Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 10, n. 1, 2023.

Trabalho apresentado no XLII CNMAC, Universidade Federal de Mato Grosso do Sul - Bonito - MS, 2023

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Growth rates for the ϵ-Entropy of RKHS generated by DPK on the sphere

Douglas Azevedo[1], Michele C. Valentino[2]
DAMAT-UTFPR, Cornélio Procópio, PR

1 Introduction

The notion of metric-entropy was created by Kolmogorov aiming to classify compact metrics sets according to their "massivity" has found several applications in many areas of Mathematics. In particular, we highlight that the connection between entropy quantities and bounded linear operators on Banach spaces, is one of the main features of the Statistical Learning Theory [1]. Moreover, DPKs are very popular in Learning Theory (see [2], for instance) and certainly the most important feature of these kernels is the fact that the class of DPKs includes the prominent Gaussian kernel. In Learning Theory, in order to estimate the error between the empirical target function and the target function (belonging to an RKHS), metric entropies quantities such as ϵ-entropy are needed in various ways ([3]). Here, we aim the lower bounds of the ϵ-entropy for the embedding, $I_K : \mathcal{H}_K \to C(S^m)$, of the unit ball of a dot product kernel Hilbert space (DPKHS) into the space of continuous functions on the unit sphere of $\mathbb{R}^{m+1}$.

Recall that, for a metric space $X = (X, d)$ and $S \subset C$, we say that U_ϵ is an ϵ-cover for S if, for all $x \in S$, there is an $u \in U_\epsilon$ such that $d(x, u) < \epsilon$. The ϵ-covering number of S, denoted by $\mathcal{N}(\epsilon, S) = \mathcal{N}(\epsilon, S, d)$ is defined as the size of the smallest ϵ-cover of S. The nth entropy number of S is defined by

$$e_n(S) = e_n(S, d) = \inf\{\epsilon > 0;\ \mathcal{N}(\epsilon, S) \leq n\}. \tag{1}$$

The (Kolmogorov) ϵ- entropy, denoted by $\mathcal{K}(\epsilon, S) = \mathcal{K}(\epsilon, S, d)$, in turn, is also defined in terms of the $\mathcal{N}(\epsilon, S)$ by

$$\mathcal{K}(\epsilon, S) = \log(\mathcal{N}(\epsilon, S)). \tag{2}$$

Above and throughout this note, log denotes the natural logarithm.

2 Dot product kernels (DPK) and its RKHS

Let $\{b_n\}$ be a summable sequence of positive numbers. The symbol $\cdot$ stands for the usual inner product of $\mathbb{R}^{m+1}$. Let S^m $(m \geq 2)$ be the unit sphere in $\mathbb{R}^{m+1}$ endowed with its induced Lebesgue measure σ_m and write $L^2(S^m) := L^2(S^m, \sigma_m)$. Kernels of the form

$$K(x, y) = \sum_{k=1}^{\infty} b_k (x \cdot y)^k, \quad x, y \in S^m, \tag{3}$$

[1]douglasa@utfpr.edu.br
[2]valentino@utfpr.edu.br

with $\sum_{k=1}^{\infty} b_k < \infty$, are called *dot product kernels (DPK)* on S^m. Clearly, K is continuous. Also, it is positive definite, that is, $\sum_{i,j=1}^{N} c_i c_j K(x_i, x_j) \geq 0$, for all $N > 1$, and every $c_i, c_j \in \mathbb{R}$, and $x_i, x_j \in S^m$. The theory of RKHS ensures that there exists a unique Hilbert space $\mathcal{H}_K$ of functions on S^m (called reproducing kernel Hilbert - RKHS) for which all linear evaluation functionals $F_x(f) := f(x),\ f \in \mathcal{H}_K,\ x \in X$ are continuous. From the results obtained in [4] for the (compact) integral operator $T : L^2(S^m) \to L^2(S^m)$

$$T(f)(x) = \int_{S^m} \left(\sum_{k=1}^{\infty} b_k (x \cdot y)^k \right) f(y)\, d\sigma_m(y), \quad x \in S^m, \quad f \in L^2(S^m), \tag{4}$$

under certain assumption on the sequence $\{b_n\}$ and via tools from harmonic analysis we are able to characterize the $\mathcal{H}_K$.

3 Estimating the ϵ-entropy

Let X and Y be Banach spaces with unit balls given by B_X and B_Y, respectively. For $\epsilon > 0$, the covering numbers of an operator $T : X \to Y$ are defined as

$$\mathcal{N}(\epsilon, T) = \mathcal{N}(\epsilon, T(B_X)) := \min \left\{ n \in \mathbb{N} : \ \exists y_1, ..., y_n \in Y \ s.t. \ T(B_X) \subset \bigcup_{j=1}^{n} (y_j + \epsilon B_Y) \right\}. \tag{5}$$

The (Kolmogorov) ϵ-entropy of an operator $T : X \to Y$ is then defined as

$$\mathcal{K}(\epsilon, T) = \mathcal{K}(\epsilon, T(B_X)) := \log(\mathcal{N}(\epsilon, T)). \tag{6}$$

A lower bound for the ϵ-entropy of the embedding $I_K : \mathcal{H}_K \to C(S^m)$ is describe in the following theorem.

Theorem 3.1. *There exist $A > 0$ with the following property: for any $\epsilon > 0$ sufficiently small,*

$$\mathcal{K}(\epsilon, I_K) \geq A \frac{\left(\log\left(\frac{1}{\epsilon}\right)\right)^{m+1}}{\left(\log\left(\log\left(\frac{1}{\epsilon}\right)\right)\right)^{m}}, \tag{7}$$

holds true.

Acknowledgement

We gratefully acknowledge the financial support from UTFPR-CP.

References

[1] F. Cucker and D. X. Zhou. **Learning Theory: An approximation theory viewpoint**. 1st ed. Cambridge: Cambridge University Press, 2007. ISBN: 9780521865593.

[2] F. Lu and H. Sun. "Positive definite dot product kernels in learning theory". In: **Advances in Computational Mathematics** 22 (2005), pp. 181–198. DOI: 10.1007/s10444-004-3140-6.

[3] D. X. Zhou. "Capacity of reproducing kernel space in learning theory". In: **IEEE Transactions on information theory** 22 (2005), pp. 181–198. DOI: 10.1007/s10444-004-3140-6.

[4] D Azevedo and V. A. Menegatto. "Sharp estimates for eigenvalues of integral operators generated by dot product kernels on the sphere". In: **Journal of Approximation Theory** 177 (2014), pp. 57–68. DOI: 10.1016/j.jat.2013.10.002.

