension ok. So, this is where we we conclude our last lecture.

(Refer Slide Time: 04:52)

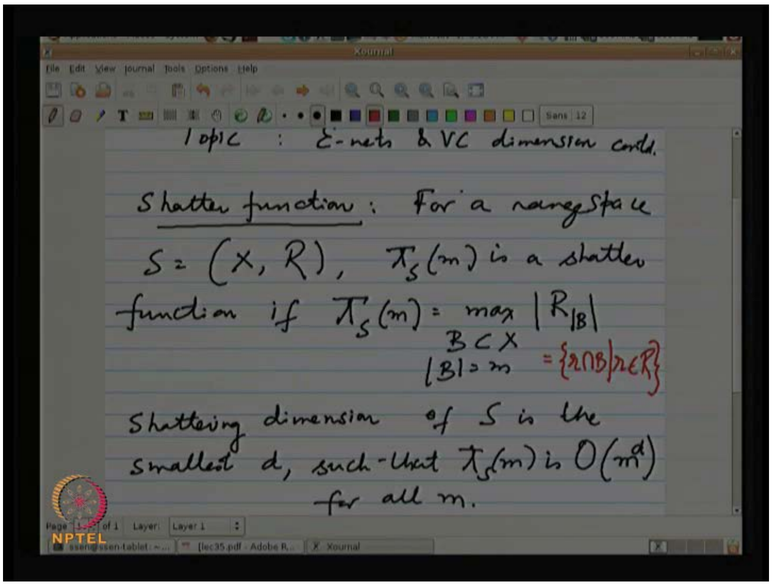


And the overall motivation was this that, you know if we somehow will be able to prove the falling property that, if you are trying to form a set cover of so called large subsets. And this set system or the range space corresponded to the set system has a finite V C dimension; this is somehow the size of the set cover can be captured only as a function of what is called the epsilon that is we are trying to capture cover all set, which are of size at least epsilon times.

So, normally if you do just you know some some simple random sampling, then will get an extra log of n factor as as the as the the size of the set cover that will require to hit all subsets of size at least epsilon. But, as we will see you know possibly tomorrow may be the next lecture that, when we have finite V C dimension we will be able to actually show the existence of set covers whose size do not depend on the number of subsets, it is only dependent on the epsilon.

So, we will need a few more definitions and some more concepts of a range spaces. Let us continue with that. So, the first thing that again I will define today is a slight variation on on the definition of what what we what we discussed in the context of shattering instead of points ok.

(Refer Slide Time: 06:44)



So, here is another definition of a twist on the shatter and that is called the shatter function. So, so for a range space S equal to (X, R). (No audio from 07:15 to 07:56) So, B is the finite subset of size m. So, look at all finite subsets of size m and we are maximizing on the number of ranges restricted to subset of size B.

So, you for any m, look at the number of ranges look at the number of ranges of the worst possible subset of size m. So, maximize the number of ranges for any subset of size m ok. And this pi S of m for every m when you set it to the maximum of this quantity that it is called the shatter function. So, for different m it could take different values depending on the configuration of the point.

So, for may be for m equal to 2 equal 2 it will be 5; may be m equal to 5 would be like 20 or whatever depends on you know what is the worst configuration of size m. That leads to the maximum number of ranges right maximum number of possible subsets as defined by the range space alright. So, this is the shattering function. And the shattering dimension is so now a certain difference between the shattering dimension and the V C dimension.

So, what is the shattering dimension? Is the smallest d, such that pi S, so this shattered function can be bounded by some kind of polynomial (No audio from 09:58 to 10:18).

(( ))

R oh restricted to B, this is this is what we defined before in the last lectures. So, you look at all the ranges; so, this is nothing but, this is nothing but all those r intersection B such that r belongs to the original range space. So, we are now only talking about subsets of of B that is all. So, we defined the shatter function like this.

And now, we are saying that the shattering dimension of this of this range space is such that, you know we want to somehow bound it; although it was not clear from the definition of you know V C dimension that is so what we looking at, we are looking at only the power set right. So, we say that you know beyond this some size of some V C dimension some some size delta, we cannot really generate the power set using the ranges right.

So, but the power set is basically exponential right 2 to the power delta. Now, just because it is not exponential, also does not imply that thing should become polynomial right. Suppose, my V C dimension is 10 which means that you know there is some subset of a size 10 such that, I have ranges from the range space that can generate all subset of this configuration of 10 points that is 2 to the power 10.

Now, when I go to 11, 12, 13 whatever I know that I cannot shatter those, because my V C V C dimension is 10 I cannot shatter a a subset of 11 points or 12 points or whatever; which means that I cannot generate all the 2 to the power 11 subset or I cannot generate all the 2 to the power 12 or whatever ok. But, it does not say anything about it that it should be polynomial alright its only that we cannot generate the power set right.

But then what turns out and that is that is very surprising thing that once we you know once we cannot shatter something, it turns out that after that we will give the proof of that; that it turns out that then we can kind of bound the number of subsets that we can generate is the ranges by polynomial. That is eventually what we will prove. And in that process, I give another intermediate definition of shattering dimension ok.

So, shattering dimension says that, you look at the worst case configuration. So, that you you are looking at the maximum number of subset that can be generated for that worst case configuration of size m, that defines my shattering function ok. Now, that shattering function, if I can bound by some polynomial m to the power d, where d is some fixed constant could be very large ok. Then that is called the shattering dimension of this. The shattering dimension of phase is not necessarily the V C dimension.

So, shattering dimension is by definition something, so that it will help us or it is suppose to help us you know bound the number of subset that you can generate by some kind of polynomial. It is not even clear from this definition that the shattering dimension actually this is that; there is such a d, it says is find the smallest d such that this is true, there may not be such a d, it is just a definition at this point.

(( ))

Yeah yeah it can be see this V C dimension can be infinite, those shattering dimension can also be infinite clearly you know the all the, where the V C dimension is infinite, then shattering dimension is also going to be infinite.

(( ))

Yeah. So, I will do that in fact I have done that already, if you if you just recollect you know what we did for you know trapezoidal maps; what did we do? We basically defining some kind of bunches of segments that we can generate get by from a samples. So, we looked at the segments, selected some of them, they are those defines trapezoid right.

So, if I ranges such trapezoids for example; so, if you look at the range space, where my x is the set of segments on the plane and my subset are trapezoids trapezoids of where you know the the parallel sides parallel to the y axis. So, that is a that is a range space, where you know we were actually generating this subsets, the possible subsets of the segments right.

Each trapezoid captured some segment and we already you know if it is a very simple simple argument. Because, if trapezoid has to be defined by at least four segments, then there maximum of n to the power 4 possible trapezoid right. And those are the only distinct subset that you can have. So, you see that, already we have this m to the power d is 4 there. So, we have an argument for the the the range space, where S is the set of segment and and the subset of trapezoid, where the shattering dimension is 4.

Now, it is also a non trivial problem many cases for many range spaces, what is the V C dimension. So, we will also look at you know may be shattering dimension is sometime easier to compute; and if you can find relationship between shattering dimension, V C dimension, then maybe we can get a bound in the V C dimension. That is the other motivation for the definition.

(( ))

No, this is not a computational problem it is a purely commentarial quantity. So, we are not trying to discuss algorithm to compute this, these are just commentarial quantities.

But, you have to determine?

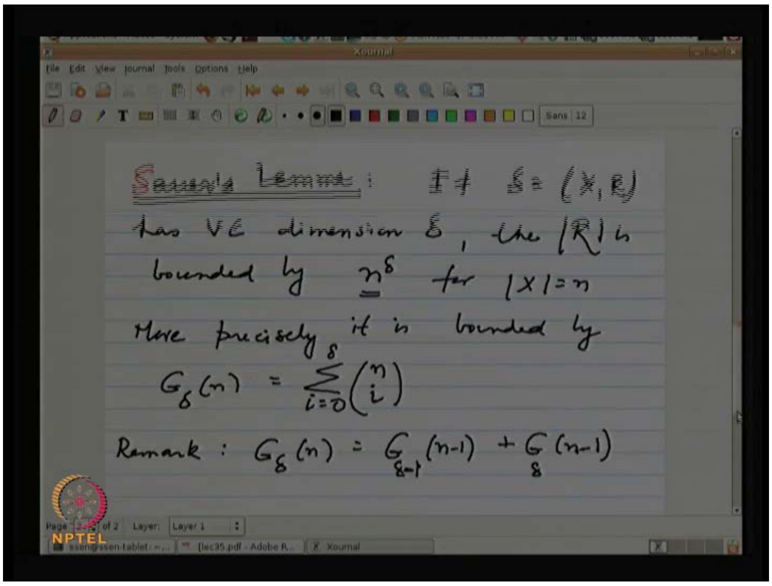
Yeah determines we have to determine. It is not an algorithmic we are not trying to determines the V C dimension algorithmically. It is you know given a range space tell me, what is a V C dimension, whatever by what other proof techniques you may have, or if you do not have or maybe you do not have a very tight bound. You can now we say V C dimension is less than 100, but I do not know whether it is 90 or 91 or whatever right.

Same thing about shattering dimension I am saying it is 4, but maybe it is less who knows. But, our you know requirements are you know we are not really looking for very tight bounds, we are only kind of looking for looking at situation, where we can get a finite bound in the V C dimension. Because, most of our you know arguments will hinge upon a finite of the V C dimension nothing more than that right.

So, the shattering dimension somehow manages to bound this function by some polynomial. And V C dimension definition does not say that, if you cannot shatter a subset we can actually bound the number of subset by polynomial, it does not say that; it only says that, if we cannot we cannot bound it by the sorry if we if we cannot generate the power set, then you know we have found the V C dimension that is all ok.

So, let us see if there is connection between we will see some connection between shattering dimension, V C dimension; and why you know that is that is surprising part here it is so, you know that if it is not the power set, than it becomes some kind of a polynomial. That is the bound, we will try to prove. So, for that there is a nice technical lemma.

(Refer Slide Time: 18:05)



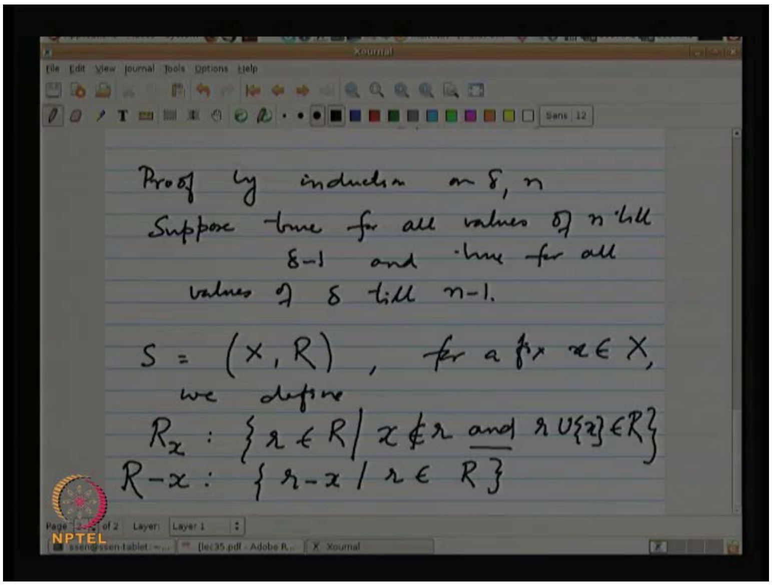
So, so it is called Sauer's Lemma. I think various people proved it, but it goes with the Sauer's Lemma. So, if S equal to (X, R) has V C dimension. So, I am trying to use consistently delta for V C dimension and d for shattering dimension. So, now comes to the points line (No audio from 18:43 to 18:57). Suddenly we are saying that, now we are claiming that, if the V C dimension is delta then and we have a set of points, finite set of points X of size n, then the maximum number of ranges is n to the power delta ok.

So, from suddenly from 2 to the power something, which was the power set, now we are talking about n to the power delta. And in fact more precisely, so this one more precisely it is bounded by some function G delta of n, it is a familiar looking function actually, n choose I; so, all possible ways of choosing delta or fewer objects from from n objects.

And just as a remark and we use that property for the proof, G delta n is G delta minus 1 (n minus 1) plus G delta minus 1 (n) right. It is a familiar recurrence for choosing n objects from choosing sorry delta objects from n objects. So, either we choose the oh sorry; so, we have not chosen sorry miss something (Refer Slide Time: 20:59). So, either we choose the n th object or we do not choose n th object right.

So, if we chosen the n th object, so then I have to choose delta minus 1 that is fine. And in other case, I just in both case should be n minus 1 right (Refer Slide Time: 21:22); in one case, I have not chosen, and one case I have chosen in case I chosen right. This I just you know plug somewhere in the proof actually, when do the proof. So, this is called Sauer's Lemma.

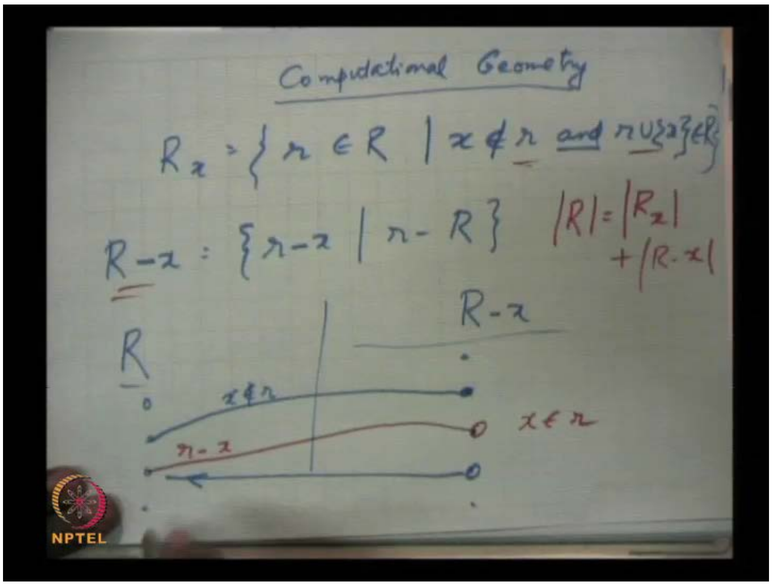
(Refer Slide Time: 21:45)



How do we prove it? We will prove by induction, I would not write the induction statement; proof by induction on delta, n. So, there is a double induction basically the 2 variables, delta and n right. So, suppose true (No audio from 22:05 to 22:32) so actually since actually I am not writing the induction statement it is a slight danger.

So, it is true for all values till delta minus 1 all values of n and true for all values of delta till n minus 1. What we do? We define you know 2, so from the given range space S equal to (X, R), we define so for a fixed x we for a fixed x R x r such that x does not belong to r and r union x belongs to R. So, I am defining 2 sets 2 ranges from the original range R. And this is I am kind of throwing out x from the ranges; so, such that, so these are two things that I have defined, it is not a good place to define this.

(Refer Slide Time: 25:44)



So, let me just make a note here; so, we will have to stare at this thing for a while. So, R x (No audio from 25:44 to 26:16), so there is an original range space R and look at this R minus x. So, there are these ranges here and some ranges here. Now, can we do define some kind of mapping between R and R minus x. See, x may or may not become belong to the range right. So, if x does not belong to the range, so this one has a easy counterpart here right.

So, if x does not belong to r. It does not easy it does not belong to r; so, R minus x and R are the same. The trouble is that, if x belongs to r; so, if x belongs to r this one x belongs to r, then it is mapped to again something which which takes out this. So, may be its mapped somewhere so may be its mapped somewhere, x belongs to r and I am mapping it to something where I am taking out x from there right.

The trouble happens, when x belongs to r, and so both of these when they get mapped to the same one, then we may have trouble. Let us we are trying to have a kind of a, we would like to prove this things, let me tell you what we are trying to do? We are trying to prove this. The number of ranges, R is equal to this plus this (Refer Slide Time: 28:26).

So, I am now trying to look at the cardinality of R minus x and R. So, when x does not belong to r, then it has a counterpart; now, x belongs to r is mapped somewhere, but something else also be mapped here right. may be why How can this two things get mapped here? So, both of them may get mapped here, when does it happen?

(( ))

Yeah yeah. So, you have a range such that, x in some sense x there is a range in R such that, this and this both of them are mapped to the same (Refer Slide Time: 29:44). So, the r union x also belongs to R, and taking out x that also belong to R, both of the ranges will get mapped to the same range in R right. If both of them are getting mapped, it means that this one this range even without x and this has the x, so both of them are getting mapped here.

So, both both those ranges; so, x it is precisely this, when when when x does not belong to r, so this r and this r both of them belong to the range, original range.

(( ))

Am I doing the other way round, just for a moment. So, I am just trying to find the correspondence between r and r minus x. So, r minus x what seems to be? So, it should be other way round.

(( ))

No no.

(( ))

No, R x we are not touching, we are only looking at this one. R x will basically you know will make sure that, this happens that is all. R x is precisely those elements where this is happening.

(( ))

Is the argument is easier to go in other direction? So, r minus x, so will it be easier. Let me try another thing just a moment.

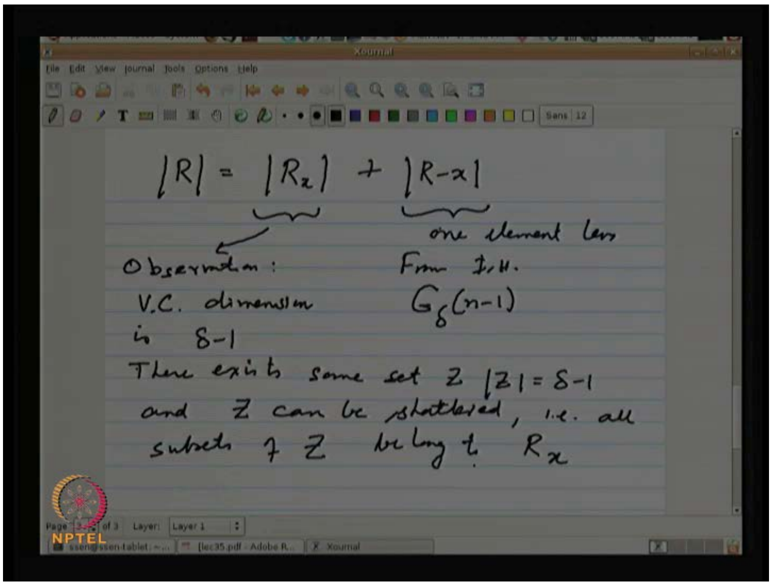
(( ))

So, it seem little easier to argue this way R and R minus x. So, yes I mean for if x does not belong to R, this is an easy mapping. And so there is some range let us say, this is r prime and others is r double prime. So, this got mapped here. And if this is equal to r prime that is what are you saying right x that will also get mapped here.

So, so this may be this is not be the right figure. And if when this happens, yeah surely this number of the ranges here should be less or equal than number of ranges here right, this is a this is a in 2 mapping. So then, so for which sub ranges this is happening? It is happening precisely for the ranges such that, x does not belong to r belongs to R. So, this range belongs to R and r union x also belongs to R right.

So, therefore, then I can write this as, this plus this; precise that, the duplicate element that are making it happen right. Now, So, I am going to add exactly those, another set of these this this will be added basically. So, then you know we will have same, so this this equality will be satisfied; so, with this (No audio from 33:52 to 34:11).

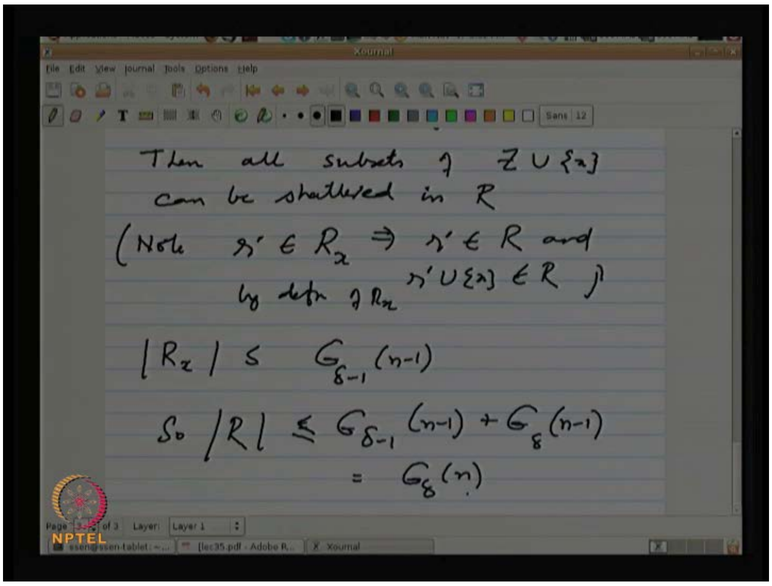
(Refer Slide Time: 34:19)



So, R is equal to R sub x plus R minus x right. So, our claim from induction hypothesis is what. So, this this is the set of the ranges that let us first look at this. So, this is the set of all ranges without the x part right. So, this is one element less, so one element less. So, from induction hypothesis this is G delta n minus 1, there is one element less. So, the claim is that, the number of ranges in R minus x is bounded by this quantity right; how about this?

So here, what we can claim is V C dimension is delta minus 1, why the delta minus 1? So, V C delta minus 1 means there exists some set let us say set z, which means that z equal to delta minus 1 and z can be shattered right, i e all subsets of z belong to R sub x.

(Refer Slide Time: 36:43)



Now, the if that is so, then I claim that, then all subsets of z union x can also be shattered, why? Because, note r prime belongs to the R x implies r prime belongs to R and r prime union x belongs to R by definition. So, I can also shatter z union x, because I can generate all the subsets which contain x and does not contain x. So, that is the definition of power sets of z union x.

I do have the all the subsets of z anyway present in r and now I have all the subsets of z union x also present in r; because that is the way that is the way this was defined right R x was defined. And therefore, this quantity can be bounded by the first one. So, R x can be bounded by it does not have the element x and it is V C dimension delta minus 1, it cannot be anything more than anything delta minus 1 n minus 1.

So, the first term is G delta minus 1 n minus 1, the second one is G delta n minus 1. So, when you add them, total number of ranges additional that, that is according to our previous observation. So, number of ranges is less than G delta minus 1 n minus 1 plus G delta n minus 1.

Sir, the definition of V C dimension says that, it cannot be shattered for greater than delta minus 1.

Yeah

(( )) we are proving that it can be shattered for delta.

No no. So, we prove proving by contradiction right. So, suppose the V C dimension of this was delta, it was more than delta minus 1 suppose delta; it means that for original range space it is delta plus 1, because I can generate all the power set of of this z union x right. So, by contradiction the V C dimension cannot be more than delta minus 1.

So, this is a G delta of n, so we got this. So, it seems like an, it is a very clever proof actually. So, one need to go over to take couple of time. So, although I have written this statements, this is true statements, may be you have to just you know just mentally sort of see why I can write this in particular this this; you can do a case analysis you know x belongs to r prime, x does not belong to r prime right. So, this will come true.

So, eventually you know what what we get out of this is that, if the V c dimension is bounded by delta, then the total number of range is for a subset of finite subset of n elements n points can be bounded by n to the power delta. So, it suddenly becomes a polynomial. (No audio from 41:02 to 41:31) Any questions about this, let us go ahead.

So, now let me establish. So, we started again let me just go back, we we started with a definition called this shattering dimension right (Refer Slide Time: 41:50). Shattering dimension is what? That, if I take any point set of size m ok; then we can bound the total number of ranges by m to the power d, for any m.

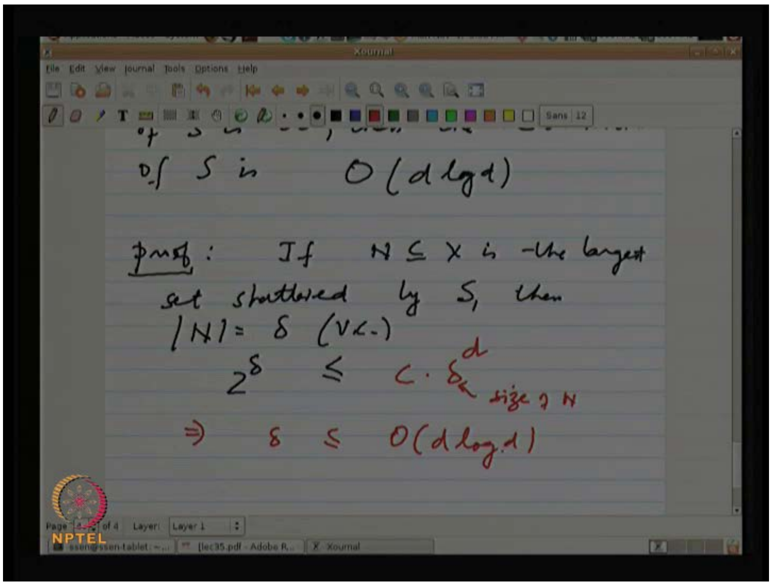
(( ))

Yeah yeah something like that yeah.

Is it that, so this is the boundary condition or is there another.

No, no no this is basically what it is that is all.

(Refer Slide Time: 42:40)



So, there is a more precise statement that, if S equal to (X, R) has shattering dimension ok. So, one one way should be clear right, if if something has V C dimension delta then then we are able to bound everything by n to the power delta. And what does shattering dimension say? That you know any point set of size m, we should be able to bound by some m to the power d ok.

So, if the V C dimension is delta, my so maybe I should write it actually. If S has a V C dimension delta, then shattering dimension is bounded by delta. We we do not want to argue the other way around. So, claim if shattering dimension of S is d, then the V C dimension of S is, is it d? This is a logarithmic factor. What is the quick proof of that? So, if let us say (No audio from 44:36 to 45:07), this is the V C dimension right; we are saying that, larger set that is shattered is is of size delta. So, 2 to the power delta is the power set of this subset n right; and that should be again somehow bounded by the shattering dimension right.

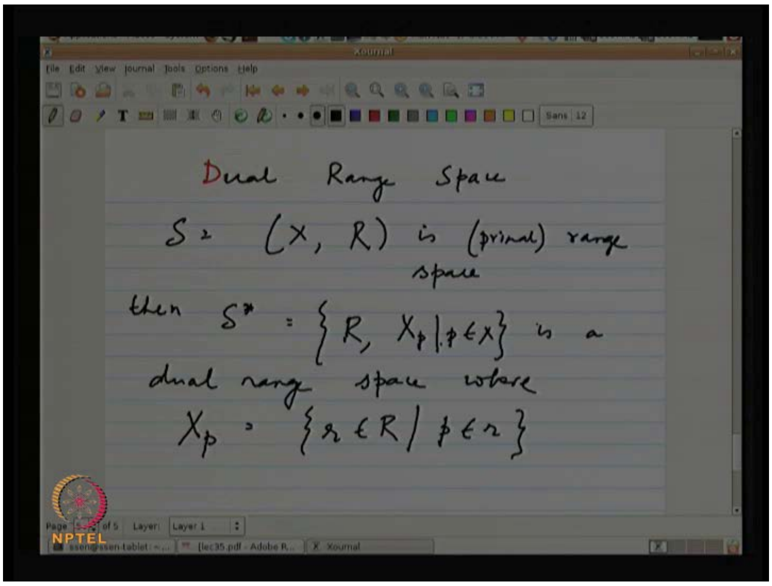
So, this is less than or equal to. So, shattering dimension says what? Shattering dimension says that, some polynomial right let us go back is that (Refer Slide Time: 45:39), it is some m to the power d. So, let us say some constant with more precise let us say it some constant times m to the power d. So, so this quantity 2 to the power delta should be less than some constant to the power this is my, this is the size of the set right, size of N; so, that to the power d is the definition of shattering dimension.

So, this implies essentially, delta is less than or equal to you know some S is d log d, you just transpose this and you get that. Again, what is the motivation for all this? You know in some for some range spaces, it may be easier to argue about the shattering dimension j like just about the trapezoidal map ok.

It may not be immediately clear that, that the V C dimension should be whatever 4 or 5 you know it is much harder to argue with the V C dimension. What is the largest set of segments that trapezoid can actually generate, all the subsets of that set of segments right. But, we argue that is 4 that are certainly bounded by 4. Let us say that is the shattering dimension is bounded by 4.

And what this says is that, V C dimension then bounded by 4 times log of 4; again, so if 1 is finite you know some log of that multiple will also be finite. It is not going to be a very large number it is pretty close to that. So, we are getting their just a few more definitions before now we go back to algorithm actually, I am ruling these definitions. So, I think we will get to the algorithm only tomorrow.

(Refer Slide Time: 47:43)



The other thing I am going to define is the notion of what is a dual range space nothing is complete without duality right, so dual range space. So, if S equal to (X, R) let us say the primal range space, then S star equal to R and X p such that, p belongs to x (No audio from 48:36 to 49:05). Let me illustrate that with an example (Refer Slide Time: 49:09 to 49:19).

(Refer Slide Time: 49:25)

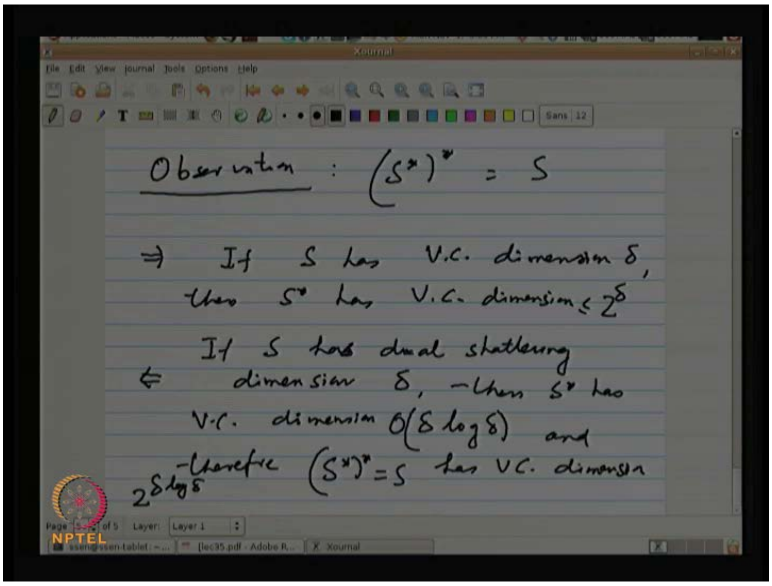
So, we have some points and these are my ranges suppose I am just drawing this ranges like this, (No audio from 49:37 to 49:51) let me call it A, B, C, D; and P 1, P 2, P 3, P 4, P 5, P 6. So, I can actually the original range we are talking about is basically the set of points, and these are the set of subset right. I can just write it as P 1, P 2, P 3, and P 6.

Again use the kind of matrix notation to denote this whole system. So, what is I have 4 ranges and you know so P 1 lies it just a incidence matrix. So, A contains P 2 and P 4 P 4 and P 3 and rest is all 0's. So, if I just write the incidence matrix that is one way of representing the same range space right. So, B contains P 3, P 4 and P 5; P 3, P 4, here I mean P 6, P 5 and then 0, 0, 0. So, I can complete this matrix. So, that is basically what I am saying, this is my primal range space.

And the dual range space is nothing but, the the transpose of this, why? So, we are saying in the dual range space my, so I interchange, so my ranges is basically become my rows A, B, C, D and then X p are basically all those ranges that contain the point p right; so, P 1, P 2, P 3 all out of P 6. So, x P 1 essentially corresponds to all those sets that has 1 in this column all right. So, P 1 let us say what P 1 lies in? P 1 lies in D, P 1 lies only in D nothing else everything else is 0.

So, X p 1 what is X p 1 then? It is just D right. P 2 is contained in A and D right, so A and D other is 0, 0. So, X p 2 is A, D. So, these are my, so in the dual range space, so it is basically you know the this is A B so x x sorry S star is A, B, C, D and then some subset of this thing D, A, D and so on, and so forth right. So, that is the dual range space. Now, clearly because the way I have drawn the picture the dual of the dual is the primal, we just again transpose it back.

(Refer Slide Time: 53:26)



Observation S star star is S. What is this significance of this dual range space? So, we will I will let me just write some couple of observation and then I will just get to that. Now, some how that turns out to be some kind of relation between the primal range space and dual range space in terms of V C dimension. If I can bound the V C dimension to the primal range space, I can bound the V C dimension in the dual range space and vice versa.

So, in one direction it is this one, if S has V C dimension delta, then S star has the dual range space has V C dimension; any guesses, where it is pretty bad things actually exponential. But, again I am saying I said before and let me re literate again we are we are kind of only looking at finiteness. So, if this is finite, that is finite. That is what matter is.

And these bounds may not be tight, there are some proofs you know which is the constructive proofs? So, these bounds are not tight really; but, it is not known whether it has to be exponentially or not.

(( ))

Yeah yeah yeah.

If S has this is another statement, dual shattering dimension. So, if S has a dual shattering dimension, so if the dual space, where S star basically has shattering dimension delta; what does it mean, then S star has V C dimension. So, what is the relationship between shattering dimension, V C dimension? V c dimension is bounded by d times of p right. So, maybe I should write let it write delta log delta right.

So, S star has V C dimension this and according to the previous statement and therefore, S star star that equal to S has V C dimension 2 to the power delta log delta, which is about delta 2 power delta. Again it is only just establishing the you know the finiteness of this, so once you go again right. So, let me stop here today; and the motivation for defining the dual range space is a is that you know we will we are getting to this you know the set cover problem actually.

So, we will will define a very clear algorithm for constructing set cover in a range space that has bounded V C dimension; and when we do that, we will actually we constructing. So, I have another definition, I could not do today, so I will have to get to that more epsilon nets basically.

So, we will construct something called as epsilon net, I define as the epsilon net right. It has to hit every large subset. So, the epsilon net that we construct will be constructed in the dual range space and therefore, we need this connections. That is why, I was building up with this definition alright.