Reproducing Kernel Hilbert Space

Hilbert Space & RKHS

- Hilbert space : a complete linear space where inner product between functions is defined.
- Reproducing Kernel Hilbert space on domain $\mathcal{X}$: Hilbert space where the evaluation functional $L_x(f) = f(x)$ is bounded
  - functional is bounded if there exists a constant M such that $|L(f)| \le M\|f\|,\ \forall f \in \mathcal{H}$
- Riesz Representation Theorem

  For every bounded linear functional L on a hillbert space $\mathcal{H}$, there exist a unique $\xi_L \in \mathcal{H}$ such that $L(f) = \langle \xi_L, f \rangle, \forall f \in \mathcal{H}$. ξ_L is called the $representer$ of L

  $\rightarrow$ This means that **every evaluation functional in RKHS have its own representator!!**
- There exist $\xi_x \in \mathcal{H},$ the representer of $L_x(\cdot)$, such that $\langle \xi_x, f \rangle = f(x), \forall f \in \mathcal{H}$
- Define $K(x,t) = \xi_x(t)$, called the ***reproducing kernel*** (RK) which is bivariate function.

  $\rightarrow$ We can take this kernel as the representer of point x in any other function in $\mathcal{H}$!!

Properties of RK

- nonnegative definite (= semi positive definite):

  for every n which is finite, and every $x_1, \cdots, x_n \in \mathcal{X}$, and every $a_1, \cdots, a_n \in \mathbb{R}$

  $$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j K(x_i, x_j) \ge 0$$

  which means that matrix composed of outcomes of $K(x_i, x_j)$ is always n.n.d

  $$a'Ka \ge 0, \forall a$$
- RK is non-negative

  $$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j K(x_i, x_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \langle K(x_i, \cdot), K(x_j, \cdot) \rangle \\ &= \left\langle a_i \sum_{i=1}^{n} K(x_i, \cdot), \sum_{j=1}^{n} a_j K(x_j, \cdot) \right\rangle \\ &= \left\| \sum_{i=1}^{n} K(x_i, \cdot) \right\|^2 \ge 0\end{aligned}$$
- The Moore- Anronszajn Theroem
  - For every RKHS $\mathcal{H}$ of functions on $\mathcal{X}$, there correspponds a unique RK $K(s,t)$
  - Conversely, for every n.n.d function $K(s,t)$ on $\mathcal{X}$, there corresponds a unique RKHS $\mathcal{H}_K$

  $\rightarrow$ **Reproducing Kernel and RKHS has one to one correspondence.**

Construct RKHS by function decomposition

- Constructing RKHS by Anronszaijn Theorem
  1. Make space of functions which has form of $f(x) = \sum_m \alpha_m K(x, y_m)$
  2. Define inner product by $\langle K(x,\cdot), K(y,\cdot)\rangle = K(x,y)$
  3. Complete that space

- Mercer-Hilbert-Schmidt Theorem (Eigen-expansion of kernel)

  Mercer Kernel : a n.n.d function on $\mathcal{X} \times \mathcal{X}$ which satisfy

  $$\int_{\mathcal{X}} \int_{\mathcal{X}} K^2(x,y)dxdy < \infty$$

  this condition is trivial when $\mathcal{X}$ is compact

- Mercer Kernel can be decomposed with continuous orthonormal eigenfunctions in L_2 and eigen value

  $$K(x,y) = \sum_{i=1}^{\infty} \gamma_i \phi_i(x) \phi_i(y)$$

- L_2 inner product of two univariate function(ϕ_i, ϕ_j) is defined as $\int \phi_i(x)\phi_j(x)dx = \delta_{ij}$
- It follows that
  1. $$\begin{aligned} \int_{\mathcal{X}} \int_{\mathcal{X}} K^2(x,y)dxdy &= \sum_i \sum_j \gamma_i \gamma_j \int \phi_i(x)\phi_j(x)dx \int \phi_i(y)\phi_j(y)dy \\ &= \sum_i \gamma_i^2 < \infty \end{aligned} \tag{1}$$
  2. $$\begin{aligned} \int_{\mathcal{X}} K(x,y)\phi_j(x)dx &= \int \sum_i \gamma_i \phi_i(x)\phi_i(y)\phi_j(x)dx \\ &= \sum_i \gamma_i \phi_i(y) \int \phi_i(x)\phi_j(x)dx \\ &= \gamma_j \phi_j(y) \end{aligned} \tag{2}$$

  $\rightarrow$ Because of These two properties, mercer kernel is RK in $\mathcal{H}_K$

- $f \in \mathcal{H}_K$ has form (by Anronszaijn thm 1)

$$
\begin{aligned}
f(x) &= \sum_m \alpha_m K(x, y_m) \\
&= \sum_m \sum_i \alpha_m \gamma_i \phi_i(x) \phi_i(y_m) \\
&= \sum_i \sum_m \alpha_m \phi_i(y_m) \gamma_i \phi_i(x) \\
&= \sum_i c_i \phi_i(x) \\
\text{where } \ c_i &= \gamma_i \sum_m \alpha_m \phi_i(y_m) \\
&= \sum_m \alpha_m \gamma_i \phi_i(y_m) \\
&= \sum_m \alpha_m (K(x, y_m), \phi_i(x)) \\
&= (\sum_m \alpha_m K(x, y_m), \phi_i(x)) \\
&= (f, \phi_i(x))
\end{aligned}
$$

- $\langle K(x,\cdot), f\rangle_{\mathcal{H}_K} = f(x)$ implies $\langle K(x,\cdot), \phi_j\rangle_{\mathcal{H}_K} = \sum_i \gamma_i \phi_i(x) \langle \phi_i, \phi_j\rangle_{\mathcal{H}_K} = \phi_j(x)$

  $\rightarrow$ by these properties, **Inner product of $\mathcal{H}_K$** can be expressed as

$$
\begin{aligned}
(\phi_k, \phi_j) &= \left(\phi_k, \sum_i \gamma_i \phi_i \left\langle \phi_i, \phi_j \right\rangle\right) \\
&= \sum_i \gamma_i (\phi_k, \phi_i) \left\langle \phi_i, \phi_j \right\rangle \\
&= \sum_i \gamma_i \left\langle \phi_i, \phi_j \right\rangle \int \phi_k(x) \phi_i(x) dx \\
&= \gamma_k \left\langle \phi_k, \phi_j \right\rangle \\
\therefore \left\langle \phi_k, \phi_j \right\rangle_{\mathcal{H}_K} &= \frac{(\phi_k, \phi_j)}{\gamma_k}
\end{aligned}
$$

$$
\begin{aligned}
\therefore \left\langle f, g \right\rangle_{\mathcal{H}_K} &= \left\langle \sum_i^{\infty} c_i \phi_i(x), \sum_j^{\infty} d_j \phi_j(x) \right\rangle \\
&= \left\langle \sum_i (f, \phi_i) \phi_i(x), \sum_j (g, \phi_j) \phi_j(x) \right\rangle \\
&= \sum_i \frac{(f, \phi_i)(g, \phi_i)}{\gamma_i} \qquad \because \text{basis functions are orthonormal} \\
&= \sum_i \frac{c_i d_i}{\gamma_i}
\end{aligned}
$$

  and finite norm constraint becomes

$$
\|f\|^2 = \sum_i^{\infty} \frac{c_i^2}{\gamma_i} < \infty
$$

- Then is $K(x,\cdot) \in \mathcal{H}_K$??

$$K(x,\cdot)=\sum_i \gamma_i\phi_i(x)\phi_i(\cdot)=\sum_i c_i\phi_i(\cdot)$$
$$\text{and}$$
$$\|K(x,\cdot)\|^2=\sum_i \frac{\gamma_i^2\phi_i(x)^2}{\gamma_i}=\sum \gamma_i\phi_i(x)\phi_i(x)=K(x,x)<\infty$$
$$\therefore K(x,\cdot)\in\mathcal{H}_K$$

- Inner product between RK and f

$$\begin{aligned}\langle K(x,\cdot),f\rangle &= \sum_i \frac{\gamma_i\phi_i(x)c_i}{\gamma_i} \\ &= \sum_i c_i\phi_i(x)=f(x)\end{aligned}$$

- Reproducing property

$$\begin{aligned}\langle K(x,\cdot),K(y,\cdot)\rangle &= \sum_i \frac{\gamma_i^2\phi_i(x)\phi_i(y)}{\gamma_i} \\ &= \sum_i \gamma_i\phi_i(x)\phi_i(y)=K(x,y)\end{aligned}$$

  or it can be shown by definition of RK

$$K(y,\cdot)\in\mathcal{H}_k$$
$$\therefore \langle K(x,\cdot),f\rangle=f(x)=K(y,x)=K(x,y)$$

Split of Hilbert space

- If two Hilbert spaces $\mathcal{H}_0$ and $\mathcal{H}_1$ equipped with inner product respectively have the only common element {0}, then we define the tensorsum Hilbert space $\mathcal{H}=\{f=f_0+f_1: f_0\in\mathcal{H}_0, f_1\in\mathcal{H}_1\}$ with inner product $\langle\cdot,\cdot\rangle=\langle\cdot,\cdot\rangle_0+\langle\cdot,\cdot\rangle_1$ and write $\mathcal{H}=\mathcal{H}_0\oplus\mathcal{H}_1$
- Sum of two n.n.d functions defined on the same domain is n.n.d
- If K_0 is RK for $\mathcal{H}_0$ and K_1 is RK for $\mathcal{H}_1$ with $\mathcal{H}_0\cap\mathcal{H}_1=\{0\}$, then $K=K_0+K_1$ is RK for $\mathcal{H}=\mathcal{H}_0\oplus\mathcal{H}_1$
- If an n.n.d function K in decomposed into two orthogonal n.n.d functions $K_0 \text{ and } K_1$, then $\mathcal{H}=\mathcal{H}_0\oplus\mathcal{H}_1$, where $\mathcal{H}_0,\mathcal{H}_1$ are RKHS corrsponding to K_0,K_1 Because of one to one correspondence of RK and RKHS,

  $\rightarrow$ **We can use each of function space as block** like ANONA

Regularization Problems Using RKHS

- we want to use norm of RKHS as penalty
- First variational problem with $J[f]=\|f\|^2_{\mathcal{H}_K}$

$$\underset{f\in\mathcal{H}_K}{min}\left[\sum_{i=1}^N L(y_i,f(x_i))+\lambda\|f\|^2_{\mathcal{H}_K}\right]$$

- The solution is finite dimension and has the form of $f(x) = \sum_i^N \alpha_i K(x_i, x)$ where x_i is observed pts.

  $\rightarrow$ We can change infinite dimension problems into finite dimension problems with N dimension

  pf)

  for any functions $\widetilde{g} \in \mathcal{H}_k$ can be decomposed as $g + p$ where $g(x) = \sum_i^N \alpha_i K(x_i, x)$ and $\rho(x) \perp K(x_i, x), \forall i = 1, \cdots, N$. Then $\rho(x_i) =< K(x_i, x), \rho(x) >= 0$

  Thus, $\widetilde{g}(x_i) = g(x_i)$ and $J[\widetilde{g}] = \|\widetilde{g}\|^2_{\mathcal{H}_K} =< g+\rho, g+\rho >= \|g\|^2 + \|\rho\|^2 \geq \|g\|^2 = J[g]$

  equality holds when $\|\rho\|^2 = 0 \rightarrow$ we can find this regardless of loss function!!

- We can change problem as

$$\underset{\alpha}{min}[L(y, K\alpha) + \lambda\alpha' K\alpha]$$

  where $K = \{K(x_i, x_j)\}$

- Second variational problam with $J[f] = \|P_1 f\|^2_{\mathcal{H}_1}$ where P_1 is projection onto $\mathcal{H}_1$

$$\underset{f \in \mathcal{H}_K}{min}\left[\sum_{i=1}^N L(y_i, f(x_i)) + \lambda\|P_1 f\|^2_{\mathcal{H}_1}\right]$$

- The solution is finite dimensional and has the form of $f(x) = \sum_{j=1}^m \beta_j \psi_j(x) + \sum_i^N \alpha_i K_1(x_i, x)$
- We can change problem as

$$\underset{\alpha,\beta}{min}[L(y, T\beta + K_1\alpha) + \lambda\alpha' K_1\alpha]$$

  where $T = \{\psi_j(x_i)\}, K_1 = \{K_1(x_i, x_j)\}$

- By using this, We can penalize only the sunset of $\mathcal{H}_k$