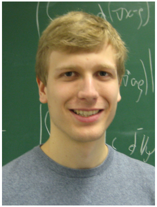
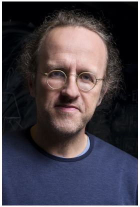
The Inductive Bias of Quantum Kernels

Jonas M. Kübler*, Simon Buchholz*, Bernhard Schölkopf

*equal contribution

EMPIRICAL INFERENCE
MAX PLANCK INSTITUTE FOR
INTELLIGENT SYSTEMS

Quantum Methods in ML

- Quantum computers operate with exponentially large Hilbert spaces.
- Older work: Use QC to speed-up linear algebra routines.[1][2]
- More recent: Use QC to define the function class (Quantum Neural Network or Quantum Kernel)

[1] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd *Quantum algorithm for linear systems of equations*, Physical Review Letters, 103(15), 2009.

[2] Carlo Ciliberto, Andrea Rocchetto, Alessandro Rudi, and Leonard Wossnig. *Statistical limits of supervised quantum learning*, Physical Review A, 102(4), 2020.

Main Messages

- No free quantum-lunch: A model that can represent exponentially many functions, and does not a priori favor few, requires exponentially large training sets.
- Prior knowledge helps: We can reduce the search space with prior knowledge. If we can encode this quantum-mechanically but not classically, we are on track for q-advantage.
- Don't forget to measure: Any q-advantage is lost if the required accuracy of estimates is exponential.

Quantum Kernels

Description of a d-qubit state via *density matrix* ρ ($2^d \times 2^d$ hermitian matrix).

Definition (Quantum Kernel)

Let $\rho : x \mapsto \rho(x)$ be a fixed feature mapping from $\mathcal{X}$ to density matrices. Then the corresponding *quantum kernel* is $k(x, x') = \mathrm{Tr}\left[\rho(x)\rho(x')\right]$.

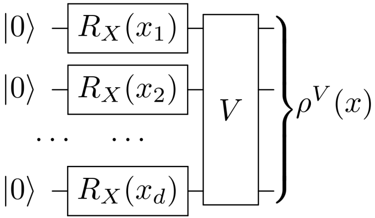
- Computes an inner product in an exponentially large space.
- Has to be estimated from measurement.
- The feature map is *fixed* independently of the data. (But hopefully well chosen for the problem).
- We can learn functions like $f_M(x) = \mathrm{Tr}\left[\rho(x)M\right]$.

An Example Kernel

$\mathcal{X} = \mathbb{R}^d$.

- Dimension $d = 1$

$$
\begin{aligned}
|\psi(x)\rangle &= R_X(x)|0\rangle \\
&= \cos(x/2)|0\rangle + i\sin(x/2)|1\rangle \\
\rho(x) &= |\psi(x)\rangle\langle\psi(x)| \\
k(x,x') &= \mathrm{Tr}\left[\rho(x)\rho(x')\right] = \cos\left(\frac{x-x'}{2}\right)^2
\end{aligned}
$$



- Dimension $d \in \mathbb{N}$

$$
k(x,x') = \prod \cos^2\left(\frac{x_i - x_i'}{2}\right)
$$

This kernel is also classically feasible.

Example (Trivial Quantum Advantage)

Let f be a scalar function that is easily computable on a quantum device but requires exponential resources to approximate classically. Generate data as $Y = f(X) + \epsilon$. The kernel $k(x, x') = f(x)f(x')$ then has an exponential advantage for learning f from data compared to classical kernels.

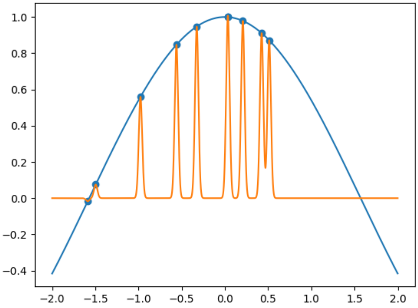
A more rigorous version of this can be found in:
Liu et al. *A rigorous and robust quantum speed-up in supervised machine learning*, Nature Physics 17, 1013–1017 (2021).

Overview setting and approach

- Kernel ridge regression (KRR) for $Y = f(X) + \varepsilon$
- When is learning with KRR easy? Depends on ...
  - ... the target function f.
  - ... the marginal distribution of X, called μ.
  - ... the kernel k.
- We use spectral techniques (Mercer decomposition) to understand learning performance
- Diversity of quantum embedding $x \to \rho(x)$ measured by purity $\mathrm{Tr}\left[\rho_\mu^2\right]$ of mean embedding $\rho_\mu = \int \rho(x)\,\mu(\mathrm{d}x)$

Main Message: No free quantum-lunch

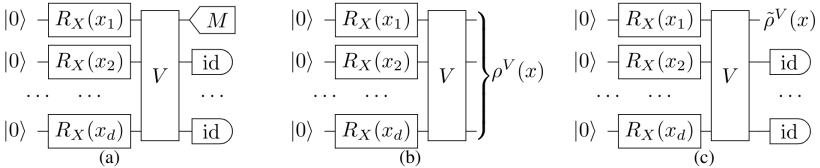
If the encoding exhaust the whole quantum Hilbert space, i.e., when the purity of mean encoding ρ_μ decays exponentially, we need exponentially many datapoints.



Fact: Exponential decay happens for many generic $x \to \rho(x)$

Projected or Biased Quantum Kernels

- Idea: define a kernel on a smaller dimensional subspace than the whole quantum Hilbert space.[3]



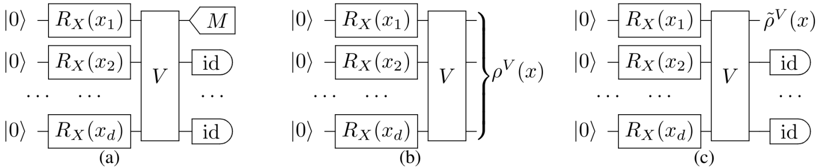
- Data is generated via *a*).

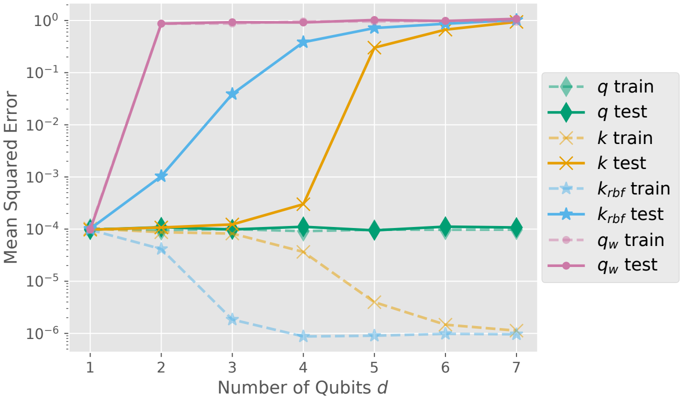
$$f(x) = \mathrm{Tr}\left[\rho^V(x)(M \otimes \mathrm{id})\right] = \mathrm{Tr}\left[\tilde{\rho}^V(x)M\right]$$

- Define the *biased kernel* $q(x, x') = \mathrm{Tr}\left[\tilde{\rho}^V(x)\tilde{\rho}^V(x')\right]$

[3] Huang, HY., Broughton, M., Mohseni, M. et al. Power of data in quantum machine learning. Nat Commun 12, 2631 (2021)

Experiments

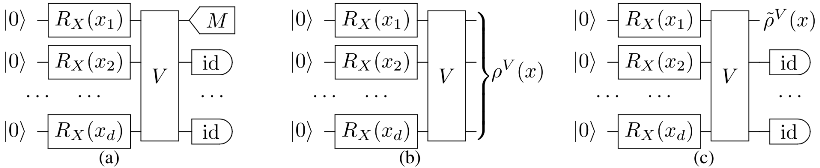




Main Message: Prior knowledge helps

We can reduce the search space with prior knowledge - "how was the problem generated?"

Quantum Advantage?



Are such biased kernels a path to quantum advantages?

- No: So far we ignored the estimation of the quantum kernels.
- Problem: Biased kernels are exponentially close to constant!
  - Why? Because $\tilde{\rho}^V$ is highly mixed.
  - Requires exponentially many measurements to extract the important information

Main Message: Don't forget to measure

- Generally it is not sufficient to measure an outcome to error ε, where ε is "something small".
- If the required error is exponentially small, we cannot harvest a q-advantage.

Thank you!

EMPIRICAL INFERENCE
MAX PLANCK INSTITUTE FOR
INTELLIGENT SYSTEMS