Neural Networks for Signal Processing-I

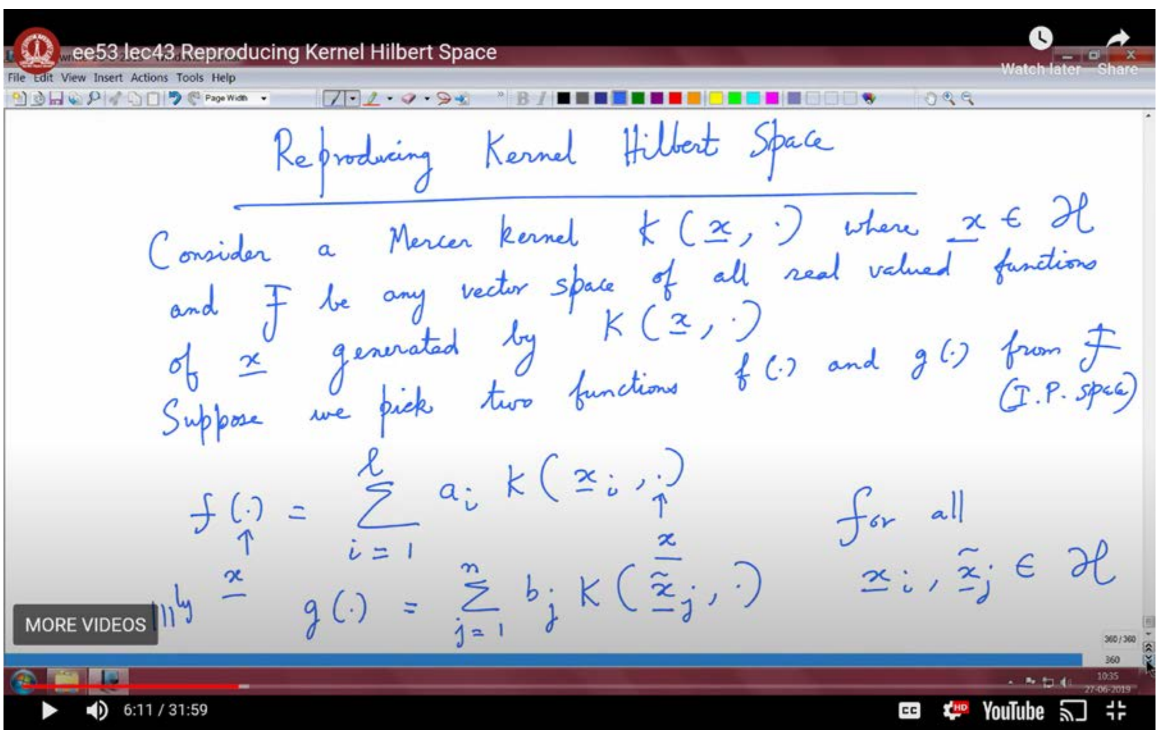
Prof. Shayan Srinivasa Garani
Department of Electronic System Engineering
Indian Institute of Science, Bengaluru

Lecture – 43

Reproducing Kernel Hilbert Space

In this module, let's delve into the concept of a Reproducing Kernel Hilbert Space (RKHS). To set the stage, recall our journey from a basic vector space, through the notions of a normed vector space and inner product space, culminating in the concept of convergence within a Hilbert space. With that foundational understanding, we're now prepared to explore what an RKHS is, particularly in the context of Mercer's theorem and the kernels that arise from it.

(Refer Slide Time: 06:11)



To begin, we need to establish some definitions. Consider a Mercer kernel k, where x is an element of a Hilbert space H. Now, remember that when we refer to a Hilbert space, it inherently implies a complete inner product space. All the detailed discussions we've had about completeness, inner products, and so forth are applicable here.

Next, let's consider any vector space F of all real-valued functions of x, which are generated by this kernel. We pick two functions f and g from this space F. The function f can be expressed as a linear combination of kernels, specifically as $f(x) = \sum_{i=1}^{l} a_i k(x_i, x)$, where the dot in the kernel notation $k(x_i, \cdot)$ represents the function of x. Similarly, the function g can be expanded as $g(x) = \sum_{j=1}^{n} b_j k\left(x_j, x\right)$.

For any x_i, x_j belonging to this Hilbert space, these functions f and g can be expanded using the kernel. Recall that this kernel can also be viewed as an inner product of two vectors, as we saw during our discussion of Support Vector Machines (SVMs). According to Mercer's theorem, these kernels can be interpreted as eigenfunction-eigenvalue pairs, further clarifying the expansion of functions using kernels.

Now, consider the bilinear form, which is essentially the inner product of f and g. This can be written as:

$$\langle f, g \rangle = \sum_{i=1}^{l} \sum_{j=1}^{n} a_i b_j k\left(x_i, x_j\right)$$

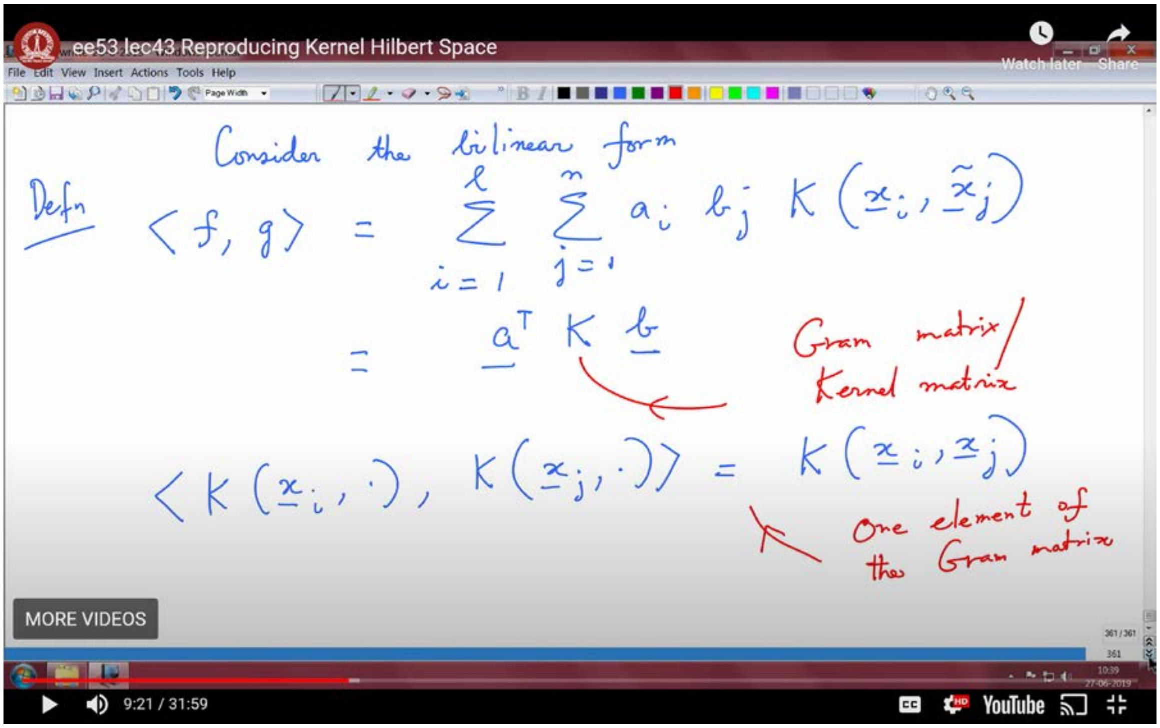
This bilinear form defines our inner product in terms of the kernel, and can also be represented as:

$$\langle f, g \rangle = a^{\top} K b$$

where K is the Gram matrix (or kernel matrix) formed from the kernel k(x_i, x_j), and a and b are vectors of coefficients a_i and b_j, respectively.

The Gram matrix K is a key concept here, as it encapsulates the relationships between the points x_i and x_j in the space via the kernel. Each element of this matrix k(x_i, x_j) corresponds to the inner product of the kernel evaluated at points x_i and x_j.

(Refer Slide Time: 09:21)



Now, let's reconsider the inner product in a slightly different form. The bilinear form discussed can also be written as:

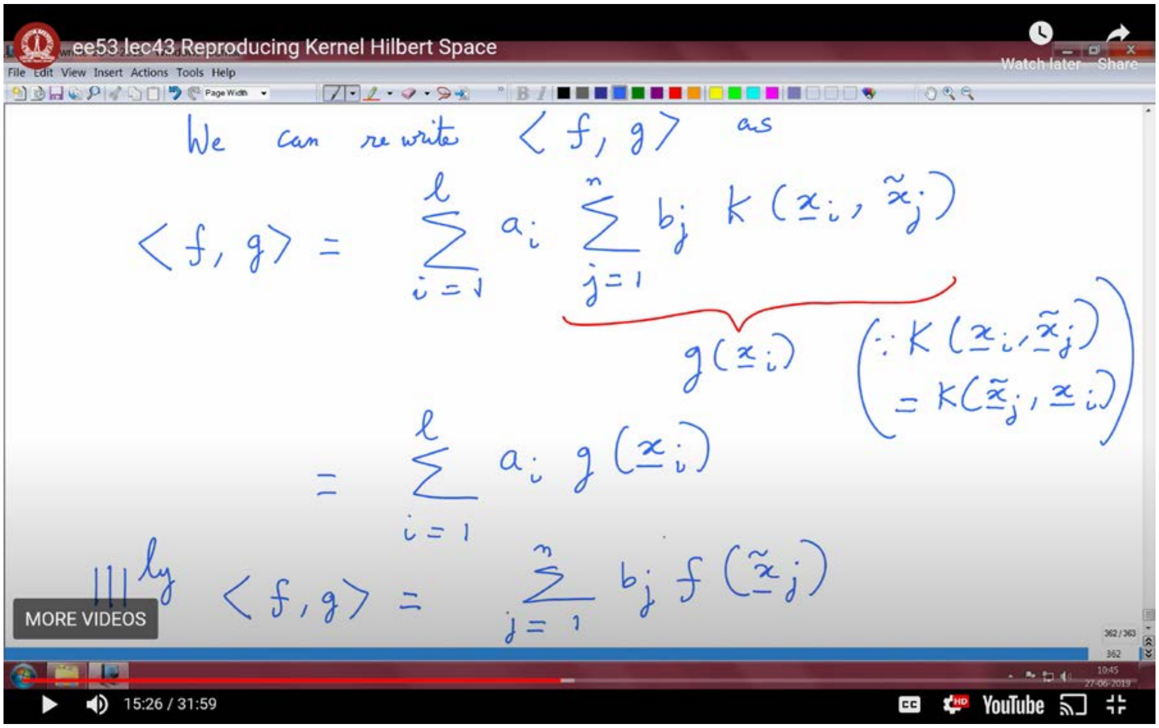
$$\langle f, g \rangle = \sum_{i=1}^{l} \sum_{j=1}^{n} a_i b_j k(x_i, x_j)$$

This representation shows that the inner product in an RKHS can be expressed directly in terms of the kernel function, further emphasizing the role of the kernel in defining the geometry of the space.

With this understanding, we're well-positioned to explore the intricacies of RKHS and how they relate to kernel methods in various applications. Let's continue this journey, keeping in mind the significance of kernels and the structure they impose on the spaces we work within.

So, here's what I'll do: I'll start by taking the definition of the bilinear form, which is given as the double summation $\sum_{i=1}^{l} a_i \sum_{j=1}^{n} b_j k(x_i, \tilde{x}_j)$. This bilinear form is important, and as you can see from the definition, it's essentially a sum over products of coefficients a_i and b_j with the kernel function applied to pairs $(x_i, \tilde{x}_j)$.

(Refer Slide Time: 15:26)

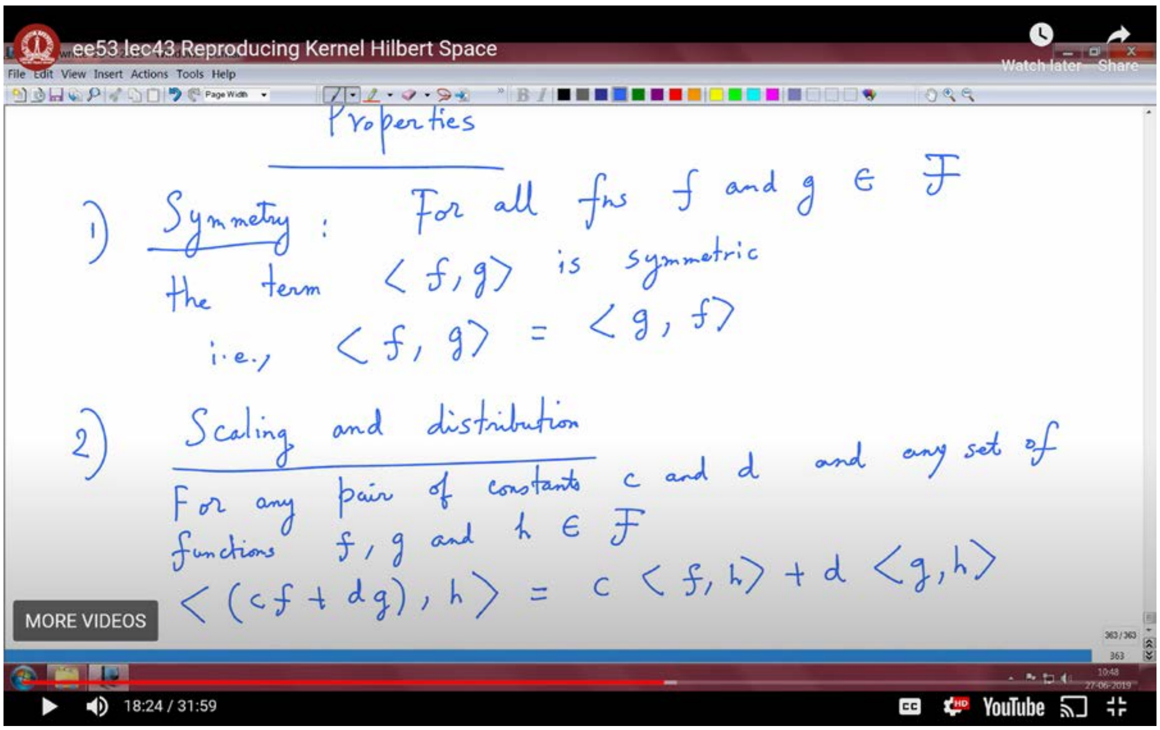


Now, let's focus on rearranging this expression. Specifically, I'll pull the a_i term outside the inner summation, so we can rewrite it in a different form. Pay close attention to the quantity $\sum_{j=1}^{n} b_j k(x_i, \tilde{x}_j)$ because, if you recall the definition, this expression is actually the function G evaluated at x_i. In other words, this sum is $G(x_i)$.

What's interesting here is the property we've utilized: the kernel function $k(x_i, \tilde{x}_j)$ is symmetric, meaning $k(x_i, \tilde{x}_j) = k(\tilde{x}_j, x_i)$. This symmetry is crucial because it allows us to exchange the arguments in the kernel function. If you recall the definition of G, it was initially written as $\sum_{j=1}^{n} b_j k(\tilde{x}_j, x)$, where the first argument is $\tilde{x}_j$ and the second is any x. But due to the symmetric property, we can switch these arguments, leading to the expression $G(x_i)$.

So, what's the takeaway? This inner product can be expressed as a linear combination of the function G evaluated at the points x_i. That's pretty fascinating, right? We began with the bilinear form of the inner product and ended up expressing it purely in terms of the function G.

(Refer Slide Time: 18:24)



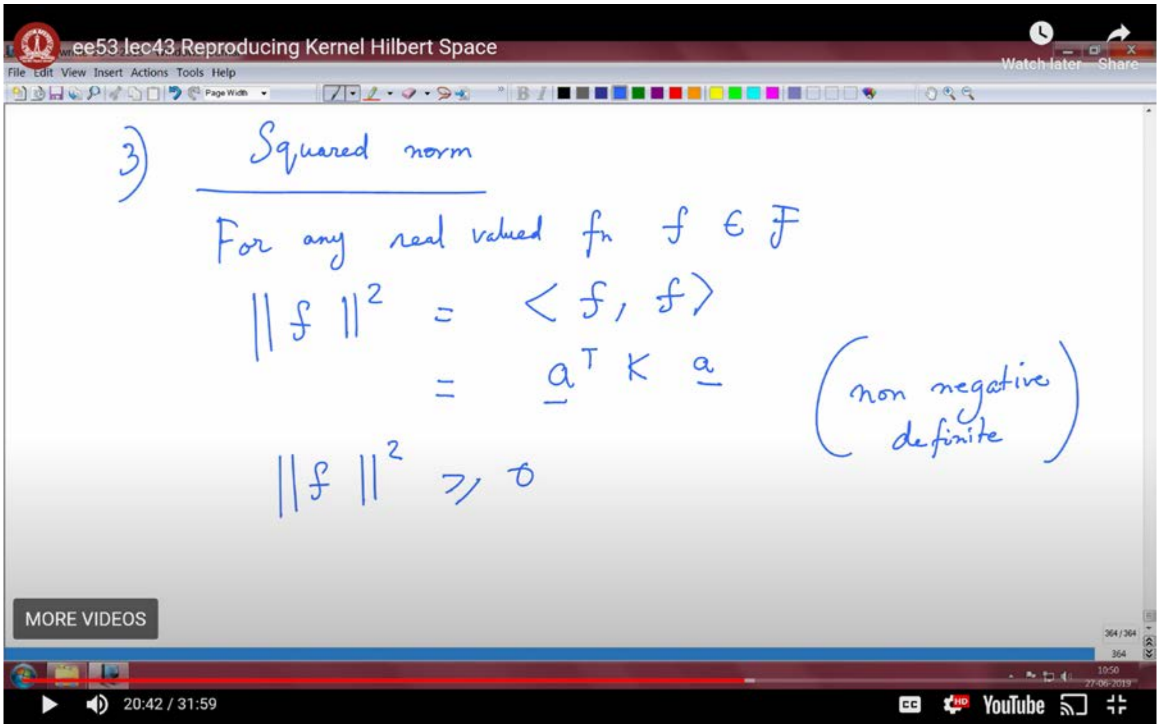
Now, a keen reader might notice that this process allows us to express the inner product starting from the bilinear form in terms of a linear combination of functions in F, evaluated over specific data points. You can follow the same steps to write the inner product of f and g as:

$$\langle f, g \rangle = \sum_{j=1}^{n} b_j f(\tilde{x}_j)$$

This implies that the definition of the bilinear form is independent of how the functions f and g are represented. In other words, the bilinear form is invariant to the particular indices used in the summation.

For instance, consider the expression $\sum_{i=1}^{l} a_i G(x_i)$. This expression is invariant to changes in the index n, the coefficient vector b, and the vector $\tilde{x}_j$. Similarly, the second form, $\sum_{j=1}^{n} b_j f(\tilde{x}_j)$, is invariant to changes in l, the vector a, and the points x_i.

(Refer Slide Time: 20:42)



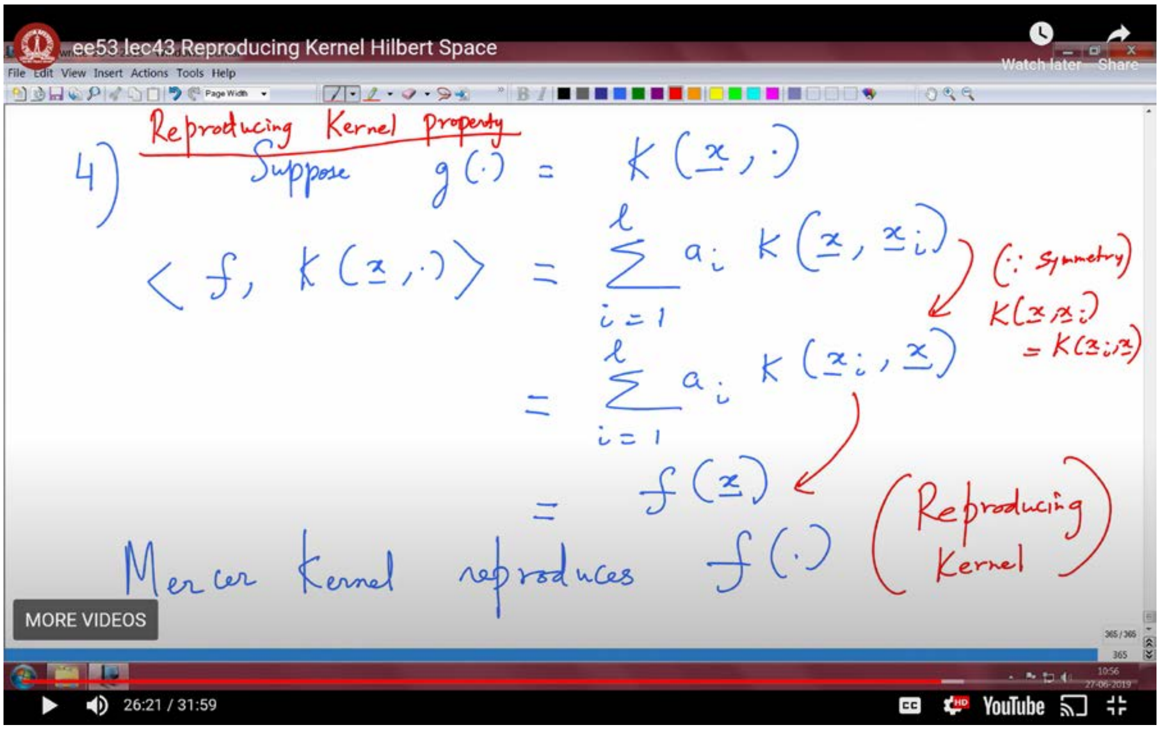
What this means is that we have two different ways to represent bilinear forms, and these properties of invariance are incredibly useful. Let's list some basic properties here. For all functions f and g that belong to this inner product space, the inner product $\langle f, g \rangle$ is symmetric. This means that taking the inner product of f and g is the same as taking the inner product of g and f. This symmetry is tied directly to the bilinear form, so if we exchange f and g in our arguments, the inner product remains unchanged.

This symmetric nature of the inner product is fundamental and highlights the consistency of the bilinear form with respect to its arguments.

We also have the scaling and distribution properties, which are crucial to understand in this context. Let me articulate the result. Consider any pair of constants, c and d, and any set of functions f, g, and h that belong to this inner product space, denoted as $\mathcal{F}$. The inner product of the linear combination of functions c f + d g with h can be expressed as:

$$\langle cf + dg, h\rangle = c\langle f, h\rangle + d\langle g, h\rangle$$

(Refer Slide Time: 26:21)



This illustrates the distribution property. As for the scaling property, if we have constants c and d, these scalars can be factored out of the inner product, similar to how it's done in vector spaces. However, it's important to be astute when considering a typical inner product space in terms of vectors. For instance, given two vectors x and y, the inner product $\langle x, y\rangle$ is essentially x^T y, which is equivalent to y^T x. But when dealing with functions that are expressed as linear combinations using kernels, we need to define this notion of an inner product using the bilinear form. From this bilinear form, the scaling and distribution properties naturally emerge.

Now, there's another essential property to discuss, the square norm property. For any real-valued function f belonging to this inner product space $\mathcal{F}$, the square of the norm $|f|^2$ is essentially the inner product of f with itself:

$$|f|^2 = \langle f, f \rangle$$

This can be expanded as:

$$|f|^2 = a^T K a$$

where a is a vector and K is a matrix. The norm is non-negative, implying that it is non-negative definite. This is a significant result.

Next, consider a function G, which is the kernel $k(x, x_i)$, where $g(x_i)$ represents the kernel acting on x and x_i, as indicated by the dot notation. Now, if we take the inner product of a function with this kernel, it becomes quite interesting.

By definition, this inner product can be expanded as:

$$\sum_{i=1}^{L} a_i k(x, x_i)$$

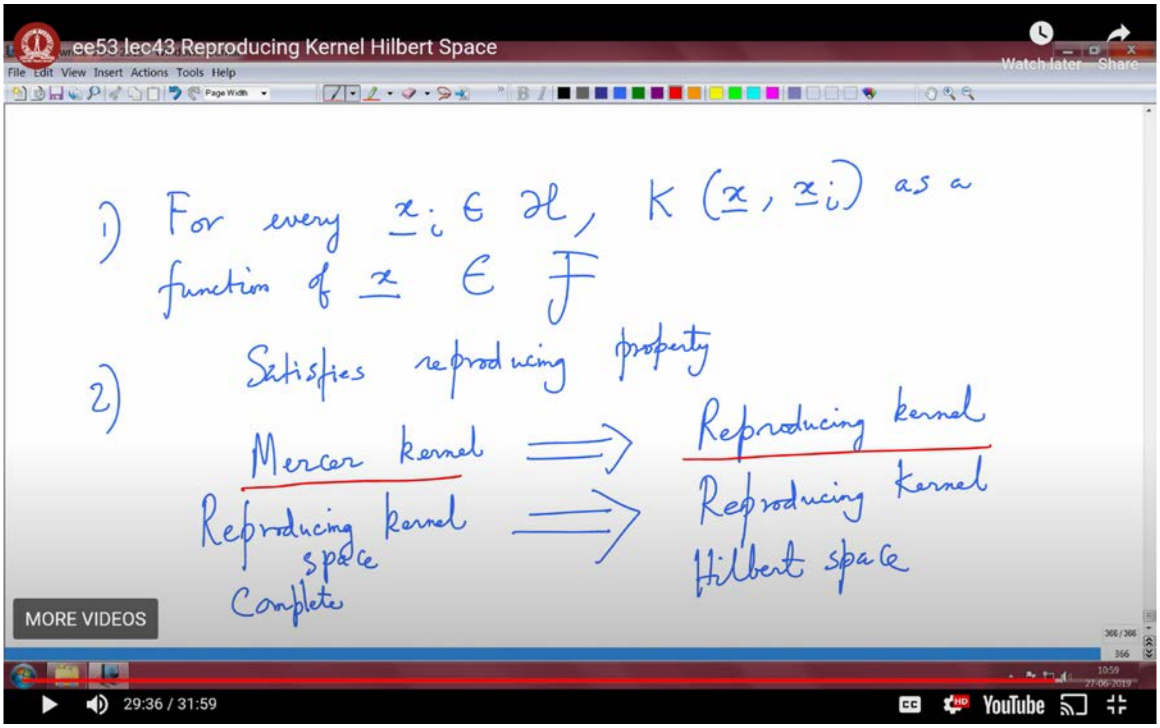
where $f(x_i)$ is our function evaluated at x_i. Now, due to the symmetric property of the kernel, we can express this as:

$$\sum_{i=1}^{L} a_i k(x_i, x)$$

which simplifies to f(x). Essentially, by retracing our steps and using the symmetry property (i.e., $k(x, x_i) = k(x_i, x)$), we see that the function can be expressed in this symmetrical form. The Mercer's kernel k can reproduce the function f(x), so when we take the inner product of the function f with the kernel, we can reconstruct the original function. Therefore, the inner product between two functions f and g is expressed in the bilinear form.

Now, I select the function g to be the kernel acting on x, where the dot notation indicates that $g(x_i)$ corresponds to the kernel $k(x, x_i)$. Essentially, $g(x_i)$ is the kernel evaluated at x and x_i. When I express this bilinear form and expand the function in terms of the kernel, leveraging the symmetric property, I retrieve the original function. This is why the Mercer kernel is known as a reproducing kernel, it has the ability to reproduce the function from which it was derived. This property is called the reproducing kernel property, and it is a fundamental concept.

(Refer Slide Time: 29:36)



This leads us to the notion of a Reproducing Kernel Hilbert Space (RKHS). Consider a function that can be expressed as a linear combination of kernels due to its membership in a Hilbert space. If I choose the function $g(x_i)$ to be the kernel $k(x, x_i)$, and then examine the inner product in bilinear form while invoking the symmetry property, it follows that taking the inner product of the function with the kernel reproduces the original function.

To summarize, for every x_i belonging to this Hilbert space, the kernel $k(x, x_i)$ as a function of x, which depends on x, also belongs to the space $\mathcal{F}$. This kernel, being an inner product kernel, satisfies the reproducing property: when you take the inner product of the function with the kernel, you obtain the function itself, which is incredibly useful.

This means that starting with a Mercer kernel, which is a reproducing kernel, allows us to retrieve the original function by selecting g as the kernel function within the inner product space $\mathcal{F}$. Therefore, the Mercer kernel is indeed a reproducing kernel.

Furthermore, a complete reproducing kernel space is a Reproducing Kernel Hilbert Space (RKHS). The completeness here is understood in the Cauchy sense, meaning any convergent sequence within this space, whether it is a function or vector, has its limit within the same space. In this context, we are dealing with functions, and thus the limiting function also belongs to the space. Since a Hilbert space is a complete inner product space, by endowing a reproducing kernel space with this completeness property, we establish it as a Reproducing Kernel Hilbert Space.

This understanding is crucial. When discussing RKHS, several key points must be kept in mind. First, one must grasp the concept of a Hilbert space in terms of convergence and completeness. Next, we need to consider how functions are expanded within the inner product space using kernels, particularly when considering the bilinear form of the inner product. By choosing one of the functions in the inner product to be the kernel, we can retrieve the original function, a property that is vital to many applications of RKHS.

In the earlier part of this lecture, I discussed Cauchy sequences in the context of series convergence and introduced the concept of a Hilbert space as a complete inner product space. Building on those foundational ideas, we then explored the notions of a reproducing kernel space and a Reproducing Kernel Hilbert Space. We’ll conclude here, and in the next session, we’ll delve into an important result: the representer theorem.