Regularization in Reproducing Kernel Banach Spaces

Guohui Song

School of Mathematical and Statistical Sciences
Arizona State University

Comp Math Seminar, September 16, 2010

Joint work with
Dr. Fred Hickernell (Illinois Institute of Technology)
Dr. Haizhang Zhang (Sun Yat-Sen University)

Outline

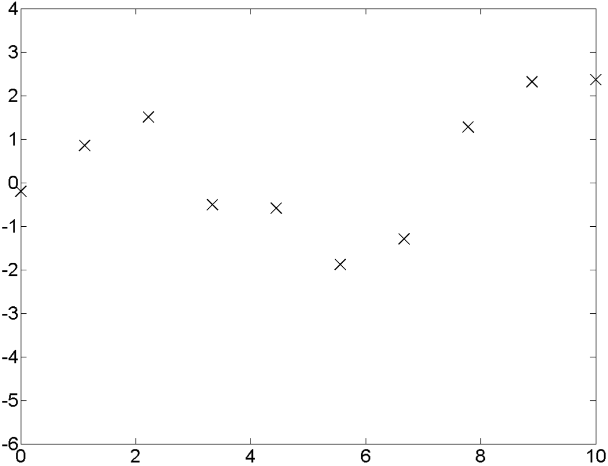
1. Scattered Data Approximation
2. Reproducing Kernel Hilbert Spaces
3. Reproducing Kernel Banach Spaces

Setting:

- Given data $\{(\boldsymbol{x}_j, y_j) : j = 1, 2, \ldots, n\}$ in $\mathbb{R}^d \times \mathbb{R}$.
- Find a function Pf which is a “good fit” to the given data.

Question 1

What is a “good fit”?



scattered data points

A Regularization Approach

- We want to control
  - the closeness to the given data
  - the complexity of the function
- A regularization approach:
  - Target function: $L(f, \mathbf{y}, \mathcal{H}) := \sum_{j=1}^{n} (f(\mathbf{x}_j) - y_j)^2 + \lambda \|f\|_{\mathcal{H}}^2$
  - $Pf := \arg\min_{f \in \mathcal{H}} L(f, \mathbf{y}, \mathcal{H})$.
  - Some fancy names: penalized least square, ridge regression, smoothing spline

Question 2

What is the hypothesis space $\mathcal{H}$?

Reproducing Kernel Hilbert Spaces (RKHS)

- We need
  - $\mathcal{H}$ is a Hilbert space
  - $\|f_n - f\|_{\mathcal{H}} \to 0 \Longrightarrow f_n(x) - f(x) \to 0$ for all $x \in \mathcal{X}$
- RKHS: a Hilbert space $\mathcal{H}$ on which the point evaluation functional is continuous.
  - $|f(x)| \leq M_x \|f\|_{\mathcal{H}}$ for all $x \in \mathcal{X}$

Question 3

Where is the "kernel"?

Kernel

- Suppose $\mathcal{X}$ is a subset of $\mathbb{R}^d$.
- K is a real-valued function on $\mathcal{X} \times \mathcal{X}$:

$$K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$$

- K is a kernel if for any positive integer m and $X := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\} \subset \mathcal{X}$, the kernel gram matrix

$$\mathsf{K}_m := [K(\boldsymbol{x}_j, \boldsymbol{x}_l) : 1 \leq j, l \leq m]$$

is symmetric and positive semi-definite.

Connections Between RKHS and Kernels

- [Aronszajn, 1950] There is a bijective mapping from the RKHS to the set of kernels such that
  - ▷ $K(\cdot, \boldsymbol{x}) \in \mathcal{H}$ for any $\boldsymbol{x} \in \mathcal{X}$,
  - ▷ $f(\boldsymbol{x}) = (f(\cdot), K(\cdot, \boldsymbol{x}))_{\mathcal{H}}$ for any $f \in \mathcal{H}$.

- Some properties of RKHS $\mathcal{H}_K$ and the kernel K
  - ▷ $\mathcal{H}_0 := \text{span}\{K(\cdot, \boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$ is dense in $\mathcal{H}_K$.
  - ▷ For any $f = \sum_{j=1}^{m} c_j K(\cdot, \boldsymbol{x}_j) \in \mathcal{H}_0$,

$$\|f\|_{\mathcal{H}_K} = \|\mathsf{K}_m^{1/2} \boldsymbol{c}\|_2.$$

Some Examples of RKHS and Kernels

- Sobolev space $H^2(\mathbb{R})$: $K(s,t) = \frac{\sqrt{3}}{3} e^{-\frac{\sqrt{3}}{2}|s-t|} \sin\left(\frac{|s-t|}{2} + \frac{\pi}{6}\right)$.
- C^0 Matérn kernel: $K(s,t) = \mathrm{e}^{-|s-t|}$
- Gaussian kernel: $K(s,t) = \mathrm{e}^{-w(s-t)^2},\ w > 0$.
- Sinc kernel: $K(s,t) = \operatorname{sinc}(s-t)$.
- Polynomial kernels: $K(s,t) = (st)^d,\ d = 1, 2, \ldots$.
- $L^2(\mathbb{R})$ is NOT a RKHS.

Regularization in the RKHS $\mathcal{H}_K$

- [Kimeldorf and Wahba, 1971] Representer Theorem :
  - Target function: $L(f, \boldsymbol{y}, \mathcal{H}_K) = \sum_{j=1}^{n} (f(\boldsymbol{x}_j) - y_j)^2 + \lambda \|f\|_{\mathcal{H}_K}^2$.
  - Let $\mathcal{S}_n := \text{span}\{K(\cdot, \boldsymbol{x}_j) : j = 1, 2, \ldots, n\}$.
  - The optimization problem reduces to finite-dimensional:

$$\min_{f \in \mathcal{H}_K} L(f, \boldsymbol{y}, \mathcal{H}_K) = \min_{f \in \mathcal{S}_n} L(f, \boldsymbol{y}, \mathcal{H}_K)$$

  - The minimizer is explicitly given:

$$Pf = \sum_{j=1}^{n} \alpha_j K(\cdot, \boldsymbol{x}_j), \qquad \text{where } \boldsymbol{\alpha} = (\mathsf{K}_n + \lambda \mathsf{I}_n)^{-1} \boldsymbol{y}.$$

Reproducing Kernel Banach Spaces

- We try to construct
  - a Banach space $\mathcal{B}$
  - point evaluation functional $\delta_{\boldsymbol{x}}$ is continuous on $\mathcal{B}$
- A specific construction
  - Let $\mathcal{B}_0 := \operatorname{span}\{K(\cdot, \boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$.
  - For any $f = \sum_{j=1}^m c_j K(\cdot, \boldsymbol{x}_j) \in \mathcal{B}_0$, define

$$\|f\|_{\mathcal{B}} := \|\boldsymbol{c}\|_1.$$

  - $\delta_{\boldsymbol{x}}$ is continuous on $\mathcal{B}_0$ if $K(\cdot, \cdot)$ is uniformly bounded.
  - Let $\mathcal{B}$ be the Banach completion of $\mathcal{B}_0$ with the norm $\|\cdot\|_{\mathcal{B}}$.

Some Properties of RKBS

- [Song2010+] Point evaluation functional is continuous on $\mathcal{B}$ if and only if

$$\sum_{j=1}^{\infty} \alpha_j K(\cdot, \boldsymbol{x}_j) = 0 \Longrightarrow \boldsymbol{\alpha} = 0.$$

- [Song2010+] Reproducing property still holds.
  - ▷ Define a bilinear form $< \cdot, \cdot >$ on $\mathcal{B}_0 \times \mathcal{B}_0$ such that

$$< \sum_{j=1}^{m} \alpha_j K(\cdot, \boldsymbol{x}_j), \sum_{j=1}^{m} \beta_j K(\cdot, \boldsymbol{x}_j) >= \boldsymbol{\alpha}^T \mathsf{K}_m \boldsymbol{\beta}$$

  - ▷ The bilinear form $< \cdot, \cdot >$ can be extended to $\mathcal{B} \times \mathcal{B}$ such that

$$< f, K(\cdot, \boldsymbol{x}) >= f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathcal{X}, f \in \mathcal{B}.$$

Regularization in RKBS

- Target function: $L(f, \boldsymbol{y}, \mathcal{B}) = \sum_{j=1}^{n} \left(f(\boldsymbol{x}_j) - y_j\right)^2 + \lambda \|f\|_{\mathcal{B}}$.
- Recall $\mathcal{S}_n = \operatorname{span}\{K(\cdot, \boldsymbol{x}_j) : j = 1, 2, \ldots, n\}$.
- Does the optimization problem reduce to finite-dimensional?

$$\min_{f \in \mathcal{B}} L(f, \boldsymbol{y}, \mathcal{B}) \overset{??}{=} \min_{f \in \mathcal{S}_n} L(f, \boldsymbol{y}, \mathcal{B})$$

- If it can reduce to finite-dimensional, how to find the minimizer $Pf = \sum_{j=1}^{n} \alpha_j K(\cdot, \boldsymbol{x}_j)$?

Regularization and Interpolation

- Define the interpolation space
  $\mathcal{I}_n(\boldsymbol{y}) = \{f \in \mathcal{B} : f(\boldsymbol{x}_j) = y_j,\ j = 1, 2, \ldots, n\}$.

- [Song2010+] The following two statements are equivalent.
  - $\min_{f \in \mathcal{B}} L(f, \boldsymbol{y}, \mathcal{B}) = \min_{f \in \mathcal{S}_n} L(f, \boldsymbol{y}, \mathcal{B}),$ for all $\boldsymbol{y} \in \mathbb{R}^n$.
  - $\min_{f \in \mathcal{I}_n(\boldsymbol{y})} \|f\|_{\mathcal{B}} = \min_{f \in \mathcal{I}_n(\boldsymbol{y}) \cap \mathcal{S}_n} \|f\|_{\mathcal{B}},$ for all $\boldsymbol{y} \in \mathbb{R}^n$.

- Note that $\mathcal{I}_n(\boldsymbol{y}) \cap \mathcal{S}_n$ has only one element when K_n is invertible.

- We only need to show that the minimal norm interpolation problem admits a minimizer in the finite-dimensional space $\mathcal{S}_n$.

Representer Theorem in RKBS

- Let $\boldsymbol{k}(\boldsymbol{x}) := (K(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, K(\boldsymbol{x}, \boldsymbol{x}_n))^T$.
- [Song2010+] Minimal norm interpolation

$$\min_{f \in \mathcal{I}_n(\boldsymbol{y})} \|f\|_{\mathcal{B}} = \min_{f \in \mathcal{I}_n(\boldsymbol{y}) \cap S_n} \|f\|_{\mathcal{B}}, \qquad \text{for all } \boldsymbol{y} \in \mathbb{R}^n$$

$$\Longleftrightarrow \quad \left\|K_n^{-1}\boldsymbol{k}(\boldsymbol{x})\right\|_1 \leq 1, \qquad \text{for all } \boldsymbol{x} \in \mathcal{X}.$$

- [Song2010+] Regularization

$$\min_{f \in \mathcal{B}} L(f, \boldsymbol{y}, \mathcal{B}) = \min_{f \in \mathcal{S}_n} L(f, \boldsymbol{y}, \mathcal{B}), \qquad \text{for all } \boldsymbol{y} \in \mathbb{R}^n$$

$$\Longleftrightarrow \quad \left\|K_n^{-1}\boldsymbol{k}(\boldsymbol{x})\right\|_1 \leq 1, \qquad \text{for all } \boldsymbol{x} \in \mathcal{X}.$$

Some Examples

- The condition $\left\|K_n^{-1}\boldsymbol{k}(\boldsymbol{x})\right\|_1 \leq 1$ is not easy to check.
- We have only been able to find two kernels satisfying it so far.
  - ▷ $K(s,t) = \min\{s,t\} - st, \quad s,t \in [0,1]$
  - ▷ $K(s,t) = \mathrm{e}^{-|s-t|}, \quad s,t \in \mathbb{R}$
- Counter examples that does not satisfy this condition
  - ▷ Gaussian kernels: $K(s,t) = \mathrm{e}^{-(s-t)^2}, \quad s,t \in \mathbb{R}$
  - ▷ Sinc Kernel: $K(s,t) = \operatorname{sinc}(s-t), \quad s,t \in \mathbb{R}$

How to find the minimizer?

- $\min_{f \in \mathcal{S}_n} L(f, \boldsymbol{y}, \mathcal{B}) = \min \left\{ \sum_{j=1}^{n} (f(\boldsymbol{x}_j) - y_j)^2 + \lambda \|\boldsymbol{c}\|_1 : f = \sum_{j=1}^{n} c_j K(\cdot, \boldsymbol{x}_j) \right\}$
- We do not have a closed form of the minimizer.
- Standard optimization methods may do, but we still need efficient methods especially for large size of data.

Thank you !