Deep learning with kernels through RKHM and the Perron–Frobenius operator

Yuka Hashimoto[1,2], Masahiro Ikeda[2,3], and Hachem Kadri[4]

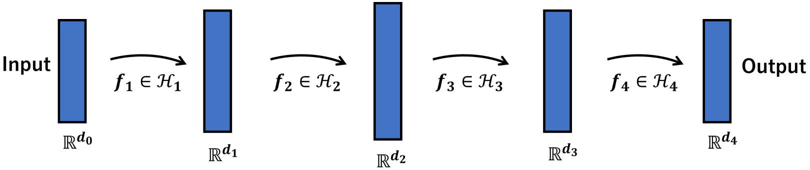
1. NTT Network Service Systems Laboratories, NTT Corporation, Tokyo, Japan
2. Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan
3. Faculty of Science and Technology, Keio University, Yokohama, Japan
4. Aix-Marseille University, CNRS, LIS, Marseille, France

Deep learning with kernels

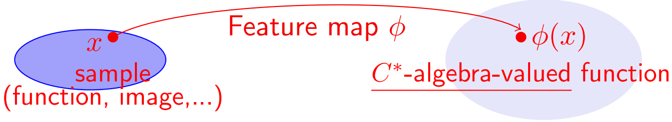
Combine the flexibility of deep neural networks with the representation power and solid theoretical understanding of kernel methods.

k_j : $\mathbb{R}^{d_j \times d_j}$-valued positive definite kernel
$\mathcal{H}_j$: vector-valued RKHS associated with k_j
$\mathcal{G}_j = \{f \in \mathcal{H}_j \mid \|f\|_{\mathcal{H}_j} \le B_j\}$ $(j = 1, \ldots, L)$
$\mathcal{G}_L^{\mathrm{deep}} = \{f_L \circ \cdots \circ f_1 \mid f_j \in \mathcal{G}_j \ (j = 1, \ldots, L)\}$

$$\text{Deep RKHS} : f = f_1 \circ \cdots \circ f_L \tag{1}$$



Generalization of deep kernel methods in RKHS to RKHM



$\mathcal{X}$
(Structured data sp.)

RKHM
(Infinite dimensional Hilbert C^*-module)

Nonlinear

Linear + C^*-algebra-valued inner product

Examples of C^*-algebra:

- $\mathbb{C}^{d\times d} = \{d$ by d matrices$\}$
- $Block((m_1, \ldots, m_M), d) =$
  $\{d$ by d block diagonal matrices with block size $(m_1, \ldots, m_M)\}$

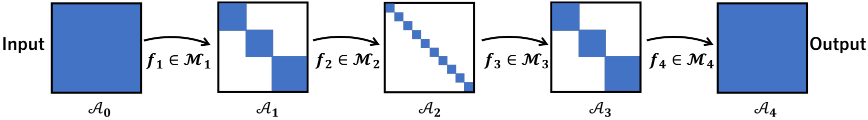
Advantages of RKHM:

- C^*-algebra-valued inner products extract information of structures.
- RKHM is a natural generalization of RKHS.
- Fundamental properties for data analysis (e.g. representer theorem).

Deep RKHM

$\mathcal{A} = \mathbb{C}^{d \times d}$, $\quad \mathcal{A}_j$: C^*-subalgebra of $\mathcal{A}$ (e.g. $Block((m_1, \ldots, m_M), d)$)
k_j : $\mathcal{A}_j$-valued positive definite kernel (ϕ_j : feature map)
$\mathcal{M}_j$: RKHM associated with k_j ($j = 1, \ldots, L$)
$P_f : \mathcal{M}_j \to \mathcal{M}_{j+1}$ (Perron–Frobenius operator) :
$\mathcal{A}$-linear operator satisfying $P_f \phi_j(x) = \phi_{j+1}(f(x))$
$\mathcal{F}_j = \{f \in \mathcal{M}_j \mid \|P_f\| \le B_j\}$ ($j = 1, \ldots, L-1$)
$\mathcal{F}_L = \{f \in \mathcal{M}_L \mid \|f\|_{\mathcal{M}_L} \le B_L\}$
$\mathcal{F}_L^{\mathrm{deep}} = \{f_L \circ \cdots \circ f_1 \mid f_j \in \mathcal{F}_j \ (j = 1, \ldots, L)\}$

$$\text{Deep RKHM} : f = f_L \circ \cdots \circ f_1 \in \mathcal{F}_L^{\mathrm{deep}} \tag{2}$$

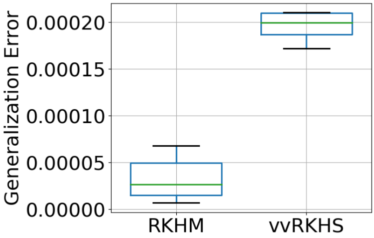


Advantages and properties of deep RKHM with the P–F operators

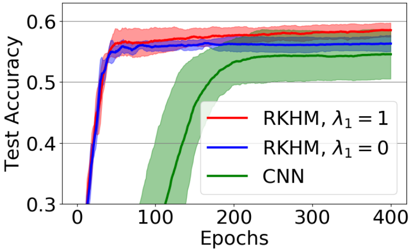
- Useful structures of matrices: Interactions among elements are induced by block diagonal structures of matrices.
- Availability of the operator norm: The operator norm alleviates the dependency of the generalization error on the output dimension.
- Connection with benign overfitting: We derived a generalization bound for deep RKHMs using Perron–Frobenius operators, which provides a connection with benign overfitting.
- Representer theorem: We proved a representer theorem of deep RKHMs involving the Perron–Frobenius operators.

Numerical results

Autoencoder with synthetic data
$(d = 10,\ n = 10,\ L = 3)$



Classification task with MNIST
$(d = 28,\ n = 20,\ L = 2)$



Conclusion

- We investigated deep kernel learning with RKHM.
- We applied Perron–Frobenius operators and the operator norm to derive a generalization bound.
- The dependence of the bound on the output dimension is milder than existing bound by virtue of the operator norm. Moreover, the application of the Perron–Frobenius operator induces a connection with benign overfitting.