Introduction of Reproducing Kernel Hilbert Space with An Application to Personalized Intervention System

Drafted by Haoda Fu, Ph.D.

Statistics for Data Scientists Series

October 14, 2018

Advanced Analytics and Data Sciences (AADS)
Eli Lilly and Company

Why

RKHS

RKHS theory is one of the most useful and elegant theories for mathematics and machine learning. It quantifies the model complexity by pointing out that the optimal solution from an infinity functional space is embedded into a finite dimensional subspace.

Why

RKHS

RKHS theory is one of the most useful and elegant theories for mathematics and machine learning. It quantifies the model complexity by pointing out that the optimal solution from an infinity functional space is embedded into a finite dimensional subspace.

Personalized Intervention and Digital Health

- **Context:** Digital technology will enable us to collect more individual patient data.
- **Decision:** The purpose of collecting these data is to generate actionable insights.
- **Reward:** The goal of these actionable insights is to maximize individual patient's outcomes.

Why

Can we leverage RKHS theory to enhance the solution for personalized intervention algorithms?

Introduction of RKHS

Logic review

Field

$\Downarrow$

Vector Spaces

$\Downarrow$

Banach Spaces

$\Downarrow$

Hilbert Spaces

$\Downarrow$

Reproducing Kernel Hilbert Spaces

Field: A field F is a structure $(F, +, \cdot, 0, 1)$

The following properties hold for a field:

1. $\forall \alpha, \beta \in F, \alpha + \beta = \beta + \alpha$ (commutativity of addition)
2. $\forall \alpha, \beta, \gamma \in F, \alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$ (commutativity of addition)
3. $\forall \alpha, \beta \in F, \alpha \cdot \beta = \beta \cdot \alpha$ (commutativity of multiplication)
4. $\forall \alpha, \beta, \gamma \in F, \alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma$ (commutativity of multiplication)
5. $\forall \alpha, \beta, \gamma \in F, \alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$ (distributivity)
6. $\exists 0 \in F, s.t. \forall \alpha \in F, 0 + \alpha = \alpha$ (additive identity)
7. $\forall \alpha \in F, \exists -\alpha \in F, s.t. \alpha + (-\alpha) = 0$ (additive inverse)
8. $\exists 1 \in F, s.t. \forall \alpha \in F, 1 \cdot \alpha = \alpha$ (multiplication identity)
9. $\forall \alpha \in F, \alpha \neq 0, \exists \alpha^{-1} \in F, s.t.\ \alpha \cdot \alpha^{-1} = 1$ (multiplicative inverse)

Example: $\mathbb{Q}$.

Vector Spaces: $+ : V \times V \mapsto V, \cdot : F \times V \mapsto V$

The following properties hold for a vector space:

1. $\forall f, g \in V, f + g = g + f$ (commutativity of addition)
2. $\forall f, g, h \in V, f + (g + h) = (f + g) + h$ (commutativity of addition)
3. $\exists 0 \in V, s.t. \forall f \in V, 0 + f = f$ (additive identity)
4. $\forall f \in V, \exists -f \in V, s.t.\ f + (-f) = 0$ (additive inverse)
5. $\forall f \in V, \forall \alpha, \beta \in F, \alpha \cdot (\beta \cdot f) = (\alpha \cdot \beta) \cdot f$ (associativity of multiplication)
6. $\forall f \in V, \forall \alpha, \beta \in F, (\alpha + \beta) \cdot f = \alpha \cdot f + \beta \cdot f$ (scalar distributivity)
7. $\forall f, g \in V, \forall \alpha \in F, \alpha \cdot (f + g) = \alpha \cdot f + \alpha \cdot g$ (vector distributivity)
8. $\forall f \in V, \exists 1 \in F, 1 \cdot f = f$ (multiplicative identity)

Example: $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^{\mathbb{R}}, \mathbb{R}^{\mathcal{X}}, C(\mathcal{X})$.

Banach Spaces

Definition (Banach Space)

A complete normed vector space is called a Banach Space.

Normed Spaces

A vector space V is called a normed space if $\forall f \in V$, $\exists \|f\| \in \mathbb{R}$, called the norm of u, such that

- $\|f\| > 0$ if $f \neq 0$, and $\|f\| = 0$ implies $f = 0$
- $\|f + g\| \leq \|f\| + \|g\|$
- $\|\alpha f\| = |\alpha| \|f\|$, where $\alpha \in \mathbb{R}$

Complete Vector Spaces

A vector space V is called complete if every Cauchy sequence in V convergent.

Example: Euclidean Spaces in $\mathbb{R}^n$, $L^p(V, \mathcal{F}, \mu)$ where $p \in [1, +\infty) - \{2\}$.

Hilbert Spaces

Definition (Hilbert Spaces)

A Hilbert space is a Banach space equipped with a dot product $\langle \cdot, \cdot \rangle$. The dot product operation satisfies the following properties, $\forall f, g, h \in \mathcal{H}$ and $\alpha \in F$

- Commutative: $\langle f, g \rangle = \langle g, f \rangle$ and $\langle f, g \rangle \in F$.
- Associative: $\langle \alpha \cdot f, g \rangle = a \cdot \langle f, g \rangle$
- Distributive: $\langle f, g + h \rangle = \langle f, g \rangle + \langle f, h \rangle$

The norm of f is defined as $\|f\| = \sqrt{\langle f, f \rangle}$.

Example: $L^2(a, b)$.

Reproducing Kernel Hilbert Spaces(RKHS)

Definition (RKHS)

A Hilbert space of real-valued functions $\mathcal{H}$ is an RKHS if every evaluation functional is bounded and continuous.

Evaluation functional

Let $\mathcal{H}$ be a Hilbert space of real-valued functions from $\mathcal{X}$ to $\mathbb{R}$, i.e. $\mathbb{R}^{\mathcal{X}}$ where $\mathcal{X}$ is an arbitrary set. For a fixed $x \in \mathcal{X}$, the evaluation functional $\mathcal{L}_x : \mathcal{H} \to \mathbb{R}$ is defined as,

$$\mathcal{L}_x(f) \triangleq f(x),$$

where $f \in \mathcal{H}$. Evaluation functionals are linear since,

$$\mathcal{L}_x(\alpha \cdot f + \beta \cdot g) = \alpha \cdot f(x) + \beta \cdot g(x) = \alpha \cdot \mathcal{L}_x f + \beta \cdot \mathcal{L}_x g.$$

Reproducing Kernel

Theorem (Riesz representation theorem)

If T be a bounded linear functional on a Hilbert space $\mathcal{H}$, then there exists a unique $g \in \mathcal{H}$ such that $T(f) = \langle g, f \rangle, \forall f \in \mathcal{H}$. The element g is called the representer of T.

Definition (Reproducing kernel)

Let $K_x \in \mathcal{H}$ be a representer, and $\langle K_x, f \rangle = f(x), \forall f \in \mathcal{H}$. The symmetric bivariate function $K(x, y) = \langle K_x, K_y \rangle = K_x(y) = K_y(x)$ is called the reproducing kernel (RK) of the space $\mathcal{H}$ and it has the reproducing property $\langle K(x, \cdot), f(\cdot) \rangle = f(x)$.

Theorem (Moore-Aronszajn theorem)

for every positive definite function $K\langle \cdot, \cdot \rangle$ on $\mathcal{X} \times \mathcal{X}$, there exists a unique RKHS and vice versa.

Why RKHS Matters - From Infinity to Finite

A general class of regularization problems has the form,

$$\underset{f \in \mathcal{H}}{\text{minimize}} \quad \left[\sum_{i=1}^{n} L\{y_i, f(x_i)\} + \lambda J(f) \right],$$

where $L(\cdot, \cdot)$ is a loss function, $J(\cdot)$ is a penalty function, and $\mathcal{H}$ is a space of functions on which $J(\cdot)$ is defined. In particular, we have $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, with the null space $\mathcal{H}_0$ consisting of, for example, low degree polynomials in that elements do not get penalized. The penalty becomes $J(f) = \|P_1 f\|^2$, where P_1 is the orthogonal projection of f onto $\mathcal{H}_1$. If $\mathcal{H}_1$ is a RKHS and $L(\cdot, \cdot)$ is continuous and convex, the solution has the form,

$$f(x) = \sum_{j=1}^{m} \beta_j h_j(x) + \sum_{i=1}^{n} \alpha_i K(x, x_i),$$

where the first term represents an expansion in $\mathcal{H}_0$.

Existence of Minimizer

A functional $F(f)$ on a Hilbert space $\mathcal{H}$ is said to be **convex** if $\forall f, g \in \mathcal{H}, F\{\alpha \cdot f + (1-\alpha) \cdot g\} \leq \alpha F(f) + (1-\alpha)F(g), \forall \alpha \in (0,1)$, and the convexity is strict if the equality holds only for $f = g$.

Theorem (Existence)

Suppose $L(f)$ is a continuous and convex functional in a Hilbert space $\mathcal{H}$ and $J(f)$ is a square (semi) norm in $\mathcal{H}$ with a null space $\mathcal{H}_0$ of finite dimension. If $L(f)$ has a unique minimizer in $\mathcal{H}_0$, then $L(f) + \lambda J(f)$ has a minimizer in $\mathcal{H}$.

Key to proof: Since $L(f) + \lambda J(f)$ must has a minimizer in a closed rectangle set,

$$R_{\rho,\gamma} \quad = \quad \{f : f \in \mathcal{H}, J(f) \leq \rho, \|f - f_0\|_0 \leq \gamma\}, \forall \rho, \gamma,$$

where f_0 is the unique minimizer in $\mathcal{H}_0$, and $\|\cdot\|_0$ is the norm on $\mathcal{H}_0$. If $L(f) + \lambda J(f)$ doesn't have a minimizer in $\mathcal{H}$, the minimizer must be on the boundary of $R_{\rho,\gamma}$, by convexity, which cannot be true $\forall \rho, \gamma$.

The Representer Theorem

Theorem (The representer theorem)

Suppose $L(f)$ is a continuous and convex functional in a Hilbert space $\mathcal{H}$ and $J(f)$ is a square (semi) norm in $\mathcal{H}$ with a null space $\mathcal{H}_0$ of finite dimension. If $L(f)$ has a unique minimizer in $\mathcal{H}_0$, then $L(f) + \lambda J(f)$ has a minimizer in $\mathcal{H}$, and

$$f(x) = \sum_{j=1}^{m} \beta_j h_j(x) + \sum_{i=1}^{n} \alpha_i K(x, x_i).$$

Key to proof: $f = f_0 + f_K + f_\eta$ where $f_0 \in \mathcal{H}_0$, $f_K \in \mathcal{H}_K \triangleq \text{span}\{K(x_1, \cdot), i = 1, c \ldots, n\}$ and $f_\eta \in \mathcal{H} \ominus (\mathcal{H}_K \oplus \mathcal{H}_0)$. Then, we have $0 = \langle K(x_i, \cdot), f_\eta(\cdot) \rangle = f_\eta(x_i)$. So,
$L\{(f_0 + f_K + f_\eta)(x_i)\} = L\{(f_0 + f_K)(x_i)\}$ and
$J(f) = J(f_K + f_\eta) = J(f_K) + J(f_\eta) \geq J(f_K)$.

Personalized Intervention