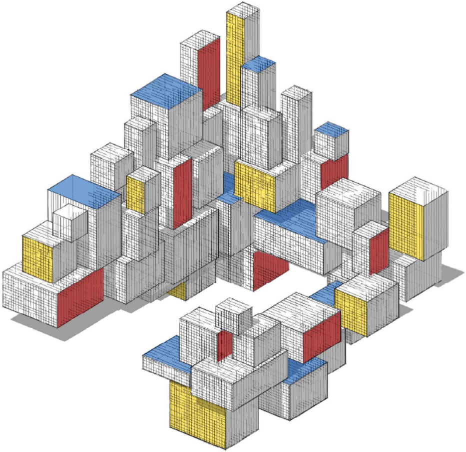
Reproducing Kernel Hilbert Spaces



Purpose

In this lecture we discuss:

- Feature maps
- Kernels
- Reproducing kernel Hilbert spaces
- Moore-Aronszajn theorem

Reproducing Kernel Hilbert Spaces

Suppose we have a feature space $\mathcal{X}$ and a Hilbert space $\mathcal{G}$ of prediction functions.

To each $\boldsymbol{x} \in \mathcal{X}$ we wish to associate a feature *function* $\kappa_{\boldsymbol{x}} \in \mathcal{G}$.

To evaluate the loss of a learner $g = g_\tau \in \mathcal{G}$, we do not need to explicitly construct g — rather, it is only required that we can evaluate g at all the feature vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of the training set τ.

A defining property of a Reproducing Kernel Hilbert Space $\mathcal{G}$ is that function evaluation at a point $\boldsymbol{x}$ can be performed by simply taking the inner product of g with feature function $\kappa_{\boldsymbol{x}}$; that is,

$$g(\boldsymbol{x}) = \langle g, \kappa_{\boldsymbol{x}} \rangle.$$

Reproducing Kernel

This property becomes particularly useful in light of the representer theorem, which states that the learner g itself can be represented as a *linear combination* of the set of feature functions $\{\kappa_{\boldsymbol{x}_i}, i = 1, \dots, n\}$.

Consequently, we can evaluate a learner g at the feature vectors $\{\boldsymbol{x}_i\}$ by taking linear combinations of $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) := \langle \kappa_{\boldsymbol{x}_i}, \kappa_{\boldsymbol{x}_j} \rangle$.

Collecting these inner products into a Gram matrix $\mathbf{K} = [\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j), i, j = 1, \dots, n]$, we will see that the feature vectors $\{\boldsymbol{x}_i\}$ only enter the loss minimization problem through $\mathbf{K}$.

Definition: Reproducing Kernel Hilbert Space

For a non-empty set $\mathcal{X}$, a Hilbert space $\mathcal{G}$ of functions $g : \mathcal{X} \to \mathbb{R}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{G}}$ is called a reproducing kernel Hilbert space (RKHS) with reproducing kernel $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ if:

1. for every $\boldsymbol{x} \in \mathcal{X}$, $\kappa_{\boldsymbol{x}} := \kappa(\boldsymbol{x}, \cdot)$ is in $\mathcal{G}$,
2. $\kappa(\boldsymbol{x}, \boldsymbol{x}) < \infty$ for all $\boldsymbol{x} \in \mathcal{X}$,
3. for every $\boldsymbol{x} \in \mathcal{X}$ and $g \in \mathcal{G}$, $g(\boldsymbol{x}) = \langle g, \kappa_{\boldsymbol{x}} \rangle_{\mathcal{G}}$.

The main (third) condition is known as the reproducing property.

The function $\kappa_{\boldsymbol{x}}$ is called the representer of evaluation.

Properties of an RKHS

The reproducing kernel of a Hilbert space of functions is *unique*.

A reproducing kernel is*symmetric* functions: $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \kappa(\boldsymbol{x}', \boldsymbol{x})$.

A reproducing kernel κ is a positive semidefinite function, meaning that for every $n \geqslant 1$ and every choice of $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$, it holds that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \, \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) \, \alpha_j \geqslant 0. \tag{1}$$

In other words, every Gram matrix $\mathbf{K}$ associated with κ is a positive semidefinite matrix; that is $\boldsymbol{\alpha}^\top \mathbf{K} \boldsymbol{\alpha} \geqslant 0$ for all $\boldsymbol{\alpha}$.

Alternative Characterization of an RKHS

Theorem: Continuous Evaluation Functionals Characterize a RKHS

An RKHS $\mathcal{G}$ on a set $\mathcal{X}$ is a Hilbert space in which every evaluation functional $\delta_{\boldsymbol{x}} : g \mapsto g(\boldsymbol{x})$ is bounded.

Conversely, a Hilbert space $\mathcal{G}$ of functions $\mathcal{X} \to \mathbb{R}$ for which every evaluation functional is bounded is an RKHS.

The fact that an RKHS has continuous evaluation functionals means that if two functions $g, h \in \mathcal{G}$ are "close" with respect to $\| \cdot \|_{\mathcal{G}}$, then their evaluations $g(\boldsymbol{x}), h(\boldsymbol{x})$ are close for every $\boldsymbol{x} \in \mathcal{X}$.

Formally, convergence in $\| \cdot \|_{\mathcal{G}}$ norm implies pointwise convergence for all $\boldsymbol{x} \in \mathcal{X}$.

Proof

Note that, since evaluation functionals $\delta_{\boldsymbol{x}}$ are linear operators, showing boundedness is equivalent to showing continuity.

Given an RKHS with reproducing kernel κ, suppose that we have a sequence $g_n \in \mathcal{G}$ converging to $g \in \mathcal{G}$, that is $\|g_n - g\|_{\mathcal{G}} \to 0$. We apply the Cauchy–Schwarz inequality and the reproducing property of κ to find that for every $\boldsymbol{x} \in \mathcal{X}$ and any n:

$$\begin{aligned} |\delta_{\boldsymbol{x}} g_n - \delta_{\boldsymbol{x}} g| = |g_n(\boldsymbol{x}) - g(\boldsymbol{x})| = |\langle g_n - g, \kappa_{\boldsymbol{x}} \rangle_{\mathcal{G}}| &\leqslant \|g_n - g\|_{\mathcal{G}} \, \|\kappa_{\boldsymbol{x}}\|_{\mathcal{G}} \\ = \|g_n - g\|_{\mathcal{G}} \sqrt{\langle \kappa_{\boldsymbol{x}}, \kappa_{\boldsymbol{x}} \rangle_{\mathcal{G}}} &= \|g_n - g\|_{\mathcal{G}} \sqrt{\kappa(\boldsymbol{x}, \boldsymbol{x})}. \end{aligned}$$

Noting that $\sqrt{\kappa(\boldsymbol{x}, \boldsymbol{x})} < \infty$ by definition for every $\boldsymbol{x} \in \mathcal{X}$, and that $\|g_n - g\|_{\mathcal{G}} \to 0$ as $n \to \infty$, we have shown continuity of $\delta_{\boldsymbol{x}}$, that is $|\delta_{\boldsymbol{x}} g_n - \delta_{\boldsymbol{x}} g| \to 0$ as $n \to \infty$ for every $\boldsymbol{x} \in \mathcal{X}$.

Proof

Conversely, suppose that evaluation functionals are bounded.

Theorem: Riesz Representation Theorem

Any bounded linear functional ϕ on a Hilbert space $\mathcal{H}$ can be represented as $\phi(h) = \langle h, g \rangle$, for some $g \in \mathcal{H}$.

Then from the Riesz representation theorem, there exists some $g_{\delta_{\boldsymbol{x}}} \in \mathcal{G}$ such that $\delta_{\boldsymbol{x}} g = \langle g, g_{\delta_{\boldsymbol{x}}} \rangle_{\mathcal{G}}$ for all $g \in \mathcal{G}$ — the *representer* of evaluation.

If we define $\kappa(\boldsymbol{x}, \boldsymbol{x}') = g_{\delta_{\boldsymbol{x}}}(\boldsymbol{x}')$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$, then $\kappa_{\boldsymbol{x}} := \kappa(\boldsymbol{x}, \cdot) = g_{\delta_{\boldsymbol{x}}}$ is an element of $\mathcal{G}$ for every $\boldsymbol{x} \in \mathcal{X}$ and $\langle g, \kappa_{\boldsymbol{x}} \rangle_{\mathcal{G}} = \delta_{\boldsymbol{x}} g = g(\boldsymbol{x})$, so that the reproducing property is verified. $\square$

Moore–Aronszajn

Any finite function $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ can serve as a reproducing kernel as long as it is finite, symmetric, and positive semidefinite.

The corresponding (unique!) RKHS $\mathcal{G}$ is the completion of the set of all functions of the form $\sum_{i=1}^{n} \alpha_i \, \kappa_{\boldsymbol{x}_i}$ where $\alpha_i \in \mathbb{R}$ for all $i = 1, \ldots, n$.

Theorem: Moore–Aronszajn

Given a non-empty set $\mathcal{X}$ and any finite symmetric positive semidefinite function $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, there exists an RKHS $\mathcal{G}$ of functions $g : \mathcal{X} \to \mathbb{R}$ with reproducing kernel κ. Moreover, $\mathcal{G}$ is unique.

Existence Proof (Sketch)

The idea is to construct a pre-RKHS $\mathcal{G}_0$ from the given function κ that has the essential structure and then to extend $\mathcal{G}_0$ to an RKHS $\mathcal{G}$.

In particular, define $\mathcal{G}_0$ as the set of finite linear combinations of functions $\kappa_{\boldsymbol{x}}, \boldsymbol{x} \in \mathcal{X}$:

$$\mathcal{G}_0 := \left\{ g = \sum_{i=1}^{n} \alpha_i \, \kappa_{\boldsymbol{x}_i} \;\middle|\; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X},\ \alpha_i \in \mathbb{R},\ n \in \mathbb{N} \right\}.$$

Define on $\mathcal{G}_0$ the following inner product:

$$\langle f, g \rangle_{\mathcal{G}_0} := \left\langle \sum_{i=1}^{n} \alpha_i \, \kappa_{\boldsymbol{x}_i}, \sum_{j=1}^{m} \beta_j \, \kappa_{\boldsymbol{x}'_j} \right\rangle_{\mathcal{G}_0} := \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \, \beta_j \, \kappa(\boldsymbol{x}_i, \boldsymbol{x}'_j).$$

Then $\mathcal{G}_0$ is an inner product space.

Existence Proof (Sketch)

In fact, $\mathcal{G}_0$ has the essential structure we require, namely that (i) evaluation functionals are bounded/continuous and (ii) Cauchy sequences in $\mathcal{G}_0$ that converge pointwise also converge in norm.

We then enlarge $\mathcal{G}_0$ to the set $\mathcal{G}$ of all functions $g : \mathcal{X} \to \mathbb{R}$ for which there exists a Cauchy sequence in $\mathcal{G}_0$ converging pointwise to g and define an inner product on $\mathcal{G}$ as the limit

$$\langle f, g \rangle_{\mathcal{G}} := \lim_{n \to \infty} \langle f_n, g_n \rangle_{\mathcal{G}_0},$$

where $f_n \to f$ and $g_n \to g$. To show that $\mathcal{G}$ is an RKHS it remains to be shown that (1) this inner product is well defined; (2) evaluation functionals remain bounded; and (3) the space $\mathcal{G}$ is complete.

Reproducing Kernels via Feature Mapping

In view of the Moore–Aronszajn theorem, to define an RKHS on a feature space $\mathcal{X}$ all all we need to specify is the reproducing kernel $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$; that is, specify a finite, symmetric, and positive semidefinite function.

The easiest way to do this is via a feature map $\boldsymbol{\phi} : \mathcal{X} \to \mathbb{R}^p$, by defining $\kappa(\boldsymbol{x}, \boldsymbol{x}') := \langle \boldsymbol{\phi}(\boldsymbol{x}), \boldsymbol{\phi}(\boldsymbol{x}') \rangle$, where $\langle \ , \ \rangle$ denotes the Euclidean inner product. This function is clearly finite and symmetric.

To verify that κ is positive semidefinite, let $\mathbf{\Phi}$ be the matrix with rows $\boldsymbol{\phi}(\boldsymbol{x}_1)^\top, \ldots, \boldsymbol{\phi}(\boldsymbol{x}_n)^\top$ and let $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]^\top \in \mathbb{R}^n$. Then,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \, \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) \, \alpha_j = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \, \boldsymbol{\phi}^\top(\boldsymbol{x}_i) \, \boldsymbol{\phi}(\boldsymbol{x}_j) \, \alpha_j = \boldsymbol{\alpha}^\top \mathbf{\Phi} \mathbf{\Phi}^\top \boldsymbol{\alpha} = \|\mathbf{\Phi}^\top \boldsymbol{\alpha}\|^2 \geqslant 0.$$

Example: Linear Kernel

Taking the identity feature map $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{x}$ on $\mathcal{X} = \mathbb{R}^p$, gives the linear kernel

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \langle \boldsymbol{x}, \boldsymbol{x}' \rangle = \boldsymbol{x}^\top \boldsymbol{x}'.$$

The RKHS of functions corresponding to the linear kernel is the space of linear functions on $\mathbb{R}^p$.

This space is isomorphic to $\mathbb{R}^p$ itself.

Nonuniqueness of Feature Maps for a Kernel

Does a given kernel function correspond uniquely to a feature map? The answer is no, as the next example shows.

Let $\mathcal{X} = \mathbb{R}$ and consider feature maps $\phi_1 : \mathcal{X} \to \mathbb{R}$ and $\boldsymbol{\phi}_2 : \mathcal{X} \to \mathbb{R}^2$, with $\phi_1(x) := x$ and $\boldsymbol{\phi}_2(x) := [x, x]^\top/\sqrt{2}$. Then

$$\kappa_{\phi_1}(x, x') = \langle \phi_1(x), \phi_1(x') \rangle = xx',$$

but also

$$\kappa_{\boldsymbol{\phi}_2}(x, x') = \langle \boldsymbol{\phi}_2(x), \boldsymbol{\phi}_2(x') \rangle = xx'.$$

Thus, we arrive at the same kernel function defined for the same underlying set $\mathcal{X}$ via two different feature maps.