REPRODUCING KERNEL HILBERT SPACES

DR. MATTHIEU R BLOCH

Monday, October 25, 2021

LOGISTICS

Drop date: October 30, 2021

My office hours tomorrow

- Tuesdays 8am-9am on BlueJeans (https://bluejeans.com/205357142)
- Come prepared!

Midterm 2:

- Moved to **Monday November 8, 2021** (gives you weekend to prepare)
- Coverage: everything since Midterm 1 (dont' forget the fundamentals though), emphasis on **regression**

WHAT'S ON THE AGENDA FOR TODAY?

Last time:

- Functional on Hilbert spaces

Today:

- Reproducing Kernel Hilbert Spaces

Reading: Romberg, lecture notes 10/11

LAST TIME: RIESZ REPRESENTATION THEOREM

Theorem (Riesz representation theorem)

Let $F : \mathcal{F} \to \mathbb{R}$ be a *continuous* linear functional on a (possible infinite dimensional) separable Hilbert space $\mathcal{F}$.

Then there exists $c \in \mathcal{F}$ such that $F(x) = \langle x, c \rangle$ for every $x \in \mathcal{F}$

Proposition. If $\{\psi_n\}_{n \geq 1}$ is an orthobasis for $\mathcal{H}$, then we can construct c above as

$$c \triangleq \sum_{n=1}^{\infty} F(\psi_n)\psi_n$$

REPRODUCING KERNEL HILBERT SPACES

Definition. (Reproducing Kernel Hilbert Spaces)

An RKHS is a Hilbert space $\mathcal{H}$ of real-valued functions $f : \mathbb{R}^d \to \mathbb{R}$ in which the sampling operation $\mathcal{S}_{\boldsymbol{\tau}} : \mathcal{H} \to \mathbb{R} : f \mapsto f(\boldsymbol{\tau})$ is continuous for every $\boldsymbol{\tau} \in \mathbb{R}^d$.

In other words, for each $\boldsymbol{\tau} \in \mathbb{R}^d$, there exists $k_{\boldsymbol{\tau}} \in \mathcal{H}$ s.t.

$$f(\boldsymbol{\tau}) = \langle f, k_{\boldsymbol{\tau}} \rangle_{\mathcal{H}} \text{ for all } f \in \mathcal{H}$$

Definition. (Kernel)

The kernel of an RKHS is

$$k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} : (\mathbf{t}, \boldsymbol{\tau}) \mapsto k_{\boldsymbol{\tau}}(\mathbf{t})$$

where $k_{\boldsymbol{\tau}}$ is the element of $\mathcal{H}$ that defines the sampling at $\boldsymbol{\tau}$.

Proposition.

A (separable) Hilbert space with orthobasis $\{\psi_n\}_{n\geq 1}$ is an RKHS with kernel $k(\mathbf{t}, \boldsymbol{\tau}) = \sum_{n=1}^{\infty} \psi_n(\boldsymbol{\tau})\psi_n(\mathbf{t})$ iff $\forall \boldsymbol{\tau} \in \mathbb{R}^d \sum_{n=1}^{\infty} |\psi_n(\tau)|^2 < \infty$

RKHS AND NON ORTHOGONAL BASIS

If $\{\phi_n\}_{n\geq1}$ is a Riesz basis for $\mathcal{H}$, we know that every $x \in \mathcal{H}$ can be written

$$x = \sum_{n\geq1} \alpha_n \phi_n \text{ with } \alpha_n \triangleq \langle x, \widetilde{\phi}_n\rangle$$

where $\{\widetilde{\phi}_n\}_{n\geq1}$ is the dual basis.

Proposition.

A (separable) Hilbert space with Riesz basis $\{\phi_n\}_{n\geq1}$ is an RKHS with kernel

$$k(\mathbf{t}, \boldsymbol{\tau}) = \sum_{n=1}^{\infty} \phi_n(\boldsymbol{\tau})\widetilde{\phi}_n(\mathbf{t})$$

iff $\forall \boldsymbol{\tau} \in \mathbb{R}^d \sum_{n=1}^{\infty} |\phi_n(\tau)|^2 < \infty$

EXAMPLES

Finite dimensional Hilbert space

Space of Lth order polynomial splines on the real line

Remark

- RKHS are more easily characterized by their kernel
- Often, we try to avoid an explicit description of the the elements in the space

KERNEL REGRESSION

Regression problem: given n pairs $(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$, solve

$$\min_{f \in \mathcal{F}} \sum_{i=1}^{n} |y_i - f(\mathbf{x}_i)|^2 + \lambda \|f\|_{\mathcal{F}}^2$$

If we restrict $\mathcal{F}$ to be an RKHS, the problem becomes

$$\min_{f \in \mathcal{F}} \sum_{i=1}^{n} \left| y_i - \langle f, x_i \rangle_{\mathcal{F}} \right|^2 + \lambda \|f\|_{\mathcal{F}}^2$$

where $x_i \triangleq k_{\mathbf{x}_i}$ provides the mapping between $\mathbb{R}^d$ and $\mathcal{F}$

$$x_i : \mathbf{R}^d \to \mathbb{R} : \mathbf{t} \mapsto k_{\mathbf{x}_i}(\mathbf{t}) = k(\mathbf{x}_i, \mathbf{t})$$

The solution is given by

$$\widehat{f} = \sum_{i=1}^{n} \widehat{\alpha}_i x_i \text{ with } \widehat{\boldsymbol{\alpha}} \triangleq (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}$$

and $\mathbf{K} \triangleq [K_{i,j}]_{1 \le i,j \le n}$ with $K_{i,j} = \langle x_i, x_j \rangle$

KERNEL REGRESSION

Kernel magic

1. $K_{ij} = \langle x_i, x_j \rangle = \left\langle k_{\mathbf{x}_i}, k_{\mathbf{x}_j} \right\rangle = k_{\mathbf{x}_i}(\mathbf{x}_j) = k(\mathbf{x}_i, \mathbf{x}_j)$
2. $\widehat{f}(\mathbf{x}) = \left\langle \widehat{f}, k_{\mathbf{x}} \right\rangle = \sum_{i=1}^{n} \widehat{\alpha_i} k(\mathbf{x}_i, \mathbf{x})$

Remarks

- We solved an infinite dimensional problem using an $n \times n$ system of equations and linear algebra
- Our solution and the evaluation only depend on the *kernel*; we never need to work directly in $\mathcal{F}$

Question: can we skip $\mathcal{F}$ entirely? how do we find "good" kernels?

ARONSZJAN'S THEOREM

Definition. (Inner product kernel)

An *inner product kernel* is a mapping $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ for which there exists a Hilbert space $\mathcal{H}$ and a mapping $\mathbf{\Phi} : \mathbb{R}^d \to \mathcal{H}$ such that

$$\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^d \quad k(\mathbf{u}, \mathbf{v}) = \langle \mathbf{\Phi}(\mathbf{u}), \mathbf{\Phi}(\mathbf{v}) \rangle_{\mathcal{H}}$$

Definition. (Positive semidefinite kernel)

A function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a *positive semidefinite* kernel if

- k is symmetric, i.e., $k(\mathbf{u}, \mathbf{v}) = k(\mathbf{v}, \mathbf{u})$
- for all $\{\mathbf{x}_i\}_{i=1}^N$, the *Gram matrix* $\mathbf{K}$ is positive semidefinite, i.e.,

$$\mathbf{x}^\intercal \mathbf{K} \mathbf{x} \geq 0 \text{ with } \mathbf{K} = [K_{i,j}] \text{ and } K_{i,j} \triangleq k(\mathbf{x}_i, \mathbf{x}_j)$$

Theorem.

A function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is an inner product kernel if and only if k is a positive semidefinite kernel.

EXAMPLES

Example. Regression using linear and quadratic functions in $\mathbb{R}^d$

Example. Regression using Radial Basis Functions

Examples of kernels

- Homogeneous polynomial kernel: $k(\mathbf{u}, \mathbf{v}) = (\mathbf{u}^\intercal \mathbf{v})^m$ with $m \in \mathbb{N}^*$
- Inhomogenous polynomial kernel: $k(\mathbf{u}, \mathbf{v}) = (\mathbf{u}^\intercal \mathbf{v} + c)^m$ with $c > 0, m \in \mathbb{N}^*$
- Radial basis function (RBF) kernel: $k(\mathbf{u}, \mathbf{v}) = \exp\left(-\frac{\| \mathbf{u} - \mathbf{v} \|^2}{2\sigma^2}\right)$ with $\sigma^2 > 0$