IIT Madras

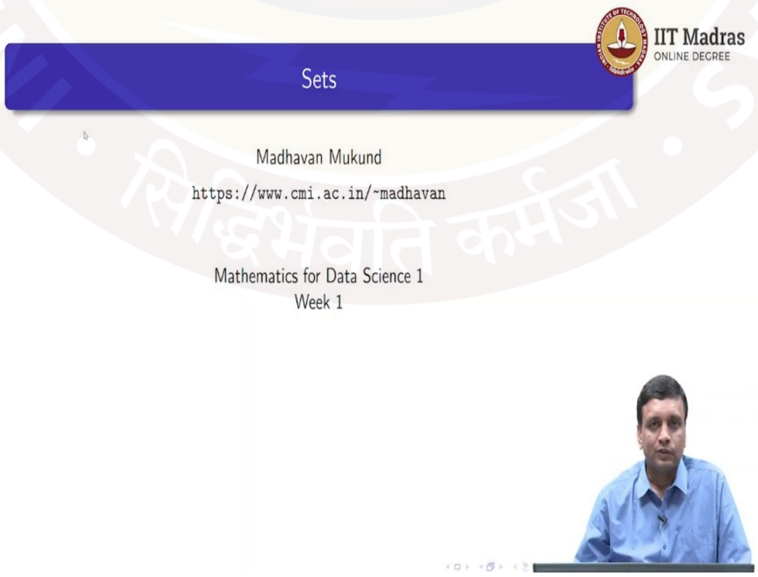
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

Week - 01
Lecture - 04
Set Theory

(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



So, we have seen numbers; we have seen natural numbers, we have seen integers, rationals, reals and we have loosely talked of them as sets of numbers. So, let us try to understand little more clearly what we mean by a set.

(Refer Slide Time: 00:26)

So, at its basic level a set is a collection of items. So, for instance, we could have a set called the days of the week which has 7 members; Sunday, Monday, Tuesday, Wednesday, Thursday, Friday and Saturday or we could take a number like 24 and list out the factors of 24 and call this a set. So, we have 1, 2, 3, 4, 6, 8, 12 and 24.

So, if you count, there are 8 factors that 24 has or we could take all the prime numbers up to a certain limit. Supposing, we want to know the prime numbers below 15, then we know that we do not have 1; but 2, 3, 5, 7 are the single digit prime numbers and then, 7, 11 and 13.

So, 2, 3, 5, 7, 11, 13 are all the primes below 15. So, this is how we talk about sets informally as just collections of items. Of course, as we have seen sets can be infinite and in particular, the infinite sets that we deal with very commonly are those which consists of the different types of numbers.

Remember that N this funny N stands for the natural numbers that is 0, 1, 2, 3, 4. Z stands for the integers. So, that is the natural numbers along with the negative integers like -1, -2, -3 and so on. Q is a peculiar symbol for the rational numbers, these are the fractions those numbers which we can write as $\frac{p}{q}$; where, p and q are both integers.

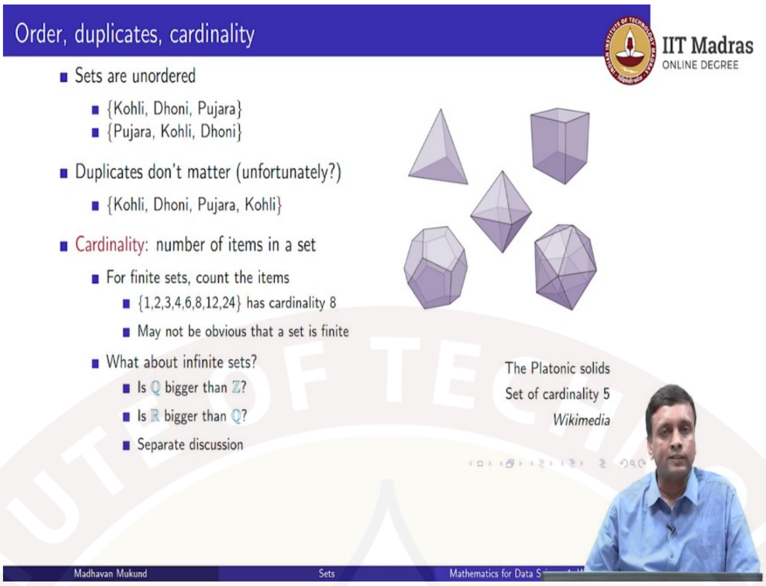
And finally, R is a set of real numbers. So, the real numbers includes all the rationals all the fractions, but also numbers that cannot be represented as fractions, such as the square root of 2 and other irrational numbers like π and e.

So, in all these things that we have seen above, it looks like there is some kind of condition which requires a set to have some uniformity; either a set consists of numbers or a set consists of days of the week or something like that. But actually mathematically there is no constraint on this. A set can have any kind of members, even a mixed membership; there is no uniformity of type.

So, for instance, we could enumerate the set of objects that appear in a painting. Now, here is a particularly famous painting, where it is not so easy to enumerate the objects because its drawn in a very abstract way. This is a painting called Three Musicians by Pablo Picasso. But we could see roughly that there are three people and that there are some musical instruments and so on and if you look very carefully, you will even find a dog.

So, notice that there is no commonality. There are people, there are musical instruments, there are chairs, there are tables, there are animals and so on. So, a set in particular can have any kind of members, it does not matter if they are mixed in type.

(Refer Slide Time: 03:00)



So, one of the important differences between say set and a sequence or a list is that the order in which we identify a set does not matter. So, normally when we talk of numbers, we tend to list them in a particular way; but as the set it does not matter. So, for instance, if you take the set of cricketers; Kohli, Dhoni and Pujara. If you reorder this set as Pujara, Kohli and Dhoni, it is the same set right.

So, the sequence in which you list the members of a set does not matter and for that matter, if you happened to accidentally write the same member twice, it does not change the set. So, in this particular set if we add Kohli a second time, as the set it does not matter. Though of course, if you are a cricket fan maybe you would like Kohli to bat twice for you.

So, when we look at a set, we might ask a basic question as to how many members it has. So, the cardinality of a set is the number of items in the set and if it is a finite set, we can just count the items. So, for instance, if you look at the factors that we listed of 24, then we can count them and say that this has cardinality 8.

Sometimes, it may not be obvious that a set is finite. You might remember from geometry that a regular polygon is one, where the all the sides are equal and all the angles are equal. So, the smallest regular polygon is an equilateral triangle in which we have 3 sides all equal and 3 internal angles of 60 degrees each. Then, we move to four sides we get a square, then we get regular pentagons, hexagons, heptagons, octagons and so on.

So, for any number of sides, you can draw a regular polygon with that many sides with equal angles on the inside. So, there is no limit. The set of regular polygons is infinite. But if we move to three dimensions, the corresponding notion to a regular polygon is what is called a platonic solid. In a platonic solid, first of all you have surfaces or sides each side is a regular polygon and all these regular polygons meet at the same angle in three dimensions.

Now, it turns out that though you might imagine that there are infinitely many regular polygons in two dimensions, there are only 5 platonic solids in three dimensions. So, this is an example of a set which turns out to be finite, even though there is no reason for it to be finite. So, these 5 platonic solids are the tetrahedron which has triangles.

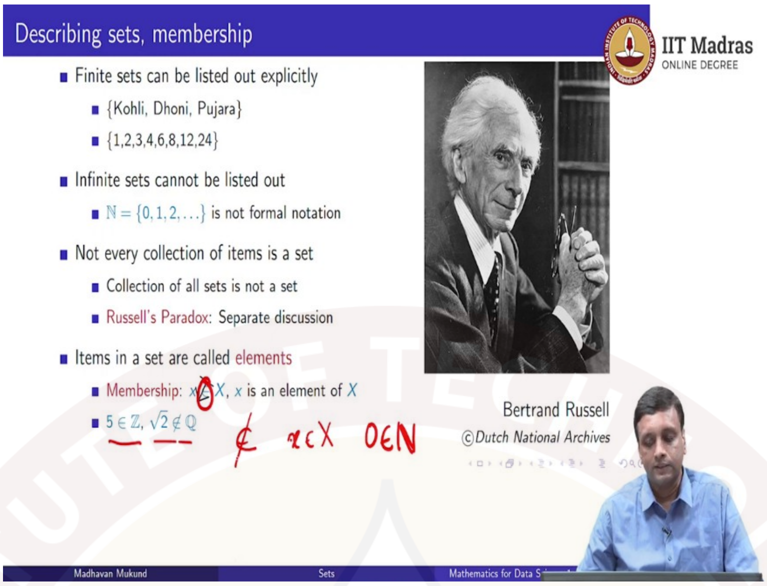
The cube which we have which has squares and then, we have an octahedron which has 8 sides which are triangles. Then, we have a dodecahedron with 12 sides and an icosahedron with 20 sides and there are no other regular solids, surprisingly it turns out.

Now, cardinality is quite easy to determine for a finite set, but what about for an infinite set? Remember that, we said that we wanted to go from integers to rational numbers because we want to talk about what happens when we divide 2 integers and the answer is not an integer and it is clear to us from our discussion that integers were discrete, we can talk about a next number and a previous number. So, there are gaps in the integers and rational were dense; between any 2 rational numbers, there is another rational number.

So, intuitively, it seems like we are adding things to the integers to get rational numbers. But can we make it formal in terms of cardinality? Are there more rational numbers than there are integers? And what happens, when we go from Q to R when we go from rational numbers to real numbers? So, remember that the real numbers, we had introduced because they were numbers such as the $\sqrt{2}$ which could not be represented as a fraction.

So, clearly there are some rational numbers which are real and some real numbers which are not rational and therefore, we have a bigger set; but again, R is really bigger than Q . So, this is a separate discussion, there will be a small separate lecture about this. But there is a way to measure cardinality of infinite sets, but it is not as straight forward as it is for finite set as you would imagine.

(Refer Slide Time: 06:58)



So, how do we describe a set? Well, we have already seen that for a finite set, we can just list out the members of the set explicitly. So, we can write out 3 numbers; Kohli, Dhoni, Pujara or 8 members the factors of 24. So, the normal notation for a list of items which form a set is to use these curly braces and to separate the items by commas.

Now, in many books and even in our lectures we will see notation like 0, 1, 2 …indicating that there is an infinite set of elements to be added which follows some kind of a pattern. So, this looks a way of listing out an infinite set, but you must understand that this is only an informal notation, this is not a formal notion.

So, you cannot write … and claim that you are listing out a infinite set. So, in fact, you need some other way of doing it and we will come to that as we go along in this lecture.

Now, it said seems reasonable that if a set is a collection of items, then we can collect anything and make it a set. It turns out that this is not quite true and this is particularly, a problem when we move to infinite sets. So, we have seen some infinite sets of numbers like naturals and reals and so on; but in general, if you take an infinite collection of objects, it may or may not form a set. In particular, Bertrand Russell showed that there is a problem, if we collect all the sets together and call it a set.

So, if we have a set of all sets, then we have a problem and this is something which is called Russell's Paradox which we will discuss in the separate lecture, but you must be careful to note that though the notion of a set is intuitive and it seems natural that any collection of objects is a set, we have to actually be a little careful in mathematics, if we are using sets in order to define what is a set and what is not a set.

But given that whatever we will see in our course, we will be fairly straight forward. So, whenever we see a collection of numbers or a collection of objects of mathematical description, we can safely assume that they are sets.

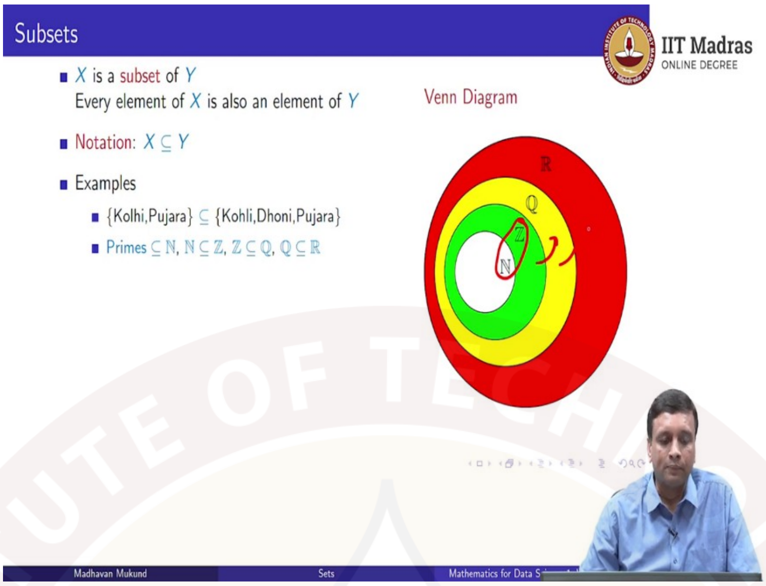
So, again some terminology. So, we have talked of different things items in a set, members of a set and so on. So, the most formal notation for the members of a set is an element. So, a set consists of elements and we write this membership of an element in a set using this $\in$ notion. So, we have this $\in$ notation which stands for element of. So, when we write $x \in X$, we mean that small x is a member of the set capital X.

So, example 0 is a member of the natural numbers right. So, $0 \in N$ is what we use. So, we can see for instance that 5 is an integer, but $\sqrt{2}$ as we claimed is not a rational number. So, an element of symbol with the line across it, means not an element of. So, 5 is an element of integer set and $\sqrt{2}$ is not a member of the set of rationals.

(Refer Slide Time: 10:02)

So, moving on from elements, we can compare sets by asking whether one set is included in another set and this is called a subset. So, $X \subseteq Y$, if every element of X is also an element of Y and this is written using this subset notation $\subseteq$. So, you have this familiar notation $X \subseteq Y$.

(Refer Slide Time: 10:27)



So, for example, if we take just 2 out of the 3 players were listed before saying Kohli and Pujara; then, this set forms the subset of our original set Kohli, Dhoni and Pujara. Similarly, if we take all the natural numbers and collect only the prime numbers. So, remember that the prime number is a number whose only factors are 1 and the number itself. So, it has exactly 2 factors; 1 and p and then, p is a prime number.

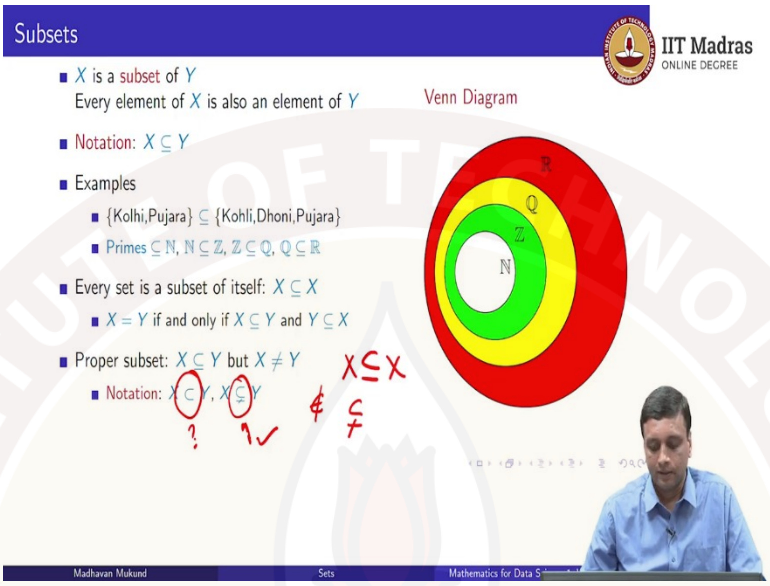
So, since some many numbers are not prime, primes is a subset of natural numbers. Since, the integers extend the natural numbers with the negative numbers, we can say that the natural numbers are included in the integers. So, $N \subseteq Z$. Similarly, we extended Z to Q. So, the set of integers is a subset of the rationals and the set of rationals is a subset of reals.

So, if you wanted to draw it, we could draw it in this particular way. So, we can raw a large circle representing the reals, a small circle inside right in the center representing the natural numbers and if one circle is included in another circle, it means that this circle is a subset of the circle outside it. So, here you can see that the natural numbers are a subset of the integers and then, from the integers, we can say that there are subset to the rationals and the rationals are a subset of the real numbers.

So, this kind of a diagram, where we represent a set by a boundary. So, this is a very abstract diagram. We are not in this case for example, listing out the elements of the set we are just indicating the extent of the set saying that the set are extends beyond Q and everything that is in Q is sitting inside R.

So, these are what are called Venn diagrams. So, a Venn diagram is a very useful way to picturize a set and relationships between sets; is one set a subset of another, is one set not a subset of another and so on.

(Refer Slide Time: 12:20)



So, we often use Venn diagrams pictorially in order to represent sets. So, notice that every set is a subset of itself because remember the definition of a subset set that X $\subseteq$Y, if every member of X is also a member of Y. So, since every element of X is also an element of X, trivially as a extreme case of this definition, every set is a subset of itself.

So, this in fact, gives us an important notion which looks obvious; but it is not so obvious, when are two sets equal. So, two sets are equal if and only if, they are actually the same set of elements. So, one way to check that two sets are equal is to check that everything in the first set belongs to the second set. So, X $\subseteq$Y and everything in the second set belongs to the first set. So, Y $\subseteq$X.

So, often this happens when we have two different ways of looking at the same set of objects. We have two different descriptions of the same set of objects and we want to check whether they are equal or not. Then, using the first description, we argue that everything which satisfies the first description also satisfies the second description and vice versa.

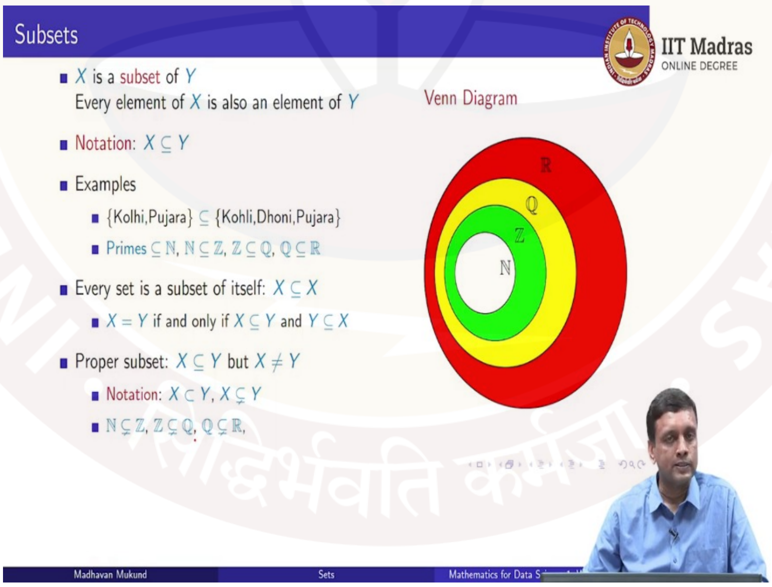
So, though this looks fairly obvious for finite sets, when it comes to infinite sets we have sometimes have to argue in an indirect way. So, this although it is an obvious statement is very important that $X = Y$ provided $X \subseteq Y$ and $Y \subseteq X$. So, sometimes we want to distinguish between the case, when X is really a proper subset of Y; that means, it does not include all of Y and that it is possibly equal to Y.

So, the subset equal to notation that we have right allows both. When we write $X \subseteq X$, what we are saying is that it is a subset, but it is actually equal. So, we are allowing both cases. So, if you want to talk about proper subsets, sometimes we use a different notation.

So, we might either drop the equal to sign , just write the subset sign $\subset$ or we might explicitly like we said not element of right. So, we are saying that this is not equal to. So, we are dropping the equal to from below the subset.

Now, this is a bit dangerous. Second symbol this not equal to this is always correct. This is sometimes used both ways. So, you have to be bit careful when we look at books when you see the single subset without the equal to whether they mean subset and equal to or proper subset.

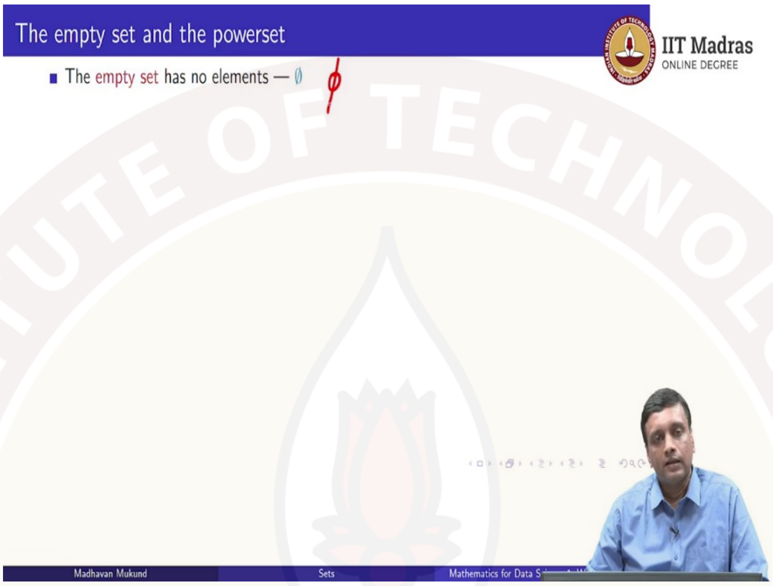
(Refer Slide Time: 14:45)



So, we know for instance that the natural numbers is a proper subset of the integers because the negative numbers are not there. Similarly, the integers are clearly a proper subset of the rationals and because the irrational numbers are not rational, the rational are a proper subset of the real numbers.

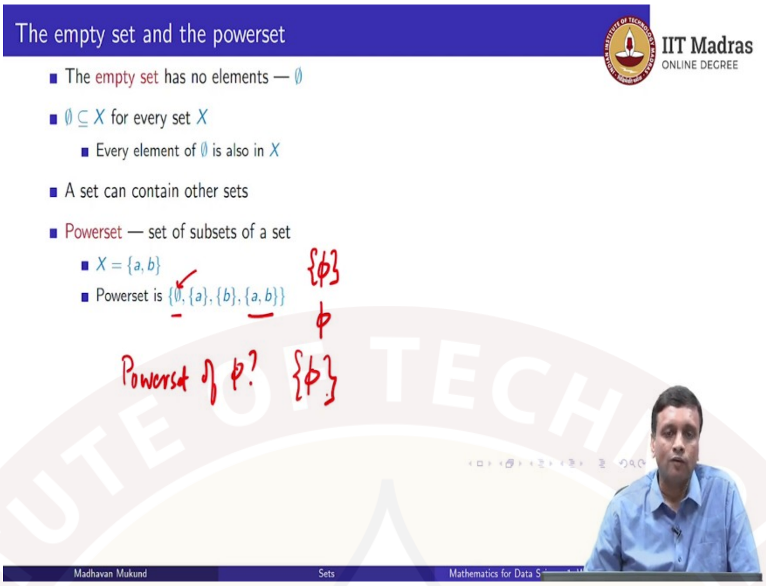
So, in most interesting cases, we will be looking at proper subsets. Sometimes, we will emphasize it by adding this cross against the equal to and sometimes, we will not and very often from context we will know whether we are talking about proper subsets or we are talking about subset which allow the full set.

(Refer Slide Time: 15:16)



Now, there is a very important set just like the 0 is very important in numbers, there is a very important set which is important set theory. It is the equivalent of 0. It is the set which has no elements. So, the set which has no elements is called the empty set and is written ∅. It is basically you can think of it as a 0 with a line across it. So, this Greek letter phi, symbolizes the empty set; so, it has no elements.

(Refer Slide Time: 15:44)



Now, what may not be very obvious is that this empty set is actually a subset of any set. Remember that we said that X ⊆ Y, if every element of X is also every is also an element of Y. Now, you might argue that an empty set has no elements. So, why is this true? Well, when we say for every and there is nothing in the set, then for every something is true right.

So, if I say that all birds with 3 legs have pink beaks, then this is actually true because we can imagine that there are no birds with three legs and therefore, every bird which actually has 3 legs will have a pink beak. But since, there are no birds with 3 legs this is actually true.

So, these kinds of vacuous statements as they are called will hold for sentences which use the word all where the set is empty. So, in particular, every element of the empty set because there are none. So, every element that could be in the empty set is also an any set X that you build. So, this empty set is a subset of every possible set. Now, though we have talked about elements and sets.

So, they are two different categories of objects. So, we have numbers and the numbers belong to a set of the type N or Q or R or Z; a set can clearly contain other sets. So, there is no restriction saying that the members of a set or the elements of a set must be some kind of discrete and indivisible objects.

So, one of the important sets of sets that we would like to look at is what is called the Powerset. So, we talked a subset. So, supposing we want to enumerate all the subsets. So, here is a two element set a comma b. So, what are all the subsets? Well, we already saw that the empty set is always a subset. So, that is one subset.

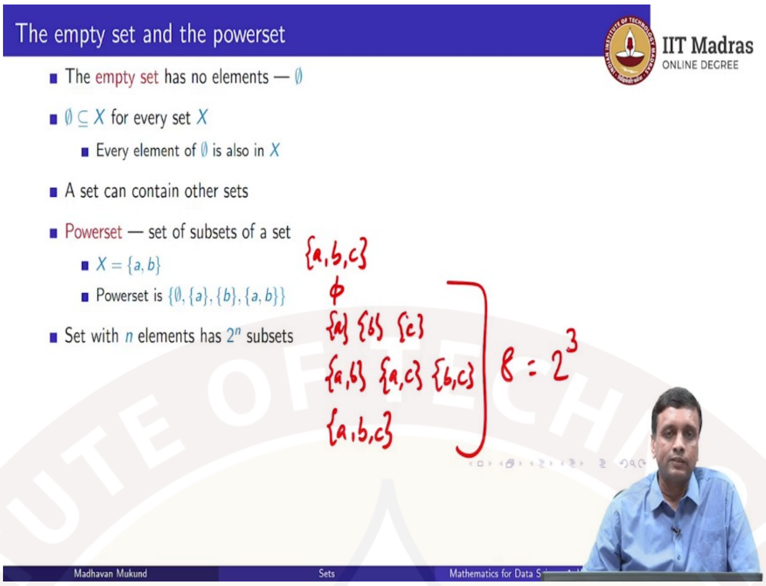
The set itself for any X, $X \subseteq X$. So, X equal to {a, b}. So, we have these two subsets which come just from the fact that empty is the subset of every set and the set itself is a subset. And then, we have two proper subsets either we can include the a and exclude the b or include the b and exclude the a. So, there are four subsets of X and if we group together these four subsets into a larger set, then we get the Powerset.

Now, notice that this itself is the set right. So, we do not write. So, this is different from this. The first is a set consisting of one element, namely the set consisting of the empty set. The lower thing is the empty set alone which is the set with no elements. So, if we put a brace around the empty set symbol, then we create a set with one element.

So, for instance, if you ask what is the power set of the empty set right. So, we know that the empty set has a power set which contains the empty set. So, we have at least one empty set as one element of the power set and there is nothing else right.

So, the full set is also the empty set, but if you duplicate an element, it is a same thing. So, in fact, the power set of the empty set is a set consisting of just one element, namely the empty set itself. So, just remember this, that the empty set on its own denotes a set with no elements, but an empty set with the brace around it is not the same thing. It is a set consisting of one element, namely the empty set.

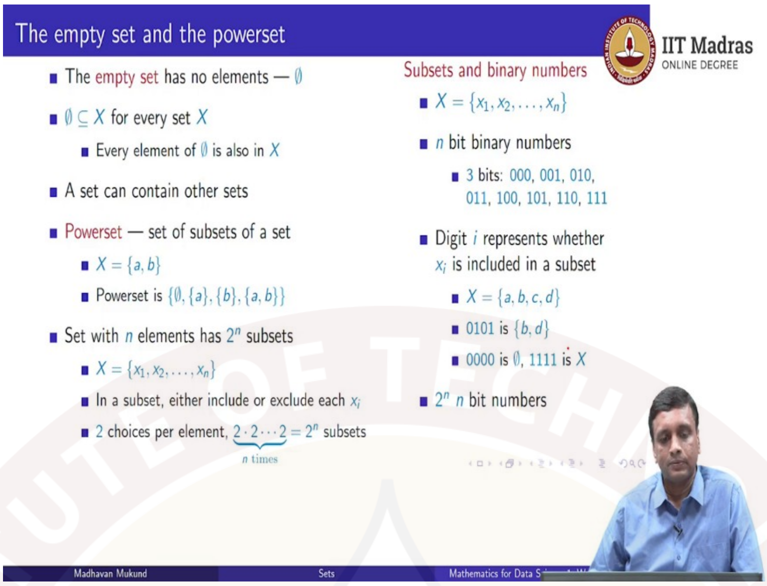
(Refer Slide Time: 19:03)



So, we saw above that if we have two elements, then the power set had four elements. So, in fact, one can generalize this and say that if we have n elements, then we would have 2^n subsets. So, for instance, if we had a, b, c right, then we would have 1 subset which is empty. We would have 3 subsets which are one element each and then, we would have 3 more subsets which are 2 elements each a, b a, c and b, c and finally, we would have the set itself right.

So, these are the only subsets. If you add these up, this is 8 which is 2^3. You can check that if you do it for a, b, c, d; then, you would have 2^4 , 16. So, why is it that a set with n element should have 2^nsubsets, no more no less?

(Refer Slide Time: 19:55)



So, here is one argument. Supposing we have n elements in the set. So, let us just call these without describing what they are specifically as x_1, $x_2$up tox_n. So, we have n distinct elements $x_1$to x_n. Remember these must be different because you cannot duplicate elements in the set. So, now, we want to construct a subset.

So, how do you construct a subset? Well for each elementx_i, we have to either include the set include $x_i$in the subset or exclude $x_i$subset. So, we have to make a choice for each $x_i$right.

So, overall, we have to make n choices right. For each $x_i$we have to decide whether to include it or exclude it from the subset. So, we have two different choices for each elements. So, we have two ways to decide whether to do something withx_1, keep it or leave it; $x_2$keep it or leave it. So, then we have two times two choices for $x_1$and $x_2$together; two times two choices forx_1, x_2, $x_3$together.

So, in general, if we have n such choices where each choice involves two options, then we have 2 into 2 into 2, n times 2^nchoices. So, each of these choices gives us different subset. So, whenever we make a different choice, we will either leave out i from the set or put anx_i. So, it will differ from the choice, where we do the other thing. So, each choice generates a separate subset. So, there are exactly 2^n subsets.

Here is another way of looking at subsets and getting to the same result. So, we can actually think of subsets in terms of binary numbers. So, let us again think of our n element set $x_1$to x_n right. So, now, supposing we look at n digit binary number. So, digit actually is comes from decimal. So, we say bit for binary digit. So, n bit binary number. So, remember in a binary number system, we have 0's and 1's and the place values represent powers of two.

So, we have the unit digits is units as usual. The next digit 2 to the power 0 is a is number of twos, number of fours, number of eights. So, it is like the decimal thing is in base 10. This is in base 2. So, now, if we look at n bit binary numbers, then for instance, if we look at 3 bit binary numbers, then we have 8 of them.

We can start with 0 0 0, then 0 0 1, 0 1 0 and so on up to 1 1 1 and again, the reason that there are 2 to the n, n bit numbers is because for each bit we can choose to put 0 or 1. So, we have two choices for the first bit, two choice for second bit and so on.

So, it is not surprising that an n bit binary number can represent 2 to the n different values from 0 to 2 to the n minus 1, if we think of them as numbers. Now, we are interested in n bit binary numbers as representing subsets. So, what we will look at is the ith bit and say that the ith bit represents the choice that we made.

If we chose to keep $x_i$in our subset, we will call it 0. If we chose to we will call it 1 for example. And if we choose to omit $x_i$from our set, we will call it 0. So, 0 represents the choice, where we leave outx_i; 1 represent the choice, where we keep x_i

So, supposing we have this four elements set a, b, c, d; then, if we look at the binary sequence or the bit sequence 0 1 0 1, the first 0 corresponds to a, so it says leave out a. The second 0 corresponds to c, so it says leave out c and for b and d we have put a 1. So, it says keep b and keep b. So, it says leave out a, keep b, leave out c, keep d. So, this 0 1 0 1 as a binary sequence corresponds to the set b comma d.

What does 0 0 0, the all 0 sequence say? The all 0 sequence says every $x_i$in the set is omitted from the subset. So, this is precisely the subset which is the empty set because it has no elements and what about the all 1 sequence? Well, the all 1 sequence says every x_i that we have is included in the final subset. So, this is the set itself. So, remember that these are the two extreme subsets; the empty set and the set itself and all the other ones come in between.

So, from this, we can see that every n bit number represents one sequence of choices. So, this gives us 2^nchoices because there are precisely2^n, n bit numbers. So, hopefully with this, you are now clear about the fact that any finite set with n elements has exactly 2^n subsets.