Quantum kernels in practice

Jennifer R Glick
Quantum Applications Researcher

Key ideas about quantum kernels

Qiskit

- Variational quantum classifiers are just *linear methods* in *quantum feature space*.

  --> use quantum kernels!

- Quantum kernels can only be expected to do better than classical kernels if they are *hard to estimate classically*.

  This is a *necessary but not sufficient* condition for a computational advantage.

Havlíček et al. Nature 567, 209 (2019)

Choosing quantum kernels in practice

Qiskit

- Take a quantum kernel that is *hard* to estimate, parametrize it, then optimize it.

  Ex: Forrelation kernel

- Take a quantum kernel that is *easy* to estimate, add entanglement, parametrize it, then optimize it.

  Ex: Gaussian kernel

- Choose a quantum kernel that *exploits structure* in your data.

  Ex: DLOG kernel

Main takeaways for this lecture

Qiskit

- There is no *a priori* reason to expect an *ad hoc* quantum kernel to perform well on a given dataset, even if it is hard to estimate classically.

- One approach to using quantum kernels in practice is to design them to exploit structural insight into a problem.

    - Quantum kernels can *exploit group structure* in data. This is a promising approach because it includes a known case of a quantum speedup. --> DLOG kernel

    - Quantum kernels can be optimized with a technique called *kernel alignment*.

A simple example of a dataset with group structure

Qiskit

1. Group: $G = SU(2)$

2. Subgroup: $S = \{1, a, a^2\}$

$s = 1$: $U_1 = \mathbb{I}$ (identity matrix)

$s = a$: $U_a = A = \exp(i\ [2\pi/3]\ X)$

$= R_x(2\pi/3)$

3. Left-cosets:

$g_1, g_2 \in G$ ➡ $g_1 = 1$: $U_1 = \mathbb{I}$

$g_2 = h$: $U_h = H$ (Hadamard)

$$C_\star = g_1 S = \{g_1 s_i\} = \{1, a, a^2\}$$

$$C_\bullet = g_2 S = \{g_2 s_i\} = \{h, ha, ha^2\}$$

4. Binary classification dataset:

$g_1 s_1,\ g_1 s_2,\ g_1 s_3$ --> +1

$\underbrace{g_2 s_1,\ g_2 s_2,\ g_2 s_3}_{x_i}$ --> $\underbrace{-1}_{y_i}$

$$T = \{(x_i, y_i)\}$$

Glick et al. arxiv:2105.03406

A simple example of a dataset with group structure

Visualizing the dataset

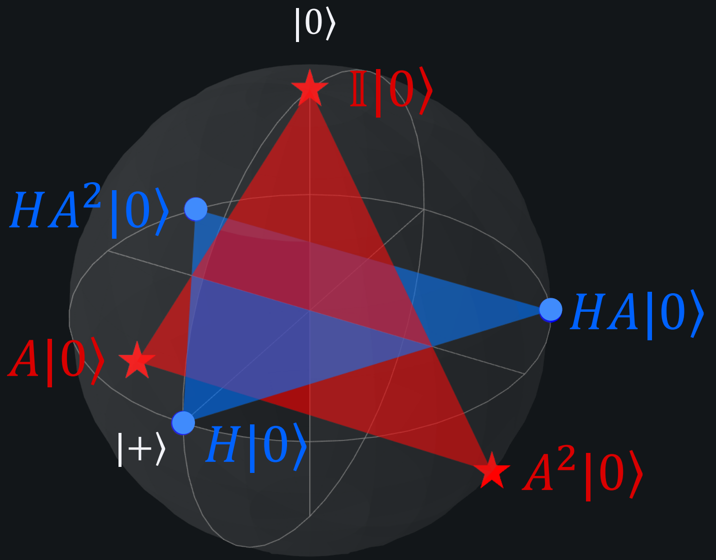
Initialize qubit in the state $|\psi\rangle = |0\rangle$

$$C_{\star}:\ U_{x_i}|0\rangle = U_{g_1}U_{s_i}|0\rangle$$

$$\rightarrow \mathbb{I}|0\rangle, A|0\rangle, A^2|0\rangle \ ,\quad A = \exp(i\,[2\pi/3]\,X)$$

$$C_{\bullet}:\ U_{x_i}|0\rangle = U_{g_2}U_{s_i}|0\rangle$$

$$\rightarrow H|0\rangle, HA|0\rangle, HA^2|0\rangle$$



A simple example of a dataset with group structure

Visualizing the dataset

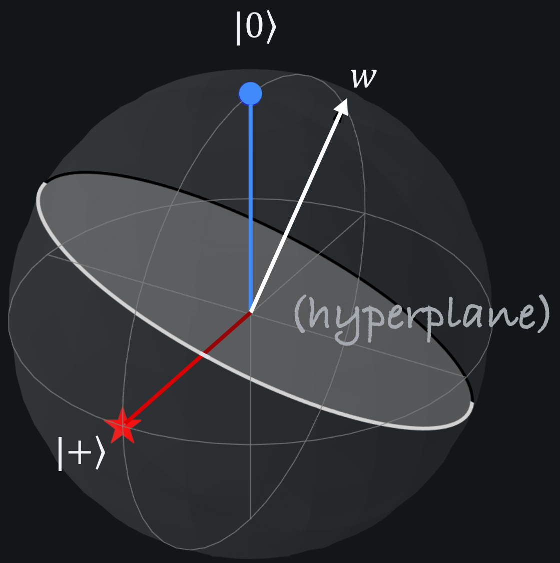
Initialize qubit in the state $|\psi\rangle = |+\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle)$

$$C_{\bigstar}: U_{g_1} U_{s_i} |+\rangle$$

$$\rightarrow \mathbb{I}|+\rangle = |+\rangle, A|+\rangle = |+\rangle, A^2|+\rangle = |+\rangle$$

$$C_{\bullet}: U_{g_2} U_{s_i} |+\rangle$$

$$\rightarrow H|+\rangle = |0\rangle, HA|+\rangle = |0\rangle, HA^2|+\rangle = |0\rangle$$



A simple example of a dataset with group structure

Why does this work?

$|+\rangle$ is a subgroup invariant state:

$$U_s|+\rangle = |+\rangle \quad \text{for all} \quad s \in S$$

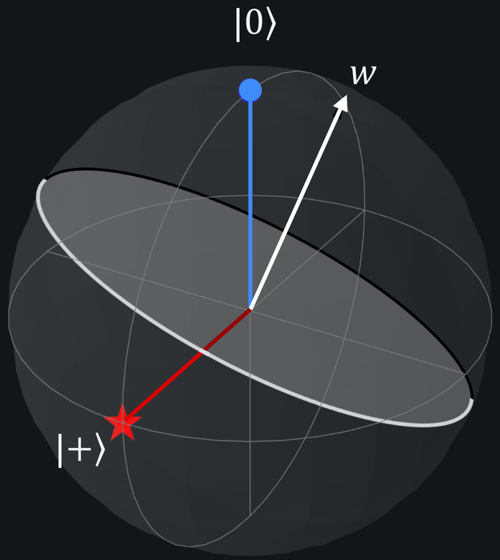
Coset elements act on this state via:

$$U_g U_s|+\rangle = U_g|+\rangle$$

This leads to two quantum feature states:

$$g_1 = 1:\ U_1|+\rangle\langle+|U_1^\dagger = |+\rangle\langle+|$$

$$g_2 = h:\ U_h|+\rangle\langle+|U_h^\dagger = |0\rangle\langle0|$$



A simple example of a dataset with group structure

Building the quantum kernel

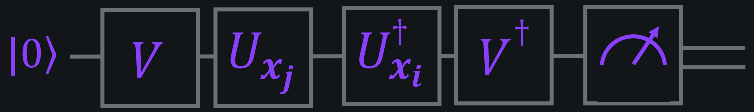
Quantum feature state for a sample x_i :

$$\Phi(x_i) = U_{x_i}|\psi\rangle\langle\psi|U_{x_i}^{\dagger} = U_{x_i}V|0\rangle\langle 0|V^{\dagger}U_{x_i}^{\dagger}$$

Quantum kernel for samples x_i , x_j :

$$K(x_i, x_j) = \mathrm{tr}\left(\Phi^{\dagger}(x_i)\Phi(x_j)\right) = \left|\langle 0|\, V^{\dagger}U_{x_i}^{\dagger}U_{x_j}V|0\rangle\right|^2$$

Quantum circuit:



$|\psi\rangle$: fiducial state

K : covariant quantum kernel

A framework for covariant quantum kernels

Qiskit

The example we just saw was extremely simple, just a single qubit.

Is there any reason to believe that this could be a promising framework?

Yes!

The kernel based on the DLOG problem can be viewed as a covariant quantum kernel.

This kernel leads to a superpolynomial speedup over all classical learners on this problem.

Group: $\mathbb{Z}_p^*$ (integers with multiplication mod p) for a prime number p

Liu et al. arxiv:2010.02174

Two key points about covariant quantum kernels

Qiskit

We know the framework of covariant quantum kernels is promising: DLOG kernel

The fiducial state:

1. May depend on the learning problem and dataset

2. Plays an important role in the performance of the classifier

Using insights from training an SVM to find a good fiducial state

Bounding the generalization error , ϵ

$$\epsilon \leq \tilde{O}(\text{primal} /\ m)$$

$$\leq \tilde{O}(\min \text{primal} /\ m)$$

$$\leq \tilde{O}(\max \text{dual} /\ m) \ = \tilde{O}(\max_\alpha F(\alpha) /\ m)$$

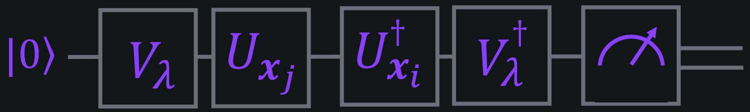
$$F(\alpha) = \sum_{i=1}^{m} \alpha_i - \underbrace{\frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \overbrace{K(x_i, x_j)}^{K_\lambda(x_i, x_j)}}_{\text{weighted kernel alignment}} \quad \text{(wolfe dual)}$$

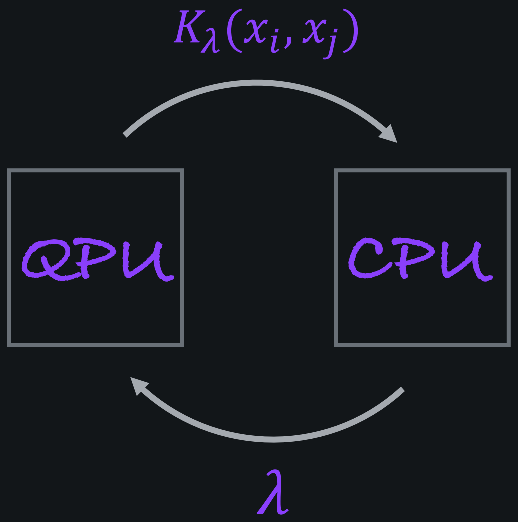
$$\leq \tilde{O}(\underbrace{\min_\lambda \max_\alpha F(\alpha, \lambda)}_{\text{Solve this!}} /\ m)$$

Vapnik, *The Nature of Statistical Learning Theory*
Cristianini et al. Adv. Neur. Info. Proc. Sys. 14 (2001)
Cortes et al. JMLR 13, 795 (2012)
Bullins et al. arxiv:1710.10230
Shawe-Taylor, Cristianini. IEEE Trans. Info. Theor. 48, 2721 (2002)

Kernel alignment for quantum kernels:
Covariant quantum kernels

Parametrize the fiducial state: $|\psi\rangle \rightarrow |\psi_\lambda\rangle = V_\lambda|0\rangle$



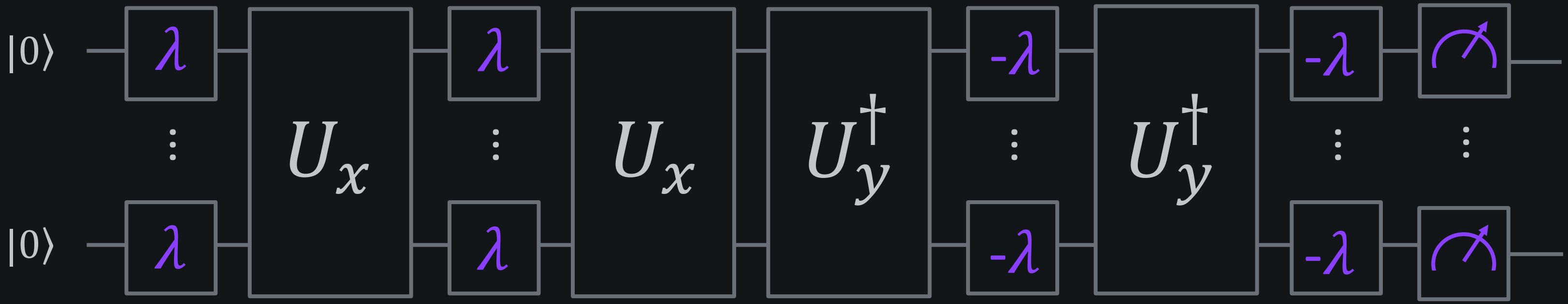


Glick et al. arxiv:2105.03406

Kernel alignment for quantum kernels:

Example of a Forrelation kernel

$K_\lambda(x, y)$:



Summary of the main takeaways

- Variational quantum classifiers are just linear methods in quantum feature space. Therefore, we can use *quantum kernels*.

- Quantum kernels can only be expected to do better than classical kernels if they are *hard to estimate classically*. This is a necessary but not sufficient condition.

- Designing quantum kernels that *exploit structure in data* is a promising step toward practical applications.

- Quantum kernels can be optimized on a given dataset with *kernel alignment*.