Vocabulary:

I withheld some technical names so that you can familiarize and appreciate the concept and meaning before the term is introduced. But it is time to learn the terms so that you can follow literature (and be able to communicate with others).

Recall that small sets are nowhere dense sets. A set A is nowhere dense if interior of its closure is empty, that is, the only open set contained in $\overline{A}$ is the empty set, that is, every non-empty open set contains points not in $\overline{A}$. Countable union of nowhere dense sets is called a set of the first category. A set which is not of first category is called set of second category. A set whose complement is of first category is called a *residual set*.

For example in the real line R the set Q of all rationals is of first category; the set A of all irrational numbers is a residual set; the interval $(0, 1)$ is a set of second category. The interval $(0, 1)$ is not residual because its complement is not of first category.

Thus residual sets are not only large sets, their complement is first category. On the other hand a set and its complement may both be second category. Of course, if you have a complete space then a set and its complement can not both be of first category because then whole space will be of first category, violating Baire.

In discussing compactness we came across the concept: Given any $\epsilon > 0$, finitely many balls of radius ϵ cover the space. When a space has this property, it is called *totally bounded*.

Recall that a set is bounded if there is a number M such that $d(x, y) \leq M$ for all $x, y \in M$. equivalently, there is an M such that $d(x_0, y) \leq M$ for all $y \in M$ where x_0 is some fixed point. if this last statement holds for one x_0 then it holds for all points x_0.

A totally bounded set is bounded. In fact take $\epsilon = 1$ and say k balls of radius one with centres at $\{x_i : 1 \leq i \leq k\}$ cover the set then for any points

x, y

$$d(x, y) \le \max\{d(x_i, x_j) : 1 \le i, j \le k\} + 2.$$

In fact, take two points x, y, say $x \in B(x_i, 1)$ and $y \in B(x_j, 1)$, then

$$d(x, y) \le d(x, x_i) + d(x_i, x_j) + d(x_j, y) \le 2 + d(x_i, x_j).$$

You should not mistakenly think that the distances are bounded by some number like $2k + 2$ or some thing.

For example if you take $X = \{0, 1000\}$ with distance $|x - y|$ then two balls of radius one cover. But the diameter is far far large.

However a bounded set need not be totally bounded. We knew that even on the real line the metric $\min\{|x - y|, 1\}$ is bounded but clearly the space is not totally bounded; for no $\epsilon < 1$ you can cover by finitely many balls of that radius.

Thus we can state criterion for compactness as: completeness plus totally bounded.

In a metric space X a set A is said to be an *ϵ-net* if every point of the space within ϵ of some point of A; that is,

$$x \in X \Rightarrow (\exists\, p \in A) \;\; d(x, p) < \epsilon.$$

Equivalently, $B(A, \epsilon) = X$.

For example in R the set of integers is an 1-net, also 3/4-net, but not a 1/2-net. An ϵ-net need not be finite. Of course, in a totally bounded set, you can get finite ϵ-net for every $\epsilon > 0$. In fact, this can be taken as the definition of totally bounded.

We considered $A \subset C[0, 1]$ in Arzela-Ascoli theorem. It has the property that there is one bound for all functions in A. We then say that the set A is uniformly bounded or equi-bounded. Thus a subset A is *equi-bounded or uniformly bounded* if there is a number M so that $|f(x)| \le M$ for all $f \in A$ and all x.

Similarly we had the property that given $\epsilon > 0$ there is one $\delta > 0$ such that $|f(s) - f(t)| < \epsilon$ for each $f \in A$ and each s, t with $|s - t| < \delta$ This property is called *equi-uniformly continuous* (or uniformly uniformly continuous, but this sounds awkward).

Thus we can state AA theorem as; A subset is compact iff it is closed, equi-bounded and equi-uniformly continuous.

We said that if a set satisfies conditions (ii) and (iii) then it may not be compact but its closure is compact. Such sets are called *pre-compact*. More precisely a set $A \subset X$ is called pre-compact if its closure is compact.

Thus a subset of $C[0,1]$ is pre-compact iff it is equi-bounded and equi-uniformly continuous. A subset of R is pre-compact iff it is bounded (in the usual metric).

Arzela-Ascoli:

No new idea is needed to see that the AA theorem holds for any compact metric space. This is what we explain now.

Let (X, d) be a compact metric space. Thus every real valued continuous function on X is bounded. Let $C(X)$ be the space of real continuous functions on X with sup metric

$$d(f, g) = \sup_x |f(x) - g(x)|.$$

Firstly this is a metric on the space. The space is complete. Proof is exactly as for $C[0,1]$. Take $\{f_n\}$ Cauchy (remember sup metric); then for each x, the sequence of numbers $\{f_n(x)\}$ is Cauchy; say $f(x)$ is its limit; argue f_n converges to f uniformly; conclude f is continuous and $d(f_n, f) \to 0$.

Arzela-Ascoli Theorem: A subset $K \subset C(X)$ is compact iff the following three conditions hold.

(i) K is closed

(ii) There is M such that $|f(x)| \leq M$ for all $f \in K$, all $x \in X$.

(iii) Given $\epsilon > 0$ there is $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ for $f \in K$ and $x, y \in X$ with $d(x, y) < \delta$.

Proof is as earlier. suppose K is compact. Since compact sets are closed (i) holds. Since $\Psi(f) = d(f, 0)$ is a continuous function on K it is bounded giving (ii). Given $\epsilon > 0$ choose finitely many balls of radius $\epsilon/3$ covering K; using uniform continuity take $\epsilon/3$ and get δ for these finitely many centers and argue that the same δ works for all $f \in K$; showing (iii).

Let now K satisfy the conditions. To show it is compact take a sequence $\{f_n\}$ in K. Need to exhibit a subsequence converging to a point of K. Remember that a compact space is separable; take a countable dense set D in X; use hypothesis (ii) and diagonal argument to get a subsequence $\{f_{n_k}\}$ such that for every point $p \in D$, the sequence of numbers $\{f_{n_k}(p)\}$ converges; use condition (iii) to argue $\{f_{n_k}\}$ converges, use condition (i) to say that this limit is in K — showing K is compact.

again as earlier if a set satisfies conditions (ii) and (iii) then its closure also satisfies those conditions and hence the closure is compact.

Compact metric spaces arise often in practice. For example the space of infinite sequences of $+1$ and -1 is a very natural phase space for a spin system consisting of very very large number of particles. For each configuration you can associate energy or magnetization and so on, very useful functions. (Actually you do so on finite products and then take limits. but let us not bother).

Stone-Weierstrass:

We shall now discuss probably the most complicated theorem of our course. This generalizes the Weierstrass theorem, you learnt last year, which says the following. Given any real valued continuous function f on the interval $[0, 1]$, and $\epsilon > 0$, there is a polynomial P such that $|P(x) - f(x)| < \epsilon$ for each $x \in [0, 1]$. With our present notation we can restate this as follows: given $f \in C[0, 1]$ and $\epsilon > 0$, there is a polynomial P such that $d(f, P) < \epsilon$. Or simply put, the set of polynomials is dense in $C[0, 1]$.

Given any polynomial P, and $\epsilon > 0$, there is a polynomial Q with rational coefficients such that $d(P, Q) < \epsilon$. Since the set of polynomials with rational coefficients is countable, we can also conclude that the space $C[0, 1]$ is separable. Thus it is a separable complete metric space or a Polish space.

The theorem that we are going to discuss is a generalisation of the above. Let X be a compact metric space and $C(X)$ be the space of real valued continuous functions on X with sup metric.

Theorem:
Let $D \subset C(X)$ where X is a compact metric space. If D satisfies he

following three conditions then it is dense in $C(X)$.

(i) D is an algebra. That is, sum, product, constant multiple of functions in D is again in D.

(ii) Constant functions are in D.

(iii) D separates points. That is, given two points $p \neq q$ in X, there is an $f \in D$ with $f(p) \neq f(q)$.

Of course we could have stated (ii) to simply say that the constant function $x \equiv 1$ is in D. Then by (i) every constant function would also be in D.

Note that if condition (iii) fails then D can not be dense. This is obvious because if every function in D takes the same value at the two points p and q then so do their limits and hence it can not be dense. For example the function $f(x) = d(p, x)$ will not be in the closure of D.

That D is an algebra simply means that it is a vector space and closed under multiplication.

This theorem gives new information even in the case of $[0, 1]$. For example, if you take any integer $k \geq 1$ then the set of polynomials

$$a_0 + a_1 x^k + a_2 x^{2k} + a_3 x^{3k} + \cdots + a_n x^{nk}; \quad n \geq 1, a_i \in R$$

is dense in $C[0, 1]$. That is, polynomials in powers of x^k are dense. For example, you can consider powers of $x^{1000000}$ and their linear combinations.

Or you can take polynomials in $\sin x$. They are dense too. That is, the set of functions

$$a_0 + a_1 \sin x + a_2 (\sin x)^2 + a_3 (\sin x)^3 + \cdots + a_n (\sin x)^n : \quad n \geq 1; a_i \in R$$

is dense. Of course these you can obtain from a clever change of variable from usual Weierstrass theorem.

Proof of the theorem:

Let $\overline{D}$ be the closure of D. We understand the set enough to conclude that it must be all of $C(X)$.

1^o. The set $\overline{D}$ is again an algebra separating points and containing constants.

Since D already has the last two properties any larger set, in particular, closure has those properties. We only need to show that it is an algebra. But

it is clear from the fact that if $f_n \to f$ and $g_n \to g$ then
$f_n + g_n \to f + g$ and $cf_n \to cf$ and $f_n g_n \to fg$. remember all these are uniform convergences.

2^0. D and hence $\overline{D}$ strongly separates points: given $p \neq q \in X$ and numbers a, b there is $f \in D$ such that $f(p) = a$ and $f(q) = b$.

Indeed take φ which takes different values at p and q and define

$$f(x) = \frac{\varphi(x) - \varphi(q)}{\varphi(p) - \varphi(q)} a + \frac{\varphi(x) - \varphi(p)}{\varphi(q) - \varphi(p)} b$$

Even though φ appears at several places, only numbers calculated from it appear mostly. The constant function $\varphi(q)$ is in D and hence so is $\varphi - \varphi(q)$ and so is the ratio appearing in the first term above and so is the first term itself. Thus it is a function in D. So is the second term. so is their sum. It takes the required values at the given points.

Thus we have control on values we want the function to take at two given points. We can still get such a function from D.

3^o. $f \in \overline{D}$ implies $|f| \in \overline{D}$.

Indeed, take $\epsilon > 0$. Let the range of f be contained in the interval $[c, d]$. By usual Weierstrass theorem get a polynomial $P(t)$ such that for all t in this interval $||t| - P(t)| < \epsilon$. In other words the function $|t|$ is approximated by a polynomial on this interval. It is clear that $||f(x)| - P(f(x))| < \epsilon$, for all $x \in X$. In other words the distance between the two functions $|f|$ and $P(f)$ is at most ϵ. Remember that if p is the polynomial,

$$P(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k$$

then $P(f)$ is the function

$$a_0 + a_1 f + a_2 f^2 + \cdots + a_k f^k.$$

This is in $\overline{D}$ because f is in it and it is an algebra. Thus there are elements close to $|f|$ which are in $\overline{D}$. The later being closed, we conclude that $|f|$ itself is in it.

4^o.

$$f, g \in \overline{D} \Rightarrow f \vee g, f \wedge g \in \overline{D}.$$

Recall

$$f \vee g(x) = \max\{f(x), g(x)\}; \quad f \wedge g = \min\{f(x), g(x)\}.$$

The claim is immediate from the previous observation and the fact

$$f \vee g = \frac{f+g+|f-g|}{2}; \quad f \wedge g = \frac{f+g-|f-g|}{2}.$$

5^o. Fix $f \in C(X)$, fix a point $p \in X$ and fix a number $\epsilon > 0$. Then there is a function $h \in \overline{D}$ such that

$$h(p) = f(p); \quad h(x) < f(x) + \epsilon$$

Of course this function depends on p and should have been denoted by h_p but to lighten the notation we did not do that. Thus we can get a function in our set which is close to a given f. Well, not really close; but does not go too far above f; it may however go too far below f.

Proof is simple. For any $q \neq p$ fix a function $\varphi_q \in \overline{D}$ which takes same values as f at both p and q. This is possible by previous observation. Let

$$U_q = \{x : \varphi(x) < f(x) + \epsilon\}.$$

Then for each q we have an open set conning the point q and of course all these sets contain p as well thus they cover X. Pick finitely many, say

$$U_{q_1};\ U_{q_2};\ U_{q_3};\ \cdots U_{q_k}.$$

Consider the function

$$h = \varphi_{q_1}\ \wedge\ \varphi_{q_2}\ \wedge\ \cdots \varphi_{q_k}.$$

Since each $\varphi_q(p) = f(p)$ we see $h(p) = f(p)$. Given any point $x \in X$ it is in some U_{q_i} and so $\varphi_{q_i}(x) < f(x) + \epsilon$ and so $h(x) < f(x) + \epsilon$. Since each $\varphi_q \in \overline{D}$ a previous observation tells us that $h \in \overline{D}$.

6^o. Fix $f \in C(X)$ and $\epsilon > 0$ the there is a $g \in \overline{D}$ such that

$$f(x) - \epsilon < g(x) < f(x) + \epsilon; \quad \forall\, x \in X.$$

In other words $d(f, g) < \epsilon$. Since this is true for every $\epsilon > 0$ and since $\overline{D}$ is closed we conclude that $f \in \overline{D}$. Since $f \in C(X)$ is arbitrary we conclude that $\overline{D} = C(X)$. In other words D is dense in $C(X)$; proving the theorem.

This is simple. For each $p \in X$, use earlier observation to get h, now let us denote it by h_p so that

$$h_p \in \overline{D}; \quad h_p(p) = f(p); \quad h_p(x) < f(x) + \epsilon\ \ \forall\, x \in X.$$

Set

$$V_p = \{x : h_p(x) > f(x) - \epsilon.\}$$

Then $p \in V_p$. In other words the family of all V_p cover the space X and take a finite sub family that covers, say,

$$V_{p_1};\ V_{p_2};\ \cdots; V_{p_k}.$$

Set

$$g = h_{p_1} \vee h_{p_2} \vee \cdots \vee h_{p_k}.$$

Given any point x, it is in some V_{p_i} and so $h_{p_i}(x) > f(x) - \epsilon$ and hence $g(x) > f(x) - \epsilon$. Also given $x \in X$ each $h_p(x) < f(x) + \epsilon$ and hence so is the finite max $g(x)$. By earlier observation, we see $g \in \overline{D}$.

This completes the proof.

The 'type' of argument to prove a little of what is needed in 5^o and then improve it in 6^o is called boot-strapping.

We have already seen implications of this theorem for continuous functions on the interval $[0, 1]$. We can use this theorem to show that for a compact metric space X, the space $C(X)$ is a separable metric space. Thus it is a Polish space (complete separable metric space).

We can also show that trigonometric polynomials are dense in the space of continuous functions f on $[0, 2\pi]$ with $f(0) = f(2\pi)$.

Space Filling Curves or Peano Curves:

Cantor startled the community by showing that R^2 has same number of points as R. But of course, he could exhibit only some identification and it was unclear whether one could get a continuous function.

Peano startled the community by showing that there is a continuous function from the unit interval $[0, 1]$ onto the unit square $[0, 1] \times [0, 1]$. After all, a continuous function from the interval to the plane should 'trace a curve' which should be 'one dimensional'. But that mental picture was proved to be wrong. This curve of Peano fills up all of the square. That is why such functions from then on are called Peano curves or space filling curves.

Soon after words Hilbert gave another geometric construction which is more fertile, in the sense, it inspired others and also gave a general method of construction. It also helped to evaluate some interesting properties such maps can possess.

One knew properties such maps can not possess. For example such a map can not be one-one. If f is a continuous map of a compact metric space X onto another compact metric Y space which is one-to one; then its inverse is already continuous and hence such a map is a homeomorphism. In fact, if V is open in X, then V^c is closed and hence compact (because X is so) and hence so is $f(V^c)$ and hence it is closed in Y and hence $f(V) = [f(V^c)]^c$ is open. This shows that f^{-1} is continuous.

In our case $X = [0,1]$ and $Y = [0,1] \times [0,1]$ and we knew that they can not be homeomorphic. It was also known that such maps can not be differentiable.

At that time when they were found, these were curious objects. But now-a-days they are finding applications in computer science. After all, by talking about point t on the line you can mean point $f(t)$ in the square. In other words, you can stay on the line and talk about two dimensional things. This was found useful in data organisation. I do not know much of it.

However, there are some interesting algorithmic questions where these maps are found useful — for example, for the travelling salesman problem (known to be 'hard' in some sense). This problem has a data of finitely many points in the unit square and asks for a tour of these points which is economical; shortest possible. There is no idea how to make a beginning of a tour. If you have a Peano curve you can make a beginning (good or bad). Take those points given to you, take 'some' points $\{t_i\}$ such that your data is $\{f(t_i)\}$ and tour in the increasing order of the points t_i. In other words the problem is linearized, an order has been brought.

I shall describe the Hilbert method but do not carry out because it needs some vocabulary. Our interest is only to make it clear to ourselves that such curves exist. We are not interested in either studying their properties or in using them. So we follow a simpler method later and completely work out. But you should know Hilbert method.

Divide the interval $[0,1]$ into four parts

$$I_1 = [0, 1/4];\ I_2 = [1/4, 1/2];\ I_3 = [1/2, 3/4];\ I_4 = [3/4, 1].$$

Divide the square also into four parts

$$R_1 = [0,1/2] \times [0,1/2]; \quad R_2 = [0,1/2] \times [1/2,1]$$

$$R_3 = [1/2,1] \times [1/2,1]; \quad R_4 = [1/2,1] \times [0,1/2].$$

Draw a curve joining by straight lines, the mid point of R_1 to mid point of R_2 to R_3 to R_4. Set up a map f that takes I_j to the part of your curve in R_j. This is our first function f_1.

Now divide each I_j into I_{jk} for $k = 1,2,3,4$, and similarly the squares. assign the four parts of the line I_j to the four parts of the square R_1 and so on and draw a curve made up of straight lines. This is our second function f_2. You need to be careful and give an algorithm on how to do it. This can be done.

To do this you need either picture or vocabulary. But the point is that these functions, so obtained, converge to a function which is continuous and its image is all of the square. We shall not pursue this process.

We exhibit our curve (due to Lebesgue) after two observations.

Remember the construction of the Cantor set. We define

$$I_0 = [0,1/3]; \qquad I_2 = [2/3,1]$$

$$I_{00} = [0,1/3^2];\ I_{02} = [2/3^2,3/3^2];\ I_{20} = 6/3^2,7/3^2];\ I_{22} = [8/3^2,9/3^2].$$

and so on.

$$C_0 = [0,1];\ C_1 = I_0 \cup I_2;\ C_2 = \cup\{I_{ij}; i,j = 0,2\}$$

$$C = \cap C_n$$

is the Cantor set.

Observation 1: There is a continuous function f defined on the Cantor set onto the unit interval $[0,1]$.

This is simple. we know that a point x is in the Cantor set iff it has triadic expansion involving only the digits zero and two. That is,

$$x = \frac{x_1}{3} + \frac{x_2}{3^2} + \frac{x_3}{3^3} + \cdots; \quad x_i = 0,2.$$

equivalently

$$x = \frac{2x_1}{3} + \frac{2x_2}{3^2} + \frac{2x_3}{3^3} + \cdots; \quad x_i = 0, 1.$$

Keep in mind that such an expansion is unique. It is worth remembering that the point $1/3$ is in C and it has the eligible ternary expansion $.0222222\cdots$. Also remember given any sequence of zeros and ones, the series $\sum(2x_i)/3^i$ converges and the number so defined is a point in $[0,1]$ and is in the Cantor set.

Here then is the function

$$f(x) = \sum \frac{x_i}{2^i}; \qquad x = \sum \frac{2x_i}{3^i}.$$

This is good definition because, given a point $x \in C$ it has a unique eligible ternary expansion (that is involving zeros and twos).

Clearly given any number $y \in [0,1]$, it has a binary expansion $\sum y_i/2^i$ with each y_i zero or one and then the point $x = \sum(2y_i)/3^i$ is in C and $f(x) = y$ to show that f is onto $[0,1]$.

Finally continuity of f is seen as follows. let $x_n \to x$, all in C. If $x \in I_0$ then first digits of the x_n after some stage must be zero. Otherwise $x \in I_2$ and after some stage the first digit of the x_n must be two.

Similarly depending to which of the four intervals I_{ij} the point x belongs, we see that the second digit of x_n must be same as that of x after some stage. In other words, given any k, $f(x_n)$ will have the same first k digits as that of $f(x)$ after some stage. This shows continuity.

Observation 2: there is a continuous one-one map g on C onto $C \times C$.

Of course any such map is a homeomorphism, as noted earlier. We are thinking $C \times C \subset [0,1] \times [0,1]$. If $x \in C$ and

$$x = \frac{x_1}{3} + \frac{x_2}{3^2} + \frac{x_3}{3^3} + \cdots; \quad x_i = 0, 2.$$

we put

$$g_1(x) = \frac{x_1}{3} + \frac{x_3}{3^2} + \frac{x_5}{3^3}\cdots; \qquad g_2(x) = \frac{x_2}{3} + \frac{x_4}{3^2} + \frac{x_6}{3^3} + \cdots;$$

$$g(x) = (g_1(x), g_2(x)).$$

Since each x_i is zero or two, we see that both $g_1(x)$ and $g_2(x)$ are in C. We have just separated the even and odd digits. This is one-one. Remember an eligible expansion is unique for points in C. This is also onto $C \times C$ because given two points in C we can interlace the digits of these two points to get a point of C whose image is the given pair.

That it is continuous is immediate because the digits are unchanged. (of course, this is to be interpreted carefully, there is a change in the 'place' of the digit).

Observation 3: Given a continuous function h on C to $S = [0, 1] \times [0, 1]$ we can extend it to a continuous map of $I = [0, 1]$ to S.

The only issue is to define the function on the points of the deleted intervals. Let $x \in (a, b)$ a deleted interval. Thus $a, b \in C$. Hence we know $h(a)$ and $h(b)$. Extend linearly in between. More precisely put

$$h((1-\theta)a + \theta b) = (1-\theta)f(a) + \theta f(b); \qquad 0 \leq \theta \leq 1$$

Since unit square is convex, h so defined still takes values in S. As the point θ goes from 0 to 1; the point $t = (1-\theta)a + \theta b$ travels from a to b and the h values travel on the straight line from $h(a)$ to $h(b)$. We are stll using the same symbol h for this extended function. Thus now h is defined on all of I and takes values in S.

We need to show that h so defined is continuous at all points. Of course, it is continuous at all points outside the Cantor set. Indeed, let $x \in I_n$, one of the deleted open intervals and $x_i \to x$. Then after some stage each $x_i \in I_n$ and so the definition given above takes over to show continuity.

Let $x \in C$. Let $x_i \to x$. We need to show that $h(x_i) \to h(x)$. Given $\epsilon > 0$ plot (a', b') around x so that for every point $y \in C \cap (a', b')$ we have $d(f(x), f(y)) < \epsilon$. This is possible because f is given to be continuous on C. Let us take $a, b \in C$, $a' < a < x < b < b'$. This is possible because every point of C is a limit point of C.

We now claim that for every point y in this interval (a, b), whether y is in C or not, we have the inequality $d(h(x), h(y)) < \epsilon$. Since this last inequality holds for points of $y \in C$ in this interval, we only need to show that it holds for points not in C. But then any such point is in a deleted interval (a_n, b_n).

Since $a, b \in C$ we conclude that

$$d(h(a_n), h(x)) \le \epsilon; \quad d(h(b_n), h(x)) \le \epsilon.$$

In other words, both $h(a_n)$ and $h(b_n)$ are in the ϵ-ball around $h(x)$. Remember balls are convex. Since $h(y)$ is on the line joining $h(a_n)$ and $h(b_n)$, we see that $d(h(y), h(x)) \le \epsilon$. This shows continuity.

We shall now exhibit space filling curve.

Take $f : C \to [0, 1]$ as in observation (1).
Take $g(x) = (g_1(x), g_2(x)) : C \to C \times C$ as in observation (2).
Define $h : C \to S$ by

$$h(x) = (h_1(x), h_2(x)); \quad h_1(x) = f(g_1(x));\ h_2(x) = f(g_2(x)).$$

The map h is a continuous function on C — it is composition of continuous functions. Given $(a, b) \in S$; pick $s, t \in C$ with $f(s) = a$ and $f(t) = b$ and pick $x \in C$ with $g(x) = (s, t)$. Then verify $h(x) = (a, b)$. This shows that h is onto S. Now use observation (3) to extend the map as a continuous map from all of I to S. Of course, it still remains onto S.

This does.

Cantor sets:

We discussed once that theoretically one could construct similar to the Cantor set, by cutting into more number of pieces. For example, at each stage you divide an interval into five subintervals of equal length; instead of three. You pick the first, third and fifth; instead of the first and last. The procedure remains same. Such sets are also called Cantor sets. Here is the precise definition.

A compact subset of R is called *perfect* if every point is a limit point of the set. For example the usual Cantor set, or the interval $[0, 1]$ are such sets. The set

$$\{0,\ 1,\ 1/2,\ 1/3,\ \cdots\}$$

is closed but not perfect.

The set Q of all rationals is not closed, but every point of it is a limit point of the set.

A perfect nowhere dense set is called *a Cantor set*. This means our set should be

closed,
bounded,
every point must be a limit point,
should not contain any (non-empty) open interval.

Thus usual Cantor set is an example of a Cantor set, it is called *the Cantor set*. The construction described above will also lead to a Cantor set. Mathematics, like music, can be improvised with your ideas. For example should we cut the interval into fixed number of pieces? First step cut into three pieces and leave even one, that is, the middle one. You select the other two intervals. Next, cut each selected part into five pieces and leave even ones, that is, the second and fourth part in each of the first level intervals. You have six intervals. Next cut each selected interval into seven parts etc.

It is all your pick! If you want you can divide into p_n subintervals at the n-th stage. Here p_n is the $(n+1)$-th prime.

You would probably think that you are getting many many very different sets. But this is only illusion as far as sets are concerned.

Here is a very interesting theorem: Such sets are all the same! More precisely, let P and Q be two such sets. Then there is a homeomorphism h of R to itself which sends P to Q. This last phrase means $h(P) = Q$. This is proved exactly like a similar theorem we proved quite a while ago, namely, given two countable dense sets in R, there is a homeo that sends one to the other. We shall not carry out the proof.

You should not come to the erroneous conclusion that all such sets are same. No. We only said that they *look same*. But different sets have *different qualities and expertise* from analysis/arithmetic point of view. We shall not pursue this line of thought. We shall leave aside our long excursion to metric spaces and return to real line once again.