



IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute
Lecture- 5A
Sets: Examples

So, we have seen some definitions of Sets and some operations on them. So, let us look at more examples to get familiar with the notation and the terminology of sets.

(Refer Slide Time: 00:25)

Sets

- A **set** is a collection of items
- Finite sets can be listed out explicitly
 - {Kohli, Dhoni, Pujara}, {1,4,9,16,25}
- Infinite sets cannot be listed out
 - $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ not formal
- Membership $x \in X$, Subset $X \subseteq Y$
 - $5 \in \mathbb{Z}, \sqrt{2} \notin \mathbb{Q}$
 - Primes $\subseteq \mathbb{N}, \mathbb{N} \subseteq \mathbb{Z}, \mathbb{Z} \subseteq \mathbb{Q}, \mathbb{Q} \subseteq \mathbb{R}$
- Powerset — set of subsets of a set
 - $X = \{a, b\}$, powerset $\{\emptyset, \{a\}, \{b\}, \{a, b\}\}$
 - Set with n elements has 2^n subsets

Venn Diagram

IIT Madras
ONLINE DEGREE

Madhavan Mukund Sets: Examples Mathematics for Data Science 1

So, remember that a set is a collection of items and when we write out a set, if it is a finite set, then we can just enumerate the items in the set by writing them within curly braces. On the other hand, if we have an infinite set, we really can not write out all the elements even though informally, we put dot, dot, dot to indicate a sequence, if that sequence is not very regular. For example, supposing it is a set of prime numbers, which does not have a clear pattern, then it is not very easy to represent it explicitly like this.

So, we saw that there will be another notation called set comprehension that we will come to. But, before that let us talk about the two basic relationships between sets and membership of a set. So, membership is denoted by this element of relation. So, small x typically denotes a

member or an element of a set, and capital X usually denotes a set itself. So, when we write x belongs to X like this, what we mean is the element x belongs to X .

So, for example, the number 5 belongs to set of integers, and $\sqrt{2}$ does not belong to the set of rationals for instance. Subset on the other hand, says that one set is included in another set, so everything that belongs to X belongs to Y . So, for instance, all the prime numbers are natural numbers, so the primes are a subset of the naturals. Every natural number is an integer, so the natural numbers are a subset of the integers. Similarly, the integers are a subset of the rationals and the rationals are subset of the reals.

And we draw this using these Venn diagrams where we draw these ovals or circles or boxes representing the extent of a set, it is a picture of a set. And then depending on whether a box intersects another box or it sits inside a box, it indicates whether the first set is a subset of the other one or they overlap and so on. So, in this particular diagram which also has colors, we have indicated the subset relationship between the different types of numbers that we have studied, the naturals, the integers, the rationals, and the reals.

And finally, one very useful thing to know about sets is the power set. So, when we take a set, we can enumerate all its subsets. So, remember that we have just defined a subset. And in particular, we have this special subset called the empty set, which is a subset of every set. The empty set has no elements in it, but we needed it for technical reasons, and it is a subset of every set. And in addition, if you have 2 elements set $\{a,b\}$, then the subsets could be the individual elements, the set containing a and the set containing b or the entire set itself.

So once again, just like the empty set is a subset of everything the set itself is also a subset of itself. And we argued that for a finite set with n elements, we will always have 2^n subsets. So here for instance we have 2 elements, so we have $2^2 = 4$ subsets. So, this is just a review of what we have already seen.

(Refer Slide Time: 03:06)

Set Comprehension



■ Squares of the even integers

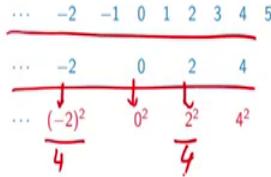
$$\{x^2 \mid x \in \mathbb{Z}, x \bmod 2 = 0\}$$

$$\{0, 4, 16, 36, 64, 100, 144, 196, 256, \dots\}$$

■ Generate Elements drawn from existing set

■ Filter Select elements that satisfy a constraint

■ Transform Modify selected elements



Madhavan Mukund

Sets: Examples

Mathematics for Data Science I, V

Now, let us look at this set comprehension notation, which is what we said we would use when we have to describe infinite sets which cannot be written down explicitly. So, this was a typical example. So, supposing we want to write down the set of all the squares of the even integers. So, the even integers are $-2, +2, 0$ as an even, $-4, +4$, and so on. But if we square them, then we know that $(-2)^2$ is the same as 2^2 is 4.

So, the set on the right which is written in this informal dot, dot, dot notation has $0^2, 2^2, 4^2, 6^2$ and so on. So, how would we write this out? Well, this is that notation on the left, which says that we take every x which belongs to the integers, check whether it is even, whether $(x \bmod 2 = 0)$, and then square it. So, let us just break this up into parts so that we remember exactly what is happening.

So first, in the set comprehension notation, we have a generator. A generator says that we are taking elements from an existing set, so we can only build new sets from old sets. So, we already have a set of integers, and we are going to try out every integer in the set, so, that is what x element of \mathbb{Z} says, is try every $x \in \mathbb{Z}$, so \mathbb{Z} generates this set. Now, all the x 's that come out are not interesting to us. So, we want to filter out those that are useful, that satisfy a given property.

In this case, the property that we are looking for is that the number is even. So, we want those x which come out of \mathbb{Z} through the generator, such that they satisfy the property that x when

divided by 2 has remainder 0, which is the property that x is even. And finally, with these x , we do not want to keep them as they are, we want to transform them. So, on the left-hand side of this vertical bar, this is the left-hand side are the actual elements of the set.

The elements of the set are generated right, then filtered through some conditions, which rule out the ones we do not want, and when the ones we keep, we can transform them. In this case, we want the squares, we do not want the even numbers, we want their squares. So, if you look on the right, this is what happened.

So, when we started the generating process, we had all the integers, then we filtered out, and we got only the even ones, and now we transform them. So, for each even number, we produced its square. And now in this process, you will notice that $(-2)^2 = 4$ and $2^2 = 4$ also. So, some elements will disappear because we do not keep duplicates.

(Refer Slide Time: 05:35)

Set Comprehension

IIT Madras
ONLINE DEGREE

- Squares of the even integers
 $\{x^2 \mid x \in \mathbb{Z}, x \bmod 2 = 0\}$
 - Generate Elements drawn from existing set
 - Filter Select elements that satisfy a constraint
 - Transform Modify selected elements

$\{0, 4, 16, 36, 64, 100, 144, 196, 256, \dots\}$

... -2 -1 0 1 2 3 4 5 ...
... -2 0 2 4 ...
~~... 4 0 4 16 ...~~



Madhavan Mukund Sets: Examples Mathematics for Data Science 1, W

Set Comprehension

IIT Madras
ONLINE DEGREE

- Squares of the even integers
 $\{x^2 \mid x \in \mathbb{Z}, x \bmod 2 = 0\}$
 - Generate Elements drawn from existing set
 - Filter Select elements that satisfy a constraint
 - Transform Modify selected elements
- More filters
 - Rationals in reduced form
 $\{p/q \mid p/q \in \mathbb{Q}, \gcd(p, q) = 1\}$

$\{0, 4, 16, 36, 64, 100, 144, 196, 256, \dots\}$

... -2 -1 0 1 2 3 4 5 ...
... -2 0 2 4 ...
... 4 0 4 16 ...

$4/10 \quad 2/5 + 2/2 : 4/10$



Madhavan Mukund Sets: Examples Mathematics for Data Science 1, W

Set Comprehension



- Squares of the even integers
 $\{x^2 \mid x \in \mathbb{Z}, x \bmod 2 = 0\}$ $\{0, 4, 16, 36, 64, 100, 144, 196, 256, \dots\}$
- Generate Elements drawn from existing set
 $\dots -2 -1 0 1 2 3 4 5 \dots$
- Filter Select elements that satisfy a constraint
 $\dots -2 0 2 4 \dots$
- Transform Modify selected elements
 $\dots 4 0 4 16 \dots$
- More filters
 - Rationals in reduced form
 $\{p/q \mid p/q \in \mathbb{Q}, \gcd(p, q) = 1\}$
 - Reals in interval $[-1, 2]$
 $0 r \in \mathbb{R}, -1 \leq r < 2\}$



Madhavan Mukund Sets Examples Mathematics for Data Science 1

So finally, when we go through this, we end up with this sequence 4, 0, 4. And then in this, we will throw away all the elements on the left, and we get the number sequence on the top. So, this is how set comprehension works.

So, we can write filters in many different ways as long as it is unambiguous, we will not be very particular about the language we use so long as there is no question about what we mean. So, for instance, we looked at this example, we have rational numbers, but some rational numbers are not in reduced form. For instance, if I write $\frac{4}{10}$, then I should actually think of this as $\frac{2}{5}$, because it is $\frac{2}{5} \times \frac{2}{2} = \frac{4}{10}$.

So, I have actually multiplied both the numerator and the denominator by 2, to go from $\frac{2}{5}$ to $\frac{4}{10}$, but it is the same rational number. So, we want the numerator and the denominator to not have any common divisors, which is the same as saying that their greatest common divisor is 1, that is nothing other than 1 divides both the top and the bottom of the fraction. So, if we take all the rational numbers, so we generate all the possible rational numbers $\frac{p}{q}$, which belong to the set of rationals.

Then, we filter out those which have no common divisor between the numerator and the denominator and we keep only those, we do not transform it in any way, we just keep it here. So,

here the transformation is just to keep it as it is, this is sometimes called the identity transformation. The identity just takes an input and produces the output the same as the input.

So, this gives the set of rationals in reduced form. So, here we have used a function, GCD. Even though we have not formally defined it here, we assume that people understand what GCD means. So, this is what we mean by saying that we can write the filter in any reasonable way, as long as people understand what it means.

Another example, we looked at are intervals. So, here we want the real numbers, which start from -1 including -1, and go up to but not including 2. So, in this case, we will use less than and less than equal to, so we will take all the reals. So, we take every possible real number, but we are not interested in all the reals, so we check whether it is greater than or equal to -1, so it includes -1 and everything above it. So, it cuts off everything which is strictly smaller than -1.

But, we also do not want it to cross 2, so we stop below 2, so it should be greater than equal to -1 or and less than 2 and if so, again we keep it without any transformation. And this notation on the top, the square bracket, and round bracket are indications of whether the endpoint is included or not. So, -1 endpoint is included, +2 endpoint is not included.

(Refer Slide Time: 08:14)

Set Comprehension ...

IIT Madras
ONLINE DEGREE

- Cubes of first 5 natural numbers

$$Y = \{n^3 \mid n \in \{0, 1, 2, 3, 4\}\}$$

- Cubes of first 500 natural numbers?

$$Y = \{n^3 \mid n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, \dots, 498, 499\}\}$$

- Use set comprehension to define first 500 natural numbers

$$X = \{n \mid n \in \mathbb{N}, n < 500\}$$



Set Comprehension ...



- Cubes of first 5 natural numbers

$$Y = \{n^3 \mid n \in \{0, 1, 2, 3, 4\}\}$$

- Cubes of first 500 natural numbers?

$$Y = \{n^3 \mid n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, \dots, 498, 499\}\}$$

- Use set comprehension to define first 500 natural numbers

$$X = \{n \mid n \in \mathbb{N}, n < 500\}$$

- Now, a more readable version

$$X = \{n \mid n \in \mathbb{N}, n < 500\}$$

$$Y = \{n^3 \mid n \in X\}$$



Madhavan Mukund

Sets, Examples

Mathematics for Data Science I, V

So, let us see why we would actually want set comprehension notation. So, let us extend our first example of squares of the even numbers to cubes. So, cube is just a number multiplied by itself 3 times. So, square is $x \times x$, a cube is $x \times x \times x$, 3 times. So, if we want the cubes of the first 5 natural numbers, we can write it out explicitly like this, we can take this generator and generate the first 5 natural numbers as 0, 1, 2, 3, 4. Remember that, in our terminology natural numbers start with 0, even though in some books, you will find that natural numbers start with 1, we always assume natural numbers start with 0.

So, the first 5 natural numbers are 0, 1, 2, 3, 4. So, this is our generator, take every n in this and transform it to n^3 without doing any further filtering. We are not asking for the first 5 odd numbers or the first 5 numbers which have some other property, we just take, taking the first 5 numbers. Now, imagine that we change this question to the first 500 natural numbers, then though we can write it out explicitly, it is rather tedious.

So, we have to replace the small list of 5 numbers by a long list of 500 numbers. And remember, we are not really allowed to write dot, dot, dot if we are being mathematically precise. So, we actually have to physically write out these 500 numbers. Now, this is not terribly convenient. On the other hand, we can define the first 500 numbers quite easily using set comprehension.

So, we can say, give me all the natural numbers, that is the generator, but restrict the natural number to be less than 500. So, remember that the first 500 natural numbers are going to be 0 up to 499. So now, this says that this set X is actually this long set here which we have written explicitly. So, we have replaced that very long and tedious expression by a much more compact expression, which captures exactly the same set. So now, we can have a much more readable version of these cubes of the first 500 natural numbers.

As an intermediate set, we generate the set X, set $X = \{n \mid n \in N, n < 500\}$. And then we take this as the generator and we say, okay, take every n which belongs to this X. So now, we know that x is restricted to 0 to 499. And then, take the cubes of these numbers, so we get n cubed in this range. So, this is one other use of set comprehension, which is to make our definitions more readable and understandable and less tedious to write.

(Refer Slide Time: 10:37)

Perfect squares

IIT Madras
ONLINE DEGREE

- Integers whose square root is also an integer

$\{z \mid z \in \mathbb{Z}, \sqrt{z} \in \mathbb{Z}\}$

Madhavan Mukund Sets: Examples Mathematics for Data Science 1, W

Perfect squares

IIT Madras
ONLINE DEGREE

- Integers whose square root is also an integer
- All squares are positive, so this is the same as
- Alternatively, generate all the perfect squares

$\{n^2 \mid n \in \mathbb{N}\}$

Madhavan Mukund Sets: Examples Mathematics for Data Science 1, W

Perfect squares



- Integers whose square root is also an integer

$$\{z \mid z \in \mathbb{Z}, \sqrt{z} \in \mathbb{Z}\}$$

- All squares are positive, so this is the same as

$$\{n \mid n \in \mathbb{N}, \sqrt{n} \in \mathbb{N}\}$$

- Alternatively, generate all the perfect squares

$$\{n^2 \mid n \in \mathbb{N}\}$$

- Extend the definition to rationals

$$\frac{9}{16} = \left(\frac{3}{4}\right)^2 \text{ is a square, } \frac{1}{2} \neq \left(\frac{p}{q}\right)^2 \text{ for any } p, q \text{ is not}$$



Madhavan Mukund

Sets: Examples

Mathematics for Data Science 1, V

Perfect squares



- Integers whose square root is also an integer

$$\{z \mid z \in \mathbb{Z}, \sqrt{z} \in \mathbb{Z}\}$$

- All squares are positive, so this is the same as

$$\{n \mid n \in \mathbb{N}, \sqrt{n} \in \mathbb{N}\}$$

- Alternatively, generate all the perfect squares

$$\{n^2 \mid n \in \mathbb{N}\}$$

- Extend the definition to rationals

$$\frac{9}{16} = \left(\frac{3}{4}\right)^2 \text{ is a square, } \frac{1}{2} \neq \left(\frac{p}{q}\right)^2 \text{ for any } p, q \text{ is not}$$

$$\{\frac{q}{p} \mid q \in \mathbb{Q}, \sqrt{q} \in \mathbb{Q}\}, \text{ or } \{q^2 \mid q \in \mathbb{Q}\}$$

- Choose the generator as required



Madhavan Mukund

Sets: Examples

Mathematics for Data Science 1, V

So, let us look at one more round of examples. So, we saw this before, we talked about perfect squares. So, we said that some integers are squares of other integers and some integers are not squares. In particular, those which are not squares, their square roots are actually irrational. We proved for instance, in our supplementary lecture, that the $\sqrt{2}$ is irrational. So, perfect square is an integer such that its square root is also an integer. So, this is what this says, give me all the integers, which satisfy the condition that their square root is also an integer.

So, the square root of small z also belongs to a set of integers, give me all set Z and call it a perfect square. Now, notice that the square must be positive, we have already discussed this

because you multiply 2 negative numbers, you get a positive number, you multiply 2 positive numbers, you again get a positive number. So, in fact, a perfect square must always be non-negative, it could be 0.

So, we could as well assume that the target set is generated by the set of natural numbers. And that, we are only interested in the positive square root, so remember that, 4 has 2 square roots, the number 4 is either $(-2) \times (-2)$, or 2×2 , but it is sufficient to know that one of its square roots is an integer because the other one will just be the same with a minus sign.

So, we can as well define the same set of perfect squares in terms of the natural numbers, we generate all the natural numbers whose square roots are also a natural numbers. Now, we can turn this around and replace the filter by a condition. So, we know that every natural number when it is squared will give us a natural number. So, all the perfect squares will be generated in that form, take a natural number, square it. So, instead of looking for those numbers whose square root is a natural number, we can just take every natural number and square it.

So, we just generate all the natural numbers and without filtering them, we just take the output square. So, this also gives us 0 , 1^2 , 2^2 , 3^2 and so on. So, these are all different ways of writing the same thing. In one case, we replace the generating set from integers to natural numbers because of the property of perfect squares. In another case, we transformed the filter into a transformation. So, instead of putting a condition on the numbers that we are generating, we took all the numbers and then squared them to get the actual perfect squares.

Now we could extend the notion of perfect squares to other sets of numbers. For instance, rationals can also admit a definition of a perfect square, so a rational will be a perfect square if it is a square of another rational. In particular, a rational could be an integer, but we will now integers can also be above and below the line, so we could have $\frac{9}{16}$, for instance as a rational number, which is $\frac{3^2}{4^2}$, so, $\frac{3}{4} \times \frac{3}{4} = \frac{9}{16}$.

So, we might want to say that this is a perfect square in the world of rationals. And not everything is a perfect square because since $\sqrt{2}$ cannot be represented a rational, it is easy to

check that half cannot be represented with a form $(\frac{p}{q})^2$. So, not every rational in this sense is a perfect square, some are, some are not. So, we can again change the definition above and replace Z and N by Q and get a reasonable definition of perfect squares in a different domain of numbers.

So, we can say give me all the rationals q such that \sqrt{q} is also a rational. Or using the second form, we can say take all the rationals and square them. So, take every q, which is a rational q and give me q^2 . So, this says that depending on how you choose the generator, you might generate the same set, or you might generate a different set. So, it is important to specify all the parts of a set comprehension correctly, so that there is no ambiguity and so that you get the set that you mean to get.