Fock spaces and the universal approximation theorem for Gaussian kernels [1]

Michio Seto (National Defense Academy)

May 20, 2023

[1] Supported by JSPS KAKENHI Grant Numbers JP20K03646.

Introduction 1

k: a real-valued function on $X \times X$.

Definition

k is called a kernel function if $k(x, y) = k(y, x)$ $(\forall x, y \in X)$ and

$\forall n \in \mathbb{N}$, $\forall \{c_j\}_{j=1}^n \subset \mathbb{R}$, $\forall \{x_j\}_{j=1}^n \subset X$,

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0.$$

Introduction 1

k: a real-valued function on $X \times X$.

Definition

k is called a kernel function if $k(x, y) = k(y, x)$ $(\forall x, y \in X)$ and

$\forall n \in \mathbb{N}$, $\forall \{c_j\}_{j=1}^n \subset \mathbb{R}$, $\forall \{x_j\}_{j=1}^n \subset X$,

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0.$$

2016 Bangalore

"Experts in RKHS are needed in the job market" (J. Ball)

Introduction 1

k: a real-valued function on $X \times X$.

Definition

k is called a kernel function if $k(x, y) = k(y, x)$ $(\forall x, y \in X)$ and

$\forall n \in \mathbb{N}$, $\forall \{c_j\}_{j=1}^{n} \subset \mathbb{R}$, $\forall \{x_j\}_{j=1}^{n} \subset X$,

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0.$$

2016 Bangalore

"Experts in RKHS are needed in the job market" (J. Ball)

Kernel trick

For given data $x_1, \ldots, x_n$, $x_j \mapsto k_{x_j} := k(\cdot, x_j)$.

Introduction 2

In the kernel method, predictions are given in the form

$$\sum_{j=1}^{n} c_j k_{x_j}(x_*) \quad (x_* : \text{a new input}).$$

Particularly, in Gaussian process regression,

$$k(x, y) = \exp(-\gamma \|x - y\|_{\mathbb{R}^d}^2) \quad (\gamma > 0)$$

plays very important roles.

- Universal Approximation Theorem (UAT) ensures representability of Gaussian kernels.
- In this talk, we shall give a very short proof of UAT for the Gaussian.

UAT for the Gaussian

Theorem (Steinwart, Micchelli-Xu-Zhang, Guella)

$\forall \gamma > 0, \quad \forall K \subset \mathbb{R}^d$ (compact), $\quad \forall \varepsilon > 0, \quad \forall f \in C_{\mathbb{R}}(K),$

$\exists c_1, \ldots, c_n \in \mathbb{R}, \quad \exists x_1, \ldots, x_n \in \mathbb{R}^d$

$$\text{s.t.} \quad \sup_{x \in K} \left| f(x) - \sum_{j=1}^{n} c_j \exp(-\gamma \|x - x_j\|_{\mathbb{R}^d}^2) \right| < \varepsilon.$$

UAT for the Gaussian

Theorem (Steinwart, Micchelli-Xu-Zhang, Guella)

$\forall \gamma > 0, \quad \forall K \subset \mathbb{R}^d$ (compact), $\quad \forall \varepsilon > 0, \quad \forall f \in C_{\mathbb{R}}(K),$

$\exists c_1, \ldots, c_n \in \mathbb{R}, \quad \exists x_1, \ldots, x_n \in \mathbb{R}^d$

$$\text{s.t.} \quad \sup_{x \in K} \left| f(x) - \sum_{j=1}^{n} c_j \exp(-\gamma \|x - x_j\|_{\mathbb{R}^d}^2) \right| < \varepsilon.$$

Simple observation

Since

$$\left| f(x) - \sum_{j=1}^{n} c_j e^{-\gamma \|x - x_j\|^2} \right| = e^{-\gamma \|x\|^2} \left| e^{\gamma \|x\|^2} f(x) - \sum_{j=1}^{n} c_j e^{-\gamma \|x_j\|} e^{2\gamma \langle x, x_j \rangle} \right|,$$

it suffices to consider exponentials of kernel functions.

Setting

X: a locally compact Hausdorff space,

k: a real-valued continuous kernel function on $X \times X$,

$\mathcal{H}_k$: the real RKHS generated by k,

$$\mathcal{F}_0(\mathcal{H}_k) := \bigoplus_{n \geq 0} \mathcal{H}_k^{\otimes n} \quad \text{(algebraic direct sum)},$$

$\mathcal{E}(\mathcal{H}_k)$: the completion of $\mathcal{F}_0(\mathcal{H}_k)$ w.r.t $\left(\sum_{n=0}^{\infty} \frac{1}{n!} \|F_n\|^2_{\mathcal{H}_k^{\otimes n}}\right)^{1/2}$,

$$F_n \circ \Delta_n(x) := F_n(x, \ldots, x) \quad (F_n \in \mathcal{H}_k^{\otimes n}),$$

$$\Gamma : \begin{pmatrix} F_0 \\ F_1 \\ F_2 \\ \vdots \end{pmatrix} \mapsto \sum_{n=0}^{\infty} \frac{1}{n!} (F_n \circ \Delta_n) \quad \left(\begin{pmatrix} F_0 \\ F_1 \\ F_2 \\ \vdots \end{pmatrix} \in \mathcal{E}(\mathcal{H}_k) \right).$$

A Very Short Proof of UAT

1. $\Gamma(\mathcal{E}(\mathcal{H}_k)) = \mathcal{H}_{\exp k}$.

$$\because \quad \Gamma \begin{pmatrix} 1 \\ k_x \\ k_x^{\otimes 2} \\ \vdots \end{pmatrix} = \sum_{n=0}^{\infty} \frac{1}{n!} (k_x^{\otimes n} \circ \Delta_n) = \sum_{n=0}^{\infty} \frac{1}{n!} k_x^n = \exp k_x .$$

A Very Short Proof of UAT

1. $\Gamma(\mathcal{E}(\mathcal{H}_k)) = \mathcal{H}_{\exp k}$.

$$\because \quad \Gamma\begin{pmatrix} 1 \\ k_x \\ k_x^{\otimes 2} \\ \vdots \end{pmatrix} = \sum_{n=0}^{\infty} \frac{1}{n!}(k_x^{\otimes n} \circ \Delta_n) = \sum_{n=0}^{\infty} \frac{1}{n!} k_x^n = \exp k_x.$$

2. $\Gamma(\mathcal{F}_0(\mathcal{H}_k))$ is an algebra.

$$\because \quad \forall F_n \in \mathcal{H}_k^{\otimes n}, \quad \forall G_m \in \mathcal{H}_k^{\otimes m},$$

$$\begin{aligned}\Gamma(F_n)\Gamma(G_m) &= \frac{1}{n!}(F_n \circ \Delta_n)\frac{1}{m!}(G_m \circ \Delta_m) \\ &= \frac{1}{n!m!}(F_n \otimes G_m) \circ \Delta_{n+m} \\ &= \frac{(n+m)!}{n!m!}\Gamma(F_n \otimes G_m).\end{aligned}$$

A Very Short Proof of UAT

3. $\Gamma(\mathcal{F}_0(\mathcal{H}_k)) \subset C_{\mathbb{R}}(X)$.

$$\because \quad \Gamma(\mathcal{F}_0(\mathcal{H}_k)) \subset \Gamma(\mathcal{E}(\mathcal{H}_k)) = \mathcal{H}_{\exp k} \subset C_{\mathbb{R}}(X).$$

A Very Short Proof of UAT

3. $\Gamma(\mathcal{F}_0(\mathcal{H}_k)) \subset C_{\mathbb{R}}(X)$.

$$\because \quad \Gamma(\mathcal{F}_0(\mathcal{H}_k)) \subset \Gamma(\mathcal{E}(\mathcal{H}_k)) = \mathcal{H}_{\exp k} \subset C_{\mathbb{R}}(X).$$

4. Conclusion.

Suppose that $\mathcal{H}_k$ separates any distinct two points in X.
Then, by the Stone-Weierstrass theorem,

$$\overline{\{g|_K : g \in \Gamma(\mathcal{F}_0(\mathcal{H}_k))\}}^{\|\cdot\|_{K,\infty}} = C_{\mathbb{R}}(K) \quad (K : \textit{compact}).$$

Hence, $\forall \varepsilon > 0$, $\forall f \in C_{\mathbb{R}}(K)$, $\exists g \in \Gamma(\mathcal{F}_0(\mathcal{H}_k))$ and

$$\exists h = \sum_{j=1}^{n} c_j \exp k_{x_j} \in \mathcal{H}_{\exp k}$$

$$\begin{aligned} \text{s.t.} \quad |f(x) - h(x)| &\leq |f(x) - g(x)| + |g(x) - h(x)| \quad (\forall x \in K) \\ &\leq \|f - g\|_{K,\infty} + \|g - h\|_{\mathcal{H}_{\exp k}} \|\exp k_x\|_{\mathcal{H}_{\exp k}} < \varepsilon. \end{aligned}$$

Further applications of $\Gamma(\mathcal{F}_0(\mathcal{H}_k)) \subset \mathcal{H}_{\exp k}$

Theorem (S)

$$k(\lambda, \mu) = \exp\left(-t\left|\frac{\lambda - \mu}{1 - \bar{\lambda}\mu}\right|^2\right) \quad (t > 0)$$

is strictly positive definite on $\mathbb{D} = \{\lambda \in \mathbb{C} : |\lambda| < 1\}$.

Further applications of $\Gamma(\mathcal{F}_0(\mathcal{H}_k)) \subset \mathcal{H}_{\exp k}$

Theorem (S)

$$k(\lambda, \mu) = \exp\left(-t \left|\frac{\lambda - \mu}{1 - \bar{\lambda}\mu}\right|^2\right) \quad (t > 0)$$

is strictly positive definite on $\mathbb{D} = \{\lambda \in \mathbb{C} : |\lambda| < 1\}$.

Definition

k is said to be strictly positive definite if

$$\forall n \in \mathbb{N},\ \forall x_1, \ldots, x_n \in X \text{ (distinct)},\ \forall (c_1, \ldots, c_n) \in \mathbb{R}^n \setminus \{\mathbf{0}\}$$

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) > 0.$$