

Robust quantum classifier with minimal overhead

Daniel K. Park^{*¶}, Carsten Blank^{†¶}, and Francesco Petruccione^{‡§}

^{*}Sungkyunkwan University Advanced Institute of Nanotechnology, Suwon, Korea

Email: dkp.quantum@gmail.com

[†]Data Cybernetics, Landsberg, Germany

Email: blank@data-cybernetics.com

[‡]School of Chemistry and Physics, University of KwaZulu-Natal, Durban, South Africa

[§]National Institute for Theoretical Physics (NITheP), KwaZulu-Natal, South Africa

Email: petruccione@ukzn.ac.za

Abstract—To witness quantum advantages in practical settings, substantial efforts are required not only at the hardware level but also on theoretical research to reduce the computational cost of a given protocol. Quantum computation has the potential to significantly enhance existing classical machine learning methods, and several quantum algorithms for binary classification based on the kernel method have been proposed. These algorithms rely on estimating an expectation value, which in turn requires an expensive quantum data encoding procedure to be repeated many times. In this work, we calculate explicitly the number of repetition necessary for acquiring a fixed success probability and show that the Hadamard-test and the swap-test circuits achieve the optimal variance in terms of the quantum circuit parameters. The variance, and hence the number of repetition, can be further reduced only via optimization over data-related parameters. We also show that the kernel-based binary classification can be performed with a single-qubit measurement regardless of the number and the dimension of the data. Finally, we show that for a number of relevant noise models the classification can be performed reliably without quantum error correction. Our findings are useful for designing quantum classification experiments under limited resources, which is the common challenge in the noisy intermediate-scale quantum era.

I. INTRODUCTION

The theory of fault-tolerant quantum computing promises tremendous opportunities with clear quantum advantages for certain computational tasks [1], [2], [3], [4]. However, the development of full-fledged quantum computing hardware remains a long-term prospect. On the road to building fault-tolerant quantum computers, noisy intermediate-scale quantum (NISQ) computers are expected to be available in the near future [5], [6], [7]. These quantum devices can execute only a limited size of quantum circuits reliably due to noise, but can surpass the capabilities of classical digital computers. An important issue in the NISQ era is to find problems and applications for which the NISQ technology can provide practical quantum advantage. Addressing this issue involves not only experimental efforts at the hardware level but also requires theoretical research to develop quantum algorithms while taking imperfections into account.

Quantum computing also has the potential to drastically improve machine learning tasks [8], [9], [10], [11], [12].

Quantum advantages in machine learning are expected naturally since quantum computers can reduce the computational cost exponentially for solving certain basic linear algebra problems [13], [4] that often appear as basic subroutines in machine learning tasks. Moreover, quantum computers can achieve exponential compression of data [14], [15], [16]. Full comprehension of which machine learning problems can be solved more efficiently with quantum algorithms remains as an important open problem.

A family of machine learning tasks for which quantum techniques are expected to outperform existing methods is the kernel-based classification [17], [18], [19], [20], [21], [22], which is a fundamental problem in pattern recognition. A common advantage of quantum computing utilized in these works is the ability to manipulate exponentially large quantum Hilbert space efficiently and evaluate the kernel function much faster than classical computers. The quantum machine learning algorithms presented in Refs. [20], [21] are of particular interest in the NISQ era since they do not require expensive subroutines for solving the convex optimization problem of support vector machine. Furthermore, they are flexible in terms of the quantum data encoding method; either amplitude encoding [10] or quantum feature mapping [19] can be used. In these algorithms, a classification score is evaluated by measuring an expectation value of certain observables, and hence repeating the same experiment multiple times is inevitable for a reliable statistics. However, due to the measurement postulate of quantum mechanics and the no-cloning theorem, a same input state must be created for every execution of the algorithm [23]. Since computational cost for preparing an arbitrary quantum input state can be substantial depending on the structure of data to be encoded, reducing the number of repetition is essential [23], especially for NISQ computing.

In this work, we explicitly calculate the number of repetition necessary for estimating the classification score with a fixed precision. We discuss various generalizations that can be made to existing kernel-based quantum classifiers and calculate the variance of the estimator to show that the classifiers presented in Refs. [20], [21] are indeed optimal with respect to the number of repetition. We also show that the kernel-based binary classification can be performed with a single-qubit measurement regardless of the number and the dimension of

[¶]Equal contribution
©2021 IEEE

the data. This is particularly useful for the systems in which the measurement error is worse than the gate error. Furthermore, we show that the binary classifications can be performed reliably under certain noise models that are important in quantum information science without having to employ expensive quantum error correction; the classification will succeed by increasing the number of repetition quadratically with respect to a relevant error rate.

II. PRELIMINARIES

A. Binary classification

Classification is a fundamental problem in machine learning. The goal of L -class classification is to infer the class label of an unseen data point $\tilde{x} \in \mathbb{C}^N$, given a labelled data set

$$\mathcal{D} = \{(x_1, y_1), \dots, (x_M, y_M)\} \subset \mathbb{C}^N \times \{0, 1, \dots, L-1\}.$$

Although the data is real-valued in usual machine learning tasks, we allow complex-valued data without loss of generality by noting that various quantum data encoding schemes utilize the quantum Hilbert space. A famous example of encoding classical information as a quantum state is the amplitude encoding which represents a classical vector $\mathbf{x}_j = (x_{1j}, \dots, x_{Nj})^T \in \mathbb{C}^N$ as a quantum state in the following form,

$$|\mathbf{x}_j\rangle := \frac{1}{\|\mathbf{x}_j\|} \sum_{i=1}^N x_{ij} |i\rangle, \quad (1)$$

using $\lceil \log_2(N) \rceil$ qubits. Similarly, a set of M data points $\mathbf{x}_1, \dots, \mathbf{x}_M$ can be encoded in $\lceil \log_2(NM) \rceil$ qubits as

$$\frac{1}{\sqrt{\sum_{ij} |x_{ij}|^2}} \sum_{j=1}^M \sum_{i=1}^N x_{ij} |i\rangle \otimes |j\rangle. \quad (2)$$

Hereinafter we focus on binary classification (i.e. $L = 2$) like majority of works on quantum kernel-based classifiers since a multi-class classification can be constructed with binary classifiers by one versus all or one versus one scheme. In addition, we will omit the Kronecker product symbol (\otimes) whenever the meaning is clear (e.g. $|i\rangle \otimes |j\rangle = |ij\rangle$).

B. Review of kernel-based quantum classifiers

This work focuses on extending and improving the kernel-based quantum classifiers presented in Refs. [20], [21], since they are more suitable for NISQ computing as mentioned in the introduction. These algorithms are referred to as *Hadamard-test classifier* (HTC) and *swap-test classifier* (STC), respectively. Construction of these algorithms can be broken into two parts: preparation of a quantum state that encodes data in a specific form and expectation value measurement. These are explained in more detail below.

The Hadamard-test classifier encodes the dataset \mathcal{D} in a quantum state as

$$|\psi_h\rangle = \frac{1}{\sqrt{2}} \sum_{j=1}^M \sqrt{a_j} (|0\rangle |\mathbf{x}_j\rangle + |1\rangle |\tilde{\mathbf{x}}\rangle) |y_j\rangle |j\rangle, \quad (3)$$

where $|\mathbf{x}_j\rangle$ and $|\tilde{\mathbf{x}}\rangle$ encodes classical training and test data vectors via an encoding of choice, and the label $y_j \in \{0, 1\}$ is represented by the computational basis of the label qubit. Without loss of generality all inputs \mathbf{x}_j and $\tilde{\mathbf{x}}$ are assumed to be normalized and have unit length. The subscript h indicates that the state is for the Hadamard-test classifier. In Ref. [20], the weights are uniform, i.e. $a_j = 1/M \forall j$, but it can be left as a variable to be optimized, similar to the treatment in support vector machines [22]. The measurement scheme utilizes a Hadamard-test, which applies a Hadamard gate on the ancilla qubit to interfere training and test data states. Finally, by measuring an expectation value of a two-qubit observable $\sigma_z^{(a)} \sigma_z^{(l)}$ on the ancilla qubit and the label qubit, one obtains

$$\langle \psi_h | H^{(a)} \sigma_z^{(a)} \sigma_z^{(l)} H^{(a)} | \psi_h \rangle = \sum_{j=1}^M (-1)^{y_j} a_j \text{Re} \langle \mathbf{x}_j | \tilde{\mathbf{x}} \rangle, \quad (4)$$

where the superscripts a and l indicate that the corresponding operator is acting on the ancilla qubit and the label qubit, respectively. From this equation, one can see that the kernel function in HTC is $k(\mathbf{x}_j, \tilde{\mathbf{x}}) = \text{Re} \langle \mathbf{x}_j | \tilde{\mathbf{x}} \rangle$, and Eq. (4) defines the classification score, which we denote by f , in an HTC. The HTC assigns a new label to the test data as

$$\tilde{y} = \frac{1}{2} \left(1 - \text{sgn} \left[\sum_{j=1}^M (-1)^{y_j} a_j \text{Re} \langle \mathbf{x}_j | \tilde{\mathbf{x}} \rangle \right] \right).$$

The HTC considers only a real part of the quantum state overlap. In order to fully exploit the ability of quantum computers to efficiently manipulate quantum states in the Hilbert space, it is desirable to construct a kernel that takes both real and imaginary parts of the quantum state into account. This motivated the birth of the swap-test classifier (STC). To see how the kernel function looks like in the STC, it is useful to express the initial state with the density matrix formalism. The initial state can be written as

$$\rho_s = |0\rangle\langle 0| \otimes \sum_{j=1}^M \left(a_j (\tilde{\rho} \otimes \rho_j) \right)^{\otimes k} \otimes |y_j\rangle\langle y_j| \otimes |j\rangle\langle j|, \quad (5)$$

where the subscript s indicates that the density matrix is for the swap-test classifier. The next step of the classifier is to apply the swap test

$$T_s = H^{(a)} \cdot \prod_{i=1}^k \text{sw}_{\text{AP}}(t_i, d_i | a = 1) \cdot H^{(a)}, \quad (6)$$

where $H^{(a)}$ represents a Hadamard gate applied to the ancilla qubit and $\text{sw}_{\text{AP}}(t_i, d_i | a)$ represents a controlled-swap gate that exchanges an i th copy of test (t_i) and training (d_i) data if the ancilla qubit state is a . Finally, the expectation value measurement of a two-qubit observable $\sigma_z^{(a)} \sigma_z^{(l)}$ results in

$$\langle \sigma_z^{(a)} \sigma_z^{(l)} \rangle = \sum_{j=1}^M (-1)^{y_j} a_j \text{Tr}(\tilde{\rho} \rho_j)^k. \quad (7)$$

Therefore, the kernel function in the STC is $k(\mathbf{x}_j, \tilde{\mathbf{x}}) = \text{Tr}(\tilde{\rho}\rho_j)^k$. When the training and test data are given as a pure state, i.e. $\tilde{\rho} = |\tilde{\mathbf{x}}\rangle\langle\tilde{\mathbf{x}}|$ and $\rho_j = |\mathbf{x}_j\rangle\langle\mathbf{x}_j|$, then the kernel is reduced to $k(\mathbf{x}_j, \tilde{\mathbf{x}}) = |\langle\tilde{\mathbf{x}}|\mathbf{x}_j\rangle|^{2k}$, which is the k th power of the quantum state fidelity. The classification score in an STC is given by Eq. (7). The STC assigns a new label to the test data as

$$\tilde{y} = \frac{1}{2} \left(1 - \text{sgn} \left[\sum_{j=1}^M (-1)^{y_j} a_j \text{Tr}(\tilde{\rho}\rho_j)^k \right] \right).$$

III. OPTIMIZATION

We seek to minimize the computational resource overhead in the HTC and STC caused by the number of repetition necessary for estimating the classification score. The number of repetition can be calculated from the Chebyshev inequality

$$\Pr[|\mu - \langle\mathcal{M}\rangle| \geq \epsilon] \leq \sigma^2 / (k\epsilon^2) \quad (8)$$

where μ is an average value obtained from k trials, $\langle\mathcal{M}\rangle$ is the expectation value to be estimated, and $\sigma^2 = \Delta\mathcal{M} = \langle\mathcal{M}^2\rangle - \langle\mathcal{M}\rangle^2$ is the variance. Since the classifier only uses the sign of the expectation value, we can choose the precision to be

$$\epsilon = \langle\mathcal{M}\rangle / c \quad (9)$$

for some constant $c > 1$. Then the desired number of repetition goes as

$$k = O(\sigma^2 / \langle\mathcal{M}\rangle^2) \quad (10)$$

to bound the error probability to a fixed constant.

In the following, we examine how to reduce k by changing the circuit design of the classifier.

A. General form of the classification score

Suppose one measures an expectation value of an observable $\mathcal{M}_\lambda = \sigma_z^{(a)} \otimes A_\lambda^{(l)}$ for some arbitrary Hermitian operator A_λ . In addition, suppose the classical label y_j is encoded in a logical label state $|\bar{y}_j\rangle$. Then it is straight-forward to see that the expectation value measured in either HTC or STC classification protocols becomes

$$\langle\mathcal{M}_\lambda\rangle = \sum_{j=1}^M a_j \langle\bar{y}_j| A_\lambda |\bar{y}_j\rangle k(\mathbf{x}_j, \tilde{\mathbf{x}}), \quad (11)$$

where the kernel function depends on whether the HTC or the STC is performed. From the above equation, one can see that the classification contrast can be increased by choosing A and $|\bar{y}_j\rangle$ such that $\langle\bar{y}_j| A_\lambda |\bar{y}_j\rangle \in \{-\lambda, \lambda\}$ for some $\lambda > 1$.

Now we choose an observable and the label qubit state based on the following rule:

$$\sum_{i=1}^{\lambda} \sigma_z^{(i)} |l\rangle^{\otimes\lambda} = (-1)^l \lambda |l\rangle^{\otimes\lambda}, \quad l \in \{0, 1\}, \quad (12)$$

where the superscript i indicates that the Pauli operator is acting on the i th qubit. By setting $A_\lambda = \sum_{i=1}^{\lambda} \sigma_z^{(i)}$ and $|\bar{y}_j\rangle =$

$|0\rangle^{\otimes\lambda}$ if \mathbf{x}_j is labelled 0 and $|\bar{y}_j\rangle = |1\rangle^{\otimes\lambda}$ if \mathbf{x}_j is labelled 1, one obtains

$$\langle\mathcal{M}_\lambda\rangle = \lambda \sum_{m=1}^M a_j (-1)^{y_j} k(\mathbf{x}_j, \tilde{\mathbf{x}}). \quad (13)$$

Therefore, the expectation value to be estimated is scaled by a factor of λ by increasing the number of label qubits by the same factor.

B. Variance calculation

We assume that we have an eigenstate decomposition (which we can find) of the observable \mathcal{M}_λ such that

$$\mathcal{M}_\lambda = m_{0,\bar{0}} P_{0,\bar{0}} + m_{0,\bar{1}} P_{0,\bar{1}} + m_{1,\bar{0}} P_{1,\bar{0}} + m_{1,\bar{1}} P_{1,\bar{1}} \quad (14)$$

where $m_{i,\bar{j}} = (-1)^{i+\bar{j}} \lambda$ and projections $P_{i,\bar{j}} = |i, \bar{j}\rangle\langle i, \bar{j}|$ for $i, \bar{j} = 0, 1$. Gathering the statistics we compute

$$\begin{aligned} \langle\mathcal{M}_\lambda\rangle &= m_{0,\bar{0}} p(0, \bar{0}) + m_{0,\bar{1}} p(0, \bar{1}) \\ &\quad + m_{1,\bar{0}} p(1, \bar{0}) + m_{1,\bar{1}} p(1, \bar{1}) \\ &= \lambda (p(0, \bar{0}) - p(0, \bar{1}) - p(1, \bar{0}) + p(1, \bar{1})) \end{aligned} \quad (15)$$

as well as

$$\begin{aligned} \langle\mathcal{M}_\lambda^2\rangle &= m_{0,\bar{0}}^2 p(0, \bar{0}) + m_{0,\bar{1}}^2 p(0, \bar{1}) \\ &\quad + m_{1,\bar{0}}^2 p(1, \bar{0}) + m_{1,\bar{1}}^2 p(1, \bar{1}) \\ &= \lambda^2 (p(0, \bar{0}) + p(0, \bar{1}) + p(1, \bar{0}) + p(1, \bar{1})) \\ &= \lambda^2 \end{aligned} \quad (16)$$

with $p(i, \bar{j}) = \text{Tr}(P_{i,\bar{j}} |\Psi_f\rangle\langle\Psi_f|)$.

By identifying Eq. (15) with Eq. (13) we see that the variance is given by

$$\begin{aligned} \sigma^2 &= \langle\mathcal{M}_\lambda^2\rangle - \langle\mathcal{M}_\lambda\rangle^2 \\ &= \lambda^2 \left(1 - \left(\sum_{j=1}^M a_j (-1)^{y_j} k(\mathbf{x}_j, \tilde{\mathbf{x}}) \right)^2 \right) \end{aligned} \quad (17)$$

which means that at the boundary, i.e. when both classes are equally far away for a test datum, the variance is maximal.

Since both σ^2 and $\langle\mathcal{M}_\lambda\rangle^2$ goes as λ^2 , the number of repetition is independent of λ according to Eq. (10). In other words, the number of repetition cannot be reduced by increasing the number of label qubits. Therefore, the best strategy with respect to the measurement of label register is to use as small number of qubits as possible. This is simply done by using a single qubit to encode the label information (i.e. $\lambda = 1$), and setting $A_\lambda^{(l)} = \sigma_z$.

C. Skewness

The skewness is a measure how much the distribution is leaning towards one side from the mean. It is the third standardized moment and with respect to the mean $\langle\mathcal{M}\rangle$ and can be expressed as

$$\mu_3^c = \frac{\langle\mathcal{M}_\lambda^3\rangle - 3\langle\mathcal{M}_\lambda\rangle\Delta\mathcal{M}_\lambda - \langle\mathcal{M}_\lambda\rangle^3}{\Delta\mathcal{M}_\lambda^{3/2}}. \quad (18)$$

For an arbitrary λ , the third moment is given by

$$\begin{aligned}\langle \mathcal{M}_\lambda^3 \rangle &= m_{0,\bar{0}}^3 p(0, \bar{0}) + m_{0,\bar{1}}^3 p(0, \bar{1}) \\ &\quad + m_{1,\bar{0}}^3 p(1, \bar{0}) + m_{1,\bar{1}}^3 p(1, \bar{1}) \\ &= \lambda^3 (p(0, \bar{0}) - p(0, \bar{1}) - p(1, \bar{0}) + p(1, \bar{1})) \\ &= \lambda^3 \left(\sum_{j=1}^M a_j (-1)^{y_j} k(\mathbf{x}_j, \tilde{\mathbf{x}}) \right).\end{aligned}\quad (19)$$

By combining the two equations above and using a notation $f = \sum_{j=1}^M a_j (-1)^{y_j} k(\mathbf{x}_j, \tilde{\mathbf{x}})$ we obtain

$$\begin{aligned}\mu_3^c &= \frac{\langle \mathcal{M}_\lambda^3 \rangle - 3 \langle \mathcal{M}_\lambda \rangle \Delta \mathcal{M} - \langle \mathcal{M}_\lambda \rangle^3}{\Delta \mathcal{M}_\lambda^{3/2}} \\ &= \frac{\lambda^3 f - 3\lambda f \lambda^2 (1 - f^2) - \lambda^3 f^3}{\lambda^3 (1 - f^2)^{3/2}} \\ &= -\frac{2f}{\sqrt{1 - f^2}}.\end{aligned}\quad (20)$$

Two remarks are to be made here. First, the skewness does not depend on λ . This result again favors the use of one-qubit register for encoding the label information. The second is that the skewness is negative with respect to the classification score f . This indicates that the probability density favors values whose absolute value is larger than $|\langle \mathcal{M}_\lambda \rangle|$. The skewness is zero only if $f = 0$ with which the algorithm cannot classify the test data; the skewness will always be non-zero when the classifier can make a decision. The necessity of the asymmetry opens an interesting research direction towards designing quantum classifiers based on the mode (e.g. majority vote) instead of the mean.

D. Generalization of the interfering circuit

Further generalization can be made in both HTC and STC by using arbitrary single qubit rotation gates instead of the Hadamard gates for creating superposition in the beginning and interference at the end on the ancilla qubit. In this case, Eq. (3) becomes

$$|\psi_h\rangle = \sum_{j=1}^M \sqrt{a_j} \left(\cos \frac{\theta_0}{2} |0\rangle |\mathbf{x}_j\rangle + \sin \frac{\theta_0}{2} e^{i\phi} |1\rangle |\tilde{\mathbf{x}}\rangle \right) |y_j\rangle |j\rangle.\quad (21)$$

The Hadamard gate at the end of the circuit for interfering two subspaces spanned by the computational basis of the ancilla qubit is also replaced with an arbitrary rotation around the y -axis of the Bloch sphere $R_y(\theta_1) = \cos(\theta_1/2)I - i \sin(\theta_1/2)\sigma_y$. This gate can be followed by an arbitrary rotation around the z -axis, but since we are also measuring in the σ_z -basis, we can neglect it as it does not alter the measurement result. For a depiction of this setup, confer to figure 2. Then the two-qubit expectation value measurement gives

$$\begin{aligned}\langle \sigma_z^{(a)} \sigma_z^{(l)} \rangle &= \sum_{j=1}^M a_j (-1)^{y_j} \left(\cos(\theta_0) \cos(\theta_1) - \sin(\theta_0) \sin(\theta_1) \right) \\ &\quad \times (\cos(\phi) \text{Re} \langle \mathbf{x}_j | \tilde{\mathbf{x}} \rangle - \sin(\phi) \text{Im} \langle \mathbf{x}_j | \tilde{\mathbf{x}} \rangle).\end{aligned}\quad (22)$$

This equation shows that the imaginary part of the state overlap $\langle \mathbf{x}_j | \tilde{\mathbf{x}} \rangle$ can also contribute to the classification result, unlike in the original HTC. We leave the use of this imaginary part for classification as an interesting future work, and only focus on the case where $\text{Im} \langle \mathbf{x}_j | \tilde{\mathbf{x}} \rangle = 0$ to mimic the original HTC.

Similarly, the general form of the expectation value for STC can be calculated. Without assuming $\text{Im} \langle \mathbf{x}_j | \tilde{\mathbf{x}} \rangle = 0$, it can be written as

$$\begin{aligned}\langle \sigma_z^{(a)} \sigma_z^{(l)} \rangle &= \sum_{j=1}^M a_j (-1)^{y_j} \left(\cos(\theta_0) \cos(\theta_1) - \sin(\theta_0) \sin(\theta_1) \right) \\ &\quad \times \cos(\phi) |\langle \mathbf{x}_j | \tilde{\mathbf{x}} \rangle|^2.\end{aligned}\quad (23)$$

Now, it is straight-forward to see that $\langle (\sigma_z^{(a)} \sigma_z^{(l)})^2 \rangle = \langle I \rangle = 1$. Thus the variance is simply $1 - \langle (\sigma_z^{(a)} \sigma_z^{(l)}) \rangle^2$, as expected from Eq. (17). Since we aim to minimize the variance, we want to maximize $f(\theta_0, \theta_1, \phi) = \langle \sigma_z^{(a)} \sigma_z^{(l)} \rangle^2$; this is our objective function. For simplicity, we can consider a special cases where $\theta_0 = \pi/2$. In this case, the objective function becomes

$$\left| \sum_{j=1}^M a_j (-1)^{y_j} \sin(\theta_1) \cos(\phi) k(\mathbf{x}_j, \tilde{\mathbf{x}}) \right|^2$$

for both HTC and STC. This is maximized with respect to θ_1 and ϕ if $|\sin(\theta_1) \cos(\phi)| = 1$. One solution is $\sin(\theta_1) = 1$ and $\cos(\phi) = -1$, which is equivalent to setting the final gate on the ancilla qubit to be the Hadamard gate.

More rigorous analysis can be done by calculating the first- and second-order partial derivatives of the objective function with respect to θ_0 , θ_1 , and ϕ . After going through some laborious calculations, it can be shown that the critical point simultaneously satisfying

$$\frac{\partial f(\theta_0, \theta_1, \phi)}{\partial \theta_0} = 0, \quad \frac{\partial f(\theta_0, \theta_1, \phi)}{\partial \theta_1} = 0, \quad \frac{\partial f(\theta_0, \theta_1, \phi)}{\partial \phi} = 0$$

is a solution to $\cos(\theta_0^*) = \cos(\theta_1^*) = \sin(\phi^*) = 0$. This condition is satisfied in the original HTC and STC in which the creation of superposition and interference on the ancilla qubit is solely done by Hadamard gates. However, to be sure that the critical point really gives the local maximum of f , the second derivative test needs to be performed. To keep further discussions simple, we assume $\sin(\phi) = 0$ so that the maximization is to be done only with respect to θ_0 and θ_1 . In this case, the above critical point satisfies

$$\frac{\partial^2 f(\theta_0, \theta_1)}{\partial \theta_0^2} = \frac{\partial^2 f(\theta_0, \theta_1)}{\partial \theta_1^2} \leq 0.\quad (24)$$

The second derivatives are zero when $\sum_j a_j (-1)^{y_j} k(\mathbf{x}_j, \tilde{\mathbf{x}}) = 0$, in which case the classification is not possible anyways. Now, if

$$\frac{\partial^2 f(\theta_0, \theta_1)}{\partial \theta_0^2} \frac{\partial^2 f(\theta_0, \theta_1)}{\partial \theta_1^2} - \left(\frac{\partial^2 f(\theta_0, \theta_1)}{\partial \theta_0 \partial \theta_1} \right)^2 > 0,\quad (25)$$

we can assure that the critical point above indeed yields the local maximum. This condition is satisfied if $|\sum_j a_j (-1)^{y_j} k(\mathbf{x}_j, \tilde{\mathbf{x}})| > |\sum_j a_j (-1)^{y_j}|$.

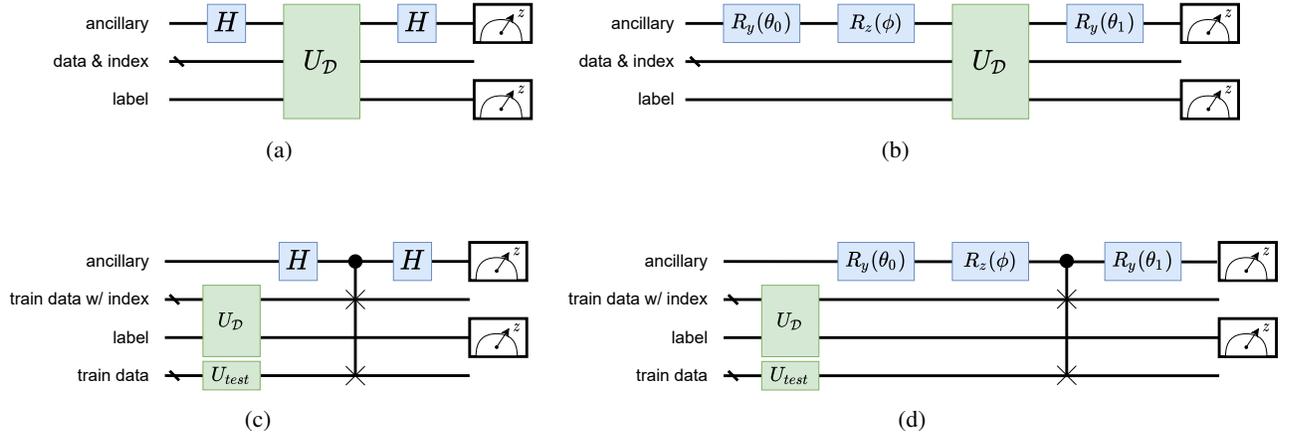


Fig. 1: **Classification Circuits.** (a) The Hadamard-test classifier depicted with the state preparation and the equal superposition of the Hadamard. (b) This is altered to now have rotational gates with arbitrary angles instead of the two Hadamard gates. (c) and (d) are analogue but with the swap-test classifier.

Having only two free parameters, another critical point is given when $\sin(\theta_0) = \sin(\theta_1) = 0$. This condition also satisfies Eq. (24), and hence a candidate for the maximizing solution. The condition in Eq. (25) is satisfied if $|\sum_j a_j(-1)^{y_j} k(\mathbf{x}_j, \tilde{\mathbf{x}})| < |\sum_j a_j(-1)^{y_j}|$. However, if $\sin(\theta_0) = \sin(\theta_1) = 0$ is used, then the classification outcome becomes $\langle \sigma_z^{(a)} \sigma_z^{(l)} \rangle = \sum_j a_j(-1)^{y_j}$, and hence does not construct the classification based on given dataset. Thus we conclude that this is not a suitable solution. Therefore, we argue that the best suited solution is given by the condition $\cos(\theta_0^*) = \cos(\theta_1^*) = \sin(\phi^*) = 0$, which is satisfied with the use of Hadamard gates. The variance is given as Eq. (17), and this is the optimal value if $|\sum_j a_j(-1)^{y_j} k(\mathbf{x}_j, \tilde{\mathbf{x}})| > |\sum_j a_j(-1)^{y_j}|$.

IV. ROBUSTNESS TO NOISE

Imperfections are unavoidable in the implementation of quantum algorithms. The theory of quantum error correction and fault-tolerance guarantees that the quantum computation can be performed reliably under noise at the cost of increasing the quantum resources (e.g. qubits and gates) as long as the physical error rate is below certain threshold value. The resource overhead is larger if the physical error rate is larger, even if it is below the fault-tolerance threshold. Typical NISQ devices will not have enough number of qubits to perform an arbitrary fault-tolerant quantum computation. Therefore, minimizing the resource overhead for fighting against noise is of critical importance. In the following, we show that under certain relevant noise models, the quantum binary classification can be performed reliably without quantum error correction while increasing the number of repetition only quadratically with respect to an effective error rate.

A. Classifier with a single-qubit measurement

Both HTC and STC rely on measuring an expectation value of a two-qubit observable $\sigma_z^{(a)} \sigma_z^{(l)}$. The same outcome can be obtained by measuring an expectation value of a single-qubit

observable after an additional two-qubit gate. We introduce a notation for the controlled bit-flip (CNOT) operation as $I^{(i)} \otimes |0\rangle\langle 0|^{(j)} + \sigma_x^{(i)} \otimes |1\rangle\langle 1|^{(j)} = cX(i|j)$ to mean that the bit-flip operation is applied to a target qubit i if the state of the control qubit j is $|1\rangle$. Then the following property holds:

$$cX(i|j)^\dagger (\sigma_z^{(i)} \otimes I^{(j)}) cX(i|j) = \sigma_z^{(i)} \otimes \sigma_z^{(j)}.$$

With the above it is trivial to see that

$$\begin{aligned} \langle \sigma_z^{(a)} \sigma_z^{(l)} \rangle &= \text{Tr}(\rho_f \sigma_z^{(a)} \sigma_z^{(l)}) \\ &= \text{Tr}(\rho_f cX(a|l)^\dagger \sigma_z^{(a)} cX(a|l)) \\ &= \text{Tr}(cX(a|l) \rho_f cX(a|l)^\dagger \sigma_z^{(a)}) \\ &= \text{Tr}(\tilde{\rho}_f \sigma_z^{(a)}), \end{aligned} \quad (26)$$

where ρ_f is the density matrix representation of the final state of either the HTC or the STC quantum circuit, and $\tilde{\rho}_f$ is the modified final state in these circuits obtained by applying the CNOT operation on the ancilla qubit with the label qubit as the control. Therefore, the HTC and STC algorithms can be performed with a single-qubit measurement with addition of a CNOT gate. The quantum circuit with this modification is depicted in Fig. 2a

As a simple demonstration, we took the toy example from Ref. [21] and simulated the quantum binary classification protocol with the one-qubit measurement scheme. The example data set consists of two training data and one test data as

$$\begin{aligned} |\mathbf{x}_1\rangle &= \frac{i}{\sqrt{2}} |0\rangle + \frac{1}{\sqrt{2}} |1\rangle, \quad y_1 = 0, \\ |\mathbf{x}_2\rangle &= \frac{i}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle, \quad y_2 = 1, \\ |\tilde{\mathbf{x}}\rangle &= \cos \frac{\theta}{2} |0\rangle - i \sin \frac{\theta}{2} |1\rangle. \end{aligned} \quad (27)$$

Simulations are carried out in two sets. First, we assume an ideal implementation without noise to verify the idea. Then

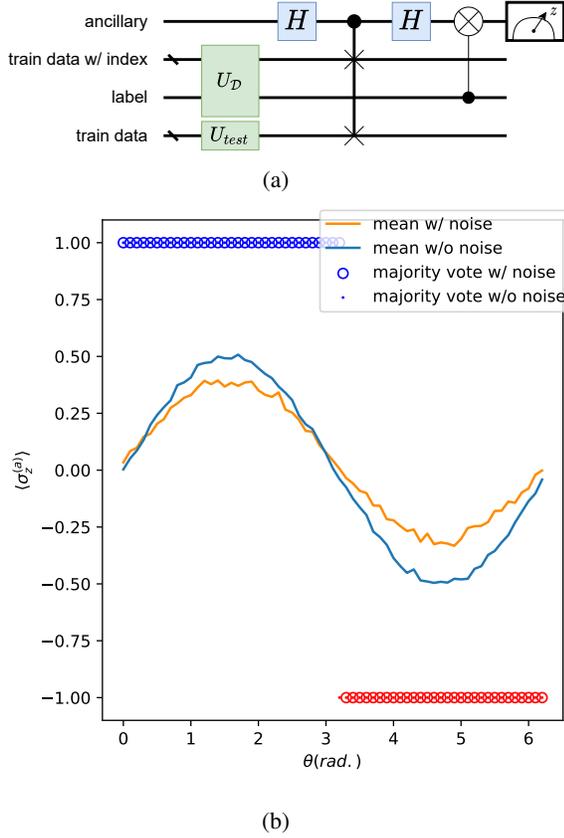


Fig. 2: **One Qubit Measurement.** (a) By adding a CNOT gate with the label qubit as control and the ancillary qubit as target, one can apply a one qubit z-measurement only. (b) Simulations of the one-qubit-measurement scheme simulated with and without noise. We compare the mean and the majority voting as the classification rule, and show that the majority voting is more favorable in practice. These results are obtained from 8192 shots and the noisy simulations are with respect to an IBM quantum devices called `ibmq_rome`.

we performed simulations with a realistic noise model. The noisy simulations implemented quantum circuits with gate decomposition (i.e. transpilation) given by a five-qubit IBM quantum device available called `ibmq_rome` and its noise model. We also tested the use of majority vote for classification in place of the expectation value as suggested in Sec. III-C. Each quantum circuit is repeated 8192 times to gather the measurement statistics. The simulation results are shown in Fig. 2b.

B. Effect of noise

As a simple noise model, let us consider the depolarizing error that acts only on the ancilla qubit at the end of the quantum circuit before the measurement. The effect of the single-qubit depolarizing channel acting on the final state $\tilde{\rho}_f$ can be studied with the Kraus representation as

$$\mathcal{E}_d(\tilde{\rho}_f) = \sum_{k=1}^4 E_k^{(a)} \tilde{\rho}_f \left(E_k^{(a)}\right)^\dagger, \quad (28)$$

where the set of Kraus operators are given as

$$E = \left\{ \sqrt{1 - \frac{3p}{4}} I, \sqrt{\frac{p}{4}} \sigma_x, \sqrt{\frac{p}{4}} \sigma_y, \sqrt{\frac{p}{4}} \sigma_z \right\}$$

with a depolarizing error rate $0 \leq p \leq 1$. Then one can calculate the expectation value under this noise model as

$$\begin{aligned} \text{Tr}\left(\mathcal{E}_d(\tilde{\rho}_f)\sigma_z^{(a)}\right) &= \text{Tr}\left(\sum_{k=1}^4 E_k^{(a)} \tilde{\rho}_f \left(E_k^{(a)}\right)^\dagger \sigma_z^{(a)}\right) \\ &= \sum_{k=1}^4 \text{Tr}\left(\tilde{\rho}_f E_k^{(a)} \sigma_z^{(a)} E_k^{(a)}\right) \\ &= (1-p) \text{Tr}\left(\tilde{\rho}_f \sigma_z^{(a)}\right), \end{aligned} \quad (29)$$

where the last line is obtained by using the commutation relations of the Pauli operators. This equation shows that under the single-qubit depolarizing noise model, the expectation value to be measured in HTC and STC are reduced by a factor of $1-p$, but importantly, it will not change the sign of the expectation value. This has an imperative consequence; the single-qubit depolarizing noise acting on the ancilla qubit of the final state can be easily mitigated since the classification only uses the sign of the measurement outcome. The same level of the classification accuracy as that of the noiseless case can be achieved by repeating the measurement $O(1/(1-p)^2)$ times. This result is deduced again from the Chebyshev inequality in Eq. (8).

Now we extend the error model to be an arbitrary Pauli channel acting on all qubits, which is the most basic noise channel ubiquitous in quantum information science [24]. That is,

$$\mathcal{E}_p = \sum_{j=1}^{4^n} c_j P_j \tilde{\rho}_f P_j^\dagger, \quad (30)$$

where $P_j \in \{I, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}$ is an element in the set of n -qubit Pauli operators, and $c_j \in \mathbb{R}$, $c_j \geq 0$, and $\sum_{j=1}^{4^n} c_j = 1$. Note that P_j can be written as $P_j = P_j^{(1)} \otimes P_j^{(2)} \otimes \dots \otimes P_j^{(n)}$. Under this noise model, the expectation value of interest becomes

$$\begin{aligned} \text{Tr}\left(\mathcal{E}_p(\tilde{\rho}_f)\sigma_z^{(a)}\right) &= \text{Tr}\left(\sum_j^{4^n} c_j P_j \tilde{\rho}_f P_j \sigma_z^{(a)}\right) \\ &= \sum_j^{4^n} c_j \text{Tr}\left(\tilde{\rho}_f P_j^{(a)} \sigma_z^{(a)} P_j^{(a)} \otimes P_j^{(2)} P_j^{(2)} \otimes \dots \otimes P_j^{(n)} P_j^{(n)}\right) \\ &= \sum_j^{4^n} c_j \text{Tr}\left(\tilde{\rho}_f P_j^{(a)} \sigma_z^{(a)} P_j^{(a)}\right) \\ &= (C_I + C_{\sigma_z} - C_{\sigma_x} - C_{\sigma_y}) \text{Tr}\left(\tilde{\rho}_f \sigma_z^{(a)}\right), \end{aligned} \quad (31)$$

where $C_i = \sum_{j|P_j^{(a)}=i} c_j$. Therefore, an arbitrary Pauli noise channel acting on the final state of the classifier only scales the expectation value by a constant factor determined by the noisy process. As long as this factor is positive, the classifier

can be made robust to noise by repeating the measurement $O(1/(C_I + C_{\sigma_z} - C_{\sigma_x} - C_{\sigma_y})^2)$ times.

V. CONCLUSION

Kernel-based quantum classification algorithms are promising candidates for NISQ applications. These algorithms are based on measuring an expectation value of an observable, which requires an experiment to be repeated many times to provide a good estimate. In this work, we investigated the possibilities and strategies to reduce this extra resource overhead. We showed explicitly that the variance in Hadamard-test and the swap-test circuits are optimal in terms of the quantum circuit design. In addition to the variance analysis, we calculated skewness and argued that the mode can be a statistically better quantity to measure for classification. We also explicitly calculated the number of repetition necessary to succeed a quantum classification protocol under relevant noise models and showed that it can be performed correctly without quantum error correction. The findings presented in this work are useful for designing NISQ experiments to perform quantum classification protocols with minimal overhead and robustness to noise.

ACKNOWLEDGMENT

This research is supported by the National Research Foundation of Korea (Grant No. 2019R111A1A01050161), Quantum Computing Development Program (Grant No. 2019M3E4A1080227), and the South African Research Chair Initiative of the Department of Science and Technology and the National Research Foundation.

REFERENCES

- [1] S. Lloyd, "Universal quantum simulators," *Science*, vol. 273, no. 5278, pp. 1073–1078, 1996. [Online]. Available: <http://www.jstor.org/stable/2899535>
- [2] C. Zalka, "Simulating quantum systems on a quantum computer," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1969, pp. 313–322, 1998.
- [3] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM review*, vol. 41, no. 2, pp. 303–332, 1999.
- [4] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," *Phys. Rev. Lett.*, vol. 103, p. 150502, Oct 2009. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.103.150502>
- [5] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018. [Online]. Available: <https://doi.org/10.22331/q-2018-08-06-79>
- [6] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum supremacy using a programmable

- superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019. [Online]. Available: <https://doi.org/10.1038/s41586-019-1666-5>
- [7] K. Wright, K. M. Beck, S. Debnath, J. M. Amini, Y. Nam, N. Grzesiak, J. S. Chen, N. C. Pimenti, M. Chmielewski, C. Collins, K. M. Hudek, J. Mizrahi, J. D. Wong-Campos, S. Allen, J. Apisdorf, P. Solomon, M. Williams, A. M. Ducore, A. Blinov, S. M. Kreikemeier, V. Chaplin, M. Keesan, C. Monroe, and J. Kim, "Benchmarking an 11-qubit quantum computer," *Nature Communications*, vol. 10, no. 1, p. 5464, 2019. [Online]. Available: <https://doi.org/10.1038/s41467-019-13534-2>
- [8] P. Wittek, *Quantum Machine Learning: What Quantum Computing Means to Data Mining*. Boston: Academic Press, 2014.
- [9] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, pp. 195 EP –, 09 2017. [Online]. Available: <https://doi.org/10.1038/nature23474>
- [10] M. Schuld and F. Petruccione, *Supervised Learning with Quantum Computers*, ser. Quantum Science and Technology. Springer International Publishing, 2018.
- [11] V. Dunjko and H. J. Briegel, "Machine learning & artificial intelligence in the quantum domain: a review of recent progress," *Reports on Progress in Physics*, vol. 81, no. 7, p. 074001, jun 2018. [Online]. Available: <https://doi.org/10.1088/1361-6633/aab406>
- [12] C. Ciliberto, M. Herbster, A. D. Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, "Quantum machine learning: a classical perspective," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 474, no. 2209, p. 20170551, 2018. [Online]. Available: <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2017.0551>
- [13] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. New York, NY, USA: Cambridge University Press, 2011.
- [14] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum random access memory," *Phys. Rev. Lett.*, vol. 100, p. 160501, Apr 2008. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.100.160501>
- [15] D. K. Park, F. Petruccione, and J.-K. K. Rhee, "Circuit-based quantum random access memory for classical data," *Scientific Reports*, vol. 9, no. 1, p. 3949, 2019. [Online]. Available: <https://doi.org/10.1038/s41598-019-40439-3>
- [16] T. M. L. Veras, I. C. S. De Araujo, K. D. Park, and A. J. Dasilva, "Circuit-based quantum random access memory for classical data with continuous amplitudes," *IEEE Transactions on Computers*, pp. 1–1, 2020.
- [17] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, 2019. [Online]. Available: <https://doi.org/10.1038/s41586-019-0980-2>
- [18] P. Rebentrost, M. Mohseni, and S. Lloyd, "Quantum support vector machine for big data classification," *Phys. Rev. Lett.*, vol. 113, p. 130503, Sep 2014. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.113.130503>
- [19] M. Schuld and N. Killoran, "Quantum machine learning in feature hilbert spaces," *Phys. Rev. Lett.*, vol. 122, p. 040504, Feb 2019. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.122.040504>
- [20] M. Schuld, M. Fingerhuth, and F. Petruccione, "Implementing a distance-based classifier with a quantum interference circuit," *EPL (Europhysics Letters)*, vol. 119, no. 6, p. 60002, 2017. [Online]. Available: <http://stacks.iop.org/0295-5075/119/i=6/a=60002>
- [21] C. Blank, D. K. Park, J.-K. K. Rhee, and F. Petruccione, "Quantum classifier with tailored quantum kernel," *npj Quantum Information*, vol. 6, no. 1, p. 41, May 2020. [Online]. Available: <https://doi.org/10.1038/s41534-020-0272-6>
- [22] D. K. Park, C. Blank, and F. Petruccione, "The theory of the quantum kernel-based binary classifier," *Physics Letters A*, vol. 384, no. 21, p. 126422, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0375960120302541>
- [23] D. K. Park, I. Sinayskiy, M. Fingerhuth, F. Petruccione, and J.-K. K. Rhee, "Parallel quantum trajectories via forking for sampling without redundancy," *New Journal of Physics*, vol. 21, no. 8, p. 083024, aug 2019. [Online]. Available: <https://doi.org/10.1088/1367-2630/ab35fb>
- [24] B. M. Terhal, "Quantum error correction for quantum memories," *Rev. Mod. Phys.*, vol. 87, pp. 307–346, Apr 2015. [Online]. Available: <https://link.aps.org/doi/10.1103/RevModPhys.87.307>