Lecture16

Tuesday, November 6, 2018 10:43 AM

HYPOTHESIS SPACE I

* **Goal**: to show that the assumption $\mathcal{H}$ is a compact subset of $C(X)$ is realizable

---

REPRODUCING KERNEL HILBERT SPACES (RKHS)

* We are interested in hypothesis spaces associated with a RKHS

- Assume X is a metric space and let $K : X \times X \to \mathbb{R}$ be

+ symmetric : $K(x, x') = K(x', x)$

+ positive-semidefinite : the Gramian matrix

$$\begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_n) \\ \vdots & & \\ K(x_n, x_1) & \cdots & K(x_n, x_n) \end{bmatrix} \geq 0$$

$\forall\, n \in \mathbb{N}$; and any $\{x_1, \dots, x_n\}$

- from semi-positiveness we can define

$$C_K := \sup_{x \in X} \sqrt{K(x,x)}$$

Ex 18.1:

$$C_K = \sup_{x, x' \in X} \sqrt{K(x,x')}$$

* $\forall\, x \in X$, we define

$$K_x(\cdot) : X \to \mathbb{R}$$
$$x' \mapsto K_x(x') := K(x, x')$$

* We define the space

$$\boxed{\mathring{\mathcal{H}}_K := \operatorname{span}\{ K_x : x \in X \}}$$

equipped with the following inner product:

$$f = \sum_{i=1}^{I} f_i K_{x_i} \in \mathring{\mathcal{H}}_K$$

$$g = \sum_{j=1}^{J} g_j K_{x_j} \in \mathring{\mathcal{H}}_K$$

$$(g, f)_{\mathcal{H}_K} := \sum_{i,j}^{I\,J} f_i\, g_j\, K(x_i, x_j)$$

<u>Thm</u> 18.1 (RKHS):

Let $\mathcal{H}_K$ be a Hilbert space with the following properties:

1) $K_x \in \mathcal{H}_K$ for $x \in X$

2) $\mathring{\mathcal{H}}_K$ is dense in $\mathcal{H}_K$

3) Reproducing property: for any $f \in \mathcal{H}_K$ and $x \in X$, we have

$$\left(K_x(x'), f(x') \right)_{\mathcal{H}_K} = f(x)$$

Then $\mathcal{H}_K$ is UNIQUE. Moreover

$$\mathcal{H}_K \subset C(X)$$

$$\mathcal{H}_K \hookrightarrow C(X) \text{ is bounded}$$

<u>Remark</u>: the corresponding between a kernel and its RKHS is one-to-one.

<u>Hypothesis space associated with a RKHS.</u>

<u>Def</u>: (Mercer)

$K : X \times X \to \mathbb{R}$ is a Mercer kernel if K is CONTINUOUS and symmetric positive semi-definite.

Prop 18.1: Suppose K is a Mercer kernel on a compact metric space X, and $\mathcal{H}_K$ is the associated RKHS. For any $R > 0$ the ball $B(R) := \{ f \in \mathcal{H}_K : \|f\|_{\mathcal{H}_K} \leq R \}$ is a CLOSED subset of $C(X)$.

- in order to prove this claim we need the following result.

Lem 18.1: (Weak compactness of closed balls in Hilbert spaces) If B is a closed ball in a Hilbert space $\mathcal{H}$, it is weakly compact. That is, every sequence $\{f_n\}_{n \in \mathbb{N}} \subset B$ has a weakly convergent subsequence $\{f_{n_k}\}_{k \in \mathbb{N}}$ i.e. there exists $f^* \in B$, such that

$$\lim_{k\to\infty} (f_{n_k}, g)_{\mathcal{H}} = (f^*, g)_{\mathcal{H}}$$

$$\forall g \in \mathcal{H}$$

<u>Proof of Prop 18.1</u>:

- From thm 18.1, it is sufficient to show that $B(R)$ is closed in $C(X)$.
- Suppose $\{f_n\}_{n\in\mathbb{N}} \subset B(R)$ converges to $f^* \in C(X)$, i.e.,

$$\lim_{n\to\infty} f_n(x) = f^*(x) \quad (*)$$

$\Rightarrow$ goal is to show that $f^* \in B(R)$

- By Lem 18.1 $B(R)$ is weakly compact $\rightarrow$ there exists a sub-sequence f_{n_k} converges weakly to $\hat{f} \in B(R)$, i.e.,

$$\lim_{k\to\infty} (f_{n_k}, g)_{\mathcal{H}_K} = (\hat{f}, g)_{\mathcal{H}_K} \quad \forall g \in \mathcal{H}_K$$

- Now take $g = K_x$

$$\lim (f_{n_i}, K_x) = (\hat{f}, K_x)$$

$$\lim_{i\to\infty} (f_{n_i}, K_x) = (f, K_x)$$

$$\| \qquad\qquad \| \text{ RKHS}$$

$$\lim_{i\to\infty} f_{n_i}(x) = \hat{f}(x)$$

$$\| (*)$$

$$f^*(x)$$

$$\Rightarrow f^*(x) = \hat{f}(x) \quad \forall x$$

$\Rightarrow f^*(x) \in B(R) \Rightarrow$ end of the proof.

Def: (Equi-continuity) a subset K of $C(X)$ is equicontinuous a $x \in X$ if

$\forall \varepsilon > 0$, $\exists$ a neighborhood B of x such that $\forall t \in B$ and $\forall f \in K$ we have: $\|f(x) - f(t)\|_\infty < \varepsilon$.

K is equi-continuous (everywhere) if K is equi-continuous everywhere in X.

Thm 18.3: (Arzelá - Ascoli thm).

Thm 18.3: (Arzelá - Ascoli thm).

Let X be compact. $K \subset C(X)$ is compact iff:

1) K is closed
2) K is bounded
3) K is equicontinuous

Thm 18.2: Suppose K is a Mercer kernel on a compact metric space X, and $\mathcal{H}_K$ is the associated RKHS. The inclusion $i_K : \mathcal{H}_K \hookrightarrow C(X)$ is COMPACT. In other words, the set $i_K(B(R))$ is compact for any $R > 0$.

Proof: Thm 18.1, and Prop 18.1, Thm 18.1 show that $i_K(B(R))$ is closed and bounded in $C(X)$. By Arzelá - Ascoli Thm, what remains is to show that $B(R)$ is equi-cont.

We have:

$$|f(x) - f(t)| = |(f, K_x - K_t)_{\mathcal{H}_K}|$$

Cauchy - Schwarz

$$\wedge\!\!\!/\!/$$

$$\|f\|_{\mathcal{H}_K} \, \|K_x - K_t\|_{\mathcal{H}_K}$$

$$\text{def } (\cdot,\cdot)_{\mathcal{H}_K} \quad \Big\| \quad f \in B(R) \Rightarrow \|f\|_{\mathcal{H}_K} \le R$$

$$R \sqrt{(K_x - K_t, K_x - K_t)_{\mathcal{H}_K}}$$

$$\|$$

$$R \left[(K_x, K_x)_{\mathcal{H}_K} - (K_x, K_t)_{\mathcal{H}_K} - (K_t, K_x)_{\mathcal{H}_K} + (K_t, K_t)_{\mathcal{H}_K} \right]^{1/2}$$

$\|$ Reproducing property

$$R \sqrt{[K_x(x) - K_x(t)] + [K_t(x) - K_t(t)]}$$

- since K is continuous on compact space $X \times X \Rightarrow K$ is uniformly continuous on $X \times X$

$\Rightarrow \forall x, \forall t, t'$ such that $\|t - t'\|_X \le \delta$ (δ is independent of x, t, t'), we have

$$|K_x(t) - K_x(t')| \le \varepsilon$$

$\Rightarrow$ if $\|x - t\|_X \le \delta$, then

$$|K_x(x) - K_x(t)| \le \varepsilon$$

$$|K_t(x) - K_t(t)| \leq \varepsilon$$

Thus:

$$|f(x) - f(t)| \leq R\sqrt{[K_x(x) - K_x(t)] + [K_t(x) - K_t(t)]}$$
$$\leq R\sqrt{2\varepsilon} \quad \forall \|x-t\|_X \leq \delta$$
$$\forall f \in i_K(B(R))$$

$\Rightarrow$ $i_K(B(R))$ is equi-continuous, and this ends the proof

Remark: In this class we take the hypothesis space as a compact subset of $C(X)$. And This is justified by this lecture.

Ex: Let $\{\varphi_i\}_{i=1}^M$ be an orthonormal subset of $L^2(0,1)$, i.e.,

$$\int_0^1 \varphi_j(x)\varphi_i(x)\,dx = (\varphi_i, \varphi_j)_{L^2} = \delta_{ij}$$

Define $K(x, x') := \sum_{j=1}^{M} \varphi_j(x)\varphi_j(x')$

Claims:

1) $K(\cdot,\cdot)$ is an SPD kernel.

2) If we define

$$\mathcal{H}_K := \left\{ \sum_{j=1}^{M} a_j \varphi_j(x) \; : \; a_1, \dots, a_M \in \mathbb{R} \right\}$$

then $\mathcal{H}_K$ is the RKHS of $K(\cdot,\cdot)$.

Proof:

1) Recall that we just need to show that

$$A := \begin{bmatrix} K(x_1, x_1) & \dots & K(x_1, x_n) \\ \vdots & & \\ K(x_n, x_1) & \dots & K(x_n, x_n) \end{bmatrix} \geq 0$$

$$\Updownarrow$$

$$\vec{a}^T A \vec{a} \geq 0 \quad \forall \, \vec{a} \in \mathbb{R}^n$$

but this is clear because

$$\vec{a}^T A \vec{a} = \sum_{i,j} a_i K(x_i, x_j) a_j$$

$$\Big\| \text{ def } K(x_i, x_j) = \sum_k \varphi_k(x_i) \varphi_k(x_j)$$

$$\sum_{i,j,k} a_i a_j \varphi_k(x_i) \varphi_k(x_j)$$

$$\|$$

$$\sum_k \left[\sum_i a_i \varphi_k(x_i) \right] \left[\sum_j a_j \varphi_k(x_j) \right]$$

$$\sum_k \left[\sum a_i \varphi_k(x_i) \right]^2 \geq 0$$

2) First notice that

$$K_x(\cdot) := K(x,\cdot) := \sum_j \underbrace{\varphi_j(x)}_{a_j} \varphi_j(\cdot) \in H_k$$

Reproducing property:

$$\left(f(\cdot), K_x(\cdot) \right)_{L^2} = \left(f(\cdot), K(x,\cdot) \right)_{L^2}$$

$f \in H_k$
$f(\cdot) = \sum_{j=1}^{M} a_j \varphi_j(\cdot)$ $\quad \| \quad$ $K(x,\cdot) = \sum_{j=1}^{M} \varphi_j(x)\varphi_j(\cdot)$

$$\left(\sum_{j=1}^{M} a_j \varphi_j(\cdot), \sum_{i=1}^{M} \varphi_i(x) \varphi_i(\cdot) \right)_{L^2}$$

$$\|$$

$$\sum_{i,j} a_j \varphi_i(x) \left(\varphi_j(\cdot), \varphi_i(\cdot) \right)_{L^2}$$

$$\| \quad \left(\varphi_i(\cdot), \varphi_j(\cdot) \right)_{L^2} = \delta_{ij}$$

$$\sum_{i=1}^{M} a_i \varphi_i(x)$$

$$\|\ \text{def } f$$

$$f(x).$$

Thm: (Representer thm): Let $\mathcal{H}_K$ be the RKHS of a SPD kernel K and let $J: \mathbb{R}^n \to \mathbb{R}$.

then:

$$\min_{\substack{f\in \mathcal{H}_K,\\ \|f\|_{\mathcal{H}_K}\le \lambda}} J\left(f(x_1),\dots,f(x_n)\right)$$

$$\Updownarrow \qquad \overline{\mathcal{H}}_K = \left\{ f \in \mathcal{H}_K \,\middle|\, f(\cdot) = g_\alpha(\cdot) = \sum_{i=1}^{n} \alpha_i K(x_i,\cdot) \right\} \subset \mathcal{H}_K$$

$$\min_{\substack{f\in \overline{\mathcal{H}}_K\\ \|f\|_{\mathcal{H}_K}\le \lambda}} J\left(f(x_1),\dots,f(x_n)\right)$$

$$\Updownarrow \qquad A_{ij} = K(x_i, x_j)$$

Gram Matrix

$$\min_{\substack{\vec{\alpha}\in \mathbb{R}^n\\ \vec{\alpha}^T A \vec{\alpha} \le \lambda^2}} J\left(g_\alpha(x_1),\dots,g_\alpha(x_n)\right)$$

Proof: Since $\overline{\mathcal{H}}_K \subset \mathcal{H}_K$ then we have the following decomposition: $f \in \mathcal{H}_K$

$$\Updownarrow$$

$$f = \bar{f} + f^{\perp} \quad \text{where } \bar{f} \in \overline{\mathcal{H}}_K, \quad f^{\perp} \in \overline{\mathcal{H}}_K^{\perp}$$

Pythagoreans thm shows

$$\|f\|_{\mathcal{H}_K}^2 = \|\bar{f}\|_{\mathcal{H}_K}^2 + \|f^{\perp}\|_{\mathcal{H}_K}^2$$

$$\|f\|^2_{\mathcal{H}_K} = \|\bar f\|^2_{\mathcal{H}_K} + \|f^\perp\|^2_{\mathcal{H}_K}$$

Now: Since $K(x_i,\cdot) \in \bar{\mathcal{H}}_K$, we have

$$f^\perp(x_i) = \underbrace{\left(f^\perp(\cdot), K(x_i,\cdot)\right)_{\mathcal{H}_K}}_{0}$$

$f^\perp \in \bar{\mathcal{H}}_K^\perp$ $\qquad$ $K(x_i,\cdot) \in \mathcal{H}_K$

Then $f(x_i) = \bar f(x_i) + f^\perp(x_i) = \bar f(x_i)$

$$\Downarrow$$

$$J(f(x_1),\dots,f(x_n)) = J(\bar f(x_1),\dots,\bar f(x_n))$$

Since $f^\perp$ does not contribute to J $\Rightarrow$ remove it from the constraint; i.e.,

$$\min_{\substack{f\in\mathcal{H}_K \\ \|\bar f\|^2 + \|f^\perp\|^2 \le \lambda^2}} J(f(x_1),\dots,f(x_n)) = \min_{\substack{f\in\mathcal{H}_K \\ \|\bar f\|^2 \le \lambda^2}} J(\bar f(x_1),\dots,\bar f(x_n))$$

* what remains is to show that

if $g_\alpha := \sum_{i=1}^n \alpha_i K(x_i,\cdot) \in \bar{\mathcal{H}}_K$

$$\|g_\alpha\|^2_{\mathcal{H}_K} \le \lambda^2 \iff \vec\alpha^T A\vec\alpha \le \lambda^2$$

$$\|$$

$$\left(\sum_i \alpha_i K(x_i, \cdot), \sum_j \alpha_j K(x_j, \cdot) \right)_{\mathcal{H}_K}$$

$$\|$$

$$\sum_{i,j} \alpha_i \alpha_j \left(K(x_i, \cdot), K(x_j, \cdot) \right)_{\mathcal{H}_K}$$

$\|$ Reproducing property

$$\sum_{i,j} \alpha_i \alpha_j K(x_i, x_j)$$

$\|$ def A

$$\sum_{i,j} \alpha_i \alpha_j A_{ij}$$

$$\|$$

$$\vec{\alpha}^T A \vec{\alpha}$$