

UNIVERSAL DISCRIMINATIVE QUANTUM NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Quantum mechanics fundamentally forbids deterministic discrimination of quantum states and processes. However, the ability to optimally distinguish various classes of quantum data is an important primitive in quantum information science. In this work, we trained near-term quantum circuits to classify data represented by quantum states using the Adam stochastic optimization algorithm. This is achieved by iterative interactions of a classical device with a quantum processor to discover the parameters of an unknown non-unitary quantum circuit. This circuit learns to simulate the unknown structure of a generalized quantum measurement, or positive-operator valued measure (POVM), that is required to optimally distinguish possible distributions of quantum inputs. Notably we used universal circuit topologies, with a theoretically motivated circuit design which guaranteed that our circuits can perform arbitrary input-output mappings. Our numerical simulations showed that quantum circuits could be trained to discriminate among various pure and mixed quantum states, exhibiting a trade-off between minimizing erroneous and inconclusive outcomes with comparable performance to theoretically optimal POVMs. We trained the circuit on different classes of quantum data and evaluated the generalization error on unseen quantum data. This generalization power hence distinguishes our work from standard circuit optimization and provides an example of quantum machine learning for a task that has inherently no classical analogue.

1 INTRODUCTION

Quantum computation are known to provide speedups in several applications over classical computation. Besides the famous Shor’s algorithm for prime number factorization, quantum computers can also produce statistical patterns that are hard for a classical computer to produce. This raises the possibility that quantum computers can also recognize patterns that are hard to recognize for classical computers, or, in general, that quantum computers can help solve classical machine learning problems more efficiently. Recently, this joint field of quantum computation and machine learning has received a considerable amount of attention. Using the circuit model of computation, several quantum algorithms have been designed that provide quadratic to exponential speedups on classical data (Biamonte et al. (2017); Ciliberto et al. (2018)).

A related and interesting area is to develop novel machine learning methods on quantum data. In general, any collection of quantum states which carry certain meaningful information can be considered as quantum data. To motivate this direction, we want to emphasize that using quantum states as a storage medium for information processing have shown to provide advantages in several aspects. Researches have shown that by coupling a quantum state with another target system, one could obtain information about the target system through this quantum state but with a better sensitivity. For example, quantum meteorology allows for a quadratic improvement over classical methods in terms of the scaling of statistical errors, i.e. the scaling of the standard deviations in estimated values in repeated measurements. Another example is quantum sensing, which provide much higher sensitivity in tasks like target detections in microwaves, i.e. quantum radar (Barzanjeh et al. (2015)), and in general, sensing electric or magnetic fields (Degen et al. (2017)). Further, it is proposed that by storing certain information of a picture on a quantum state, pattern recognition can be achieved with

much fewer damage done to this picture if it is very sensitive to the exposure of light (Schaller & Schützhold (2006)).

Certain types of quantum data are special in that they are inherently quantum mechanical. They could be produced from materials fabricated in certain exotic quantum state or could be the result of quantum information processing procedures. Here we could conjecture the inherent advantage of using quantum computers to recognize and classify these natural quantum data. For example, topological materials made in the exotic topological phase have non-classical electronic properties and are promising materials to build fault-tolerant quantum computers (Qi & Zhang (2011); Karzig et al. (2017)). However, predicting the phase of topological materials has been a very challenging problem, but recently it was shown that quantum neural networks could be used to recognize the phase of a quantum state (Cong et al. (2018)). In addition, the promised security of quantum communication protocols and a surge of ideas in quantum communication networks (Kimble (2008); Ren et al. (2017)) further stimulates the research into the topic of quantum data.

In this work, we explored the general problem of classifying quantum data. This problem could be considered as an extension of the well-researched field *quantum state discrimination*, which identifies a quantum state among a set of possible candidates of which one has complete knowledge *a priori*. Quantum state discrimination already found applications in many fields such as quantum key distribution (Bennett (1992a)) and cryptography (Barnett & Croke (2009); Bergou (2007)). A key challenge for the discrimination of quantum states is that a deterministic discrimination is impossible when the complex vectors representing the candidate states are not orthogonal, i.e. when their overlaps are non-zero. Therefore, the central task of quantum state discrimination is to find the optimal discriminative measurement that one should perform for those states.

However, it is not possible to directly apply quantum state discrimination to classify the quantum data. Firstly, it is inappropriate to assume that one possesses the complete knowledge of the data *a priori*, which are often only samples generated from the data collection process. Also, even with all the quantum data available, obtaining the complete knowledge of them (i.e. the density matrices describing them) are very expensive. In addition, *quantum state discrimination* often fails to give the optimal discriminative measurement in an analytically closed form, unless the quantum states are already orthogonal or possess certain symmetry properties (Barnett & Croke (2009)). In cases it fails, one may use numerical optimization algorithms to find the optimal measurement. However, the exponential increase of the dimension of the density matrices renders the numerical optimization inefficient.

Due to the limitations of quantum state discrimination, it is then a natural question whether we can use a quantum computer to help with the optimization procedure. Here we utilize a hybrid quantum-classical approach to directly learn the design of a shallow quantum circuit for the classification of quantum data. In this hybrid scheme, a classical computer is used to interactively change a quantum circuit and optimize the output of the quantum computer. Recently a string of works focusing on this hybrid scheme for training quantum circuits for a wide range of tasks have been proposed (Banchi et al. (2016); Wan et al. (2017); Innocenti et al. (2018); Romero et al. (2017); Mitarai et al. (2018); Farhi & Neven (2018); Verdon et al. (2017); Li & Benjamin (2017); Grant et al. (2018); Schuld et al. (2018)), with major players in quantum computing hardware releasing software toolkits for hybrid quantum-classical models. This makes the hybrid scheme a promising area in quantum computation.

Our approach described here is novel in two perspectives. First, we used a quantum circuit ansatz that is motivated for implementation on near-term devices. This ansatz is shallow in depths but possess the full power of quantum computation, i.e. it can perform any unitary transformation allowed in quantum mechanics. It comprises gates from a universal gate set consisting of C-NOT and single-qubit gates, motivated by the fact that their implementations are known for the currently most used experimental architectures. It is nearly optimal in terms of the number of C-NOT gates it has, which is also an important feature for an implementation on near-term devices. Second, unlike previous works on quantum state discrimination, we focused on the generalization ability of our circuit, i.e., we trained the circuit on a specific range of the parameters with the goal of maximizing its generalization performance, hence considering a learning framework. This distinguishes our work from the pure optimization problem for the state discrimination task, which is optimizing the circuit to distinguish only a concrete set of states. We showed that this universal quantum circuit can be trained as a discriminator for classification of non-orthogonal quantum data sampled from

various different probability distributions. Our discriminator can achieve a near-zero error rate by producing inconclusive signals.

2 DATASET AND OUR APPROACH

In this work, we propose a novel approach to train a universal quantum circuit to classify quantum data, which are stored in qubits. In this section, we first explain the mathematical descriptions of quantum states, operations and measurements on these states. We then specify the quantum data we used in this work for classification. Next, we outline the approach we took to optimize a universal quantum circuit which was used to classify the quantum data. We defer the detailed decomposition of this quantum circuit in Appendix A.

Mathematical descriptions Quantum states are described by density matrices, which are Hermitian, positive semi-definite complex matrices of unit trace, and commonly denoted by the symbol ρ . For the quantum states which stores certain useful information, we may assume that they could be parameterized by certain parameters which follow certain probability distributions specific to the information they carried. For classification, we are normally presented with an unknown quantum state, belonging to families of quantum states, each described mathematically by:

$$\rho_i(a_i), a_i \sim \alpha_i, \quad (2.1)$$

where i is the label for different families, $\rho_i(a_i)$ is the density matrix describing a quantum datum in the family i , parametrized by a_i . The parameters a_i are assumed to follow the probability distribution α_i . When training a quantum circuit for classification, a_i are samples drawn according to the distribution α_i .

Transformations on quantum states are described by a complex unitary matrix U , which transforms a quantum state ρ according to the rule:

$$\rho \rightarrow U\rho U^\dagger. \quad (2.2)$$

A measurement on the quantum state ρ is described by a set of matrices $\{M_j\}$, which are Hermitian, positive semi-definite and sum to the identity matrix, i.e. $\sum_j M_j = \mathbb{1}$. Here j labels the possible measurement outcomes, and the probability p_j that this particular measurement outcome is detected is,

$$p_j = \text{Tr}(M_j\rho), \quad (2.3)$$

where Tr is the trace operator. Such a collection of matrices M_j is commonly called a *positive-operator valued measure* (POVM). A common example of POVM is a *projection-valued measure* (PVM). In the case of a PVM, each M_j is a projector into some linear subspace and different M_j are orthogonal to each other, i.e. $M_j M_i = \delta_{ij} M_j$. Any POVM could be realized by a quantum circuit, which consists of a series of unitary matrices (transformations) and measurements in the computational basis. Conversely, a quantum circuit which is parameterized by some parameters and has measurements could also represent many different POVMs. There exists a quantum circuit which could represent any POVM with a fixed number of possible measurement outcomes. Such a circuit is called a universal discriminator in this paper, and a specific one which we used is discussed in Appendix A.

Our Dataset For this work, we restricted our attention to the classification of two families of quantum states stored in 2 qubits. Our first family consists of pure states, parametrized by a real number $a \in [0, 1]$:

$$\psi_1(a) = \left(\sqrt{1-a^2}, 0, a, 0 \right), \rho_1 = |\psi_1(a)\rangle\langle\psi_1(a)|. \quad (2.4)$$

The second family consists of mixed states $\rho_2(b)$ where $b \in [0, 1]$. Specifically,

$$\psi_{2/3} = \left(0, \pm\sqrt{1-b^2}, b, 0\right), \rho_2 = \frac{1}{2} |\psi_2\rangle\langle\psi_2| + \frac{1}{2} |\psi_3\rangle\langle\psi_3|. \quad (2.5)$$

The overlap between ψ_1 and $\psi_{2/3}$ is ab , and hence our data in two families are non-orthogonal. In the case when a is a fixed value, and $b = \frac{1}{\sqrt{2}}$, the maximal success rate for unambiguously discriminating between ρ_1 and ρ_2 has been theoretically studied, and an experimental demonstration is available (Mohseni et al. (2004)). The specific distribution we have tested in our experiments are summarized in Table 1. To generate the data for the training, validation, and testing of our circuits, we randomly sampled points from the corresponding distributions.

	Family 1 ($\rho_1(a)$)	Family 2 ($\rho_2(b)$)
Case 1	$a \in [0, 1]$	$b \in [0, 1]$
Case 2	$a \approx 0.25$	$b \approx \frac{1}{\sqrt{2}}$
Case 3	$a \approx 0.25$	$b \in [0, 1]$
Case 4	$a \in [0, 1]$	$b \approx \frac{1}{\sqrt{2}}$

Table 1: A summary of different test cases we classified in this work. Here $a(b) \in [0, 1]$ represents that $a(b)$ follows a uniform distribution in $[0, 1]$. $a(b) \approx 0.25(\frac{1}{\sqrt{2}})$ represents that $a(b)$ follows a normal distribution with mean $0.25(\frac{1}{\sqrt{2}})$, standard deviation 0.05 , and which is truncated in $[0, 1]$.

Approach. Overall, there are two major strategies to cope with our inherent inability to perform deterministic discrimination of quantum states: (a) *Minimum-error discrimination*: In this strategy, the task is to minimize the probability that the inevitable errors occur in the classification. (b) *Unambiguous discrimination*: In this strategy, the discriminator has one more output prediction than the number of classes it tries to classify: an inconclusive outcome. The task is to eliminate the error rate of the discriminator while minimizing the probability of this inevitable inconclusive outcome. A pure unambiguous discrimination with strictly zero error rate is not guaranteed to be possible for arbitrary quantum data. From the perspective of numerical optimization, one needs to allow for some small but non-zero errors to happen.

In this work, we used the machine learning approach to train a universal quantum circuit capable of giving any quantum measurements with four possible measurement outcomes $m_{i_2 i_1}$, where $i_1, i_2 \in \{0, 1\}$ are the measurement outcomes of the first and the second qubit respectively. The parameterization of this circuit is discussed in Appendix A. By assuming that input $\rho_1(a)$ produces the output m_{00} or m_{10} , input ρ_2 produces the output m_{01} , and assuming that m_{11} is the inconclusive output, this circuit acts as a discriminator. Therefore, we could trivially define various probabilities (success probability P_{suc} , error probability P_{err} , and inconclusive probability P_{inc}) with respect to an input training data with known class label. For example, when ρ_1 is the input, the probability of detecting m_{01} is the P_{err} , the probability of detecting m_{11} is the P_{inc} . In this work, we performed experiments on simulated quantum computers, where these probabilities were available. We note that on real quantum computers, these probabilities can be estimated through repeated measurements on replicated data, up to some precision.

To train the circuit, we used a heuristically motivated loss function defined in Eq. 2.6, which is the averaged absolute difference between the desired probabilities and the measured probabilities. It contains hyperparameters α_{err} and α_{inc} to balance between the erroneous outcomes and the inconclusive outcomes:

$$\begin{aligned}
 J = & \sum_i \frac{1}{|S_i|} \sum_{a_i \in S_i} |P_{\text{suc}}(\rho_i(a_i)) - 1| \\
 & + \alpha_{\text{err}} \sum_i \frac{1}{|S_i|} \sum_{a_i \in S_i} |P_{\text{err}}(\rho_i(a_i)) - 0| \\
 & + \alpha_{\text{inc}} \sum_i \frac{1}{|S_i|} \sum_{a_i \in S_i} |P_{\text{inc}}(\rho_i(a_i)) - 0|. \quad (2.6)
 \end{aligned}$$

Here we assumed that for each family of quantum states, we had been supplied with a set S_i of training samples, where each class was labeled by i . Then $|S_i|$ denotes the number of samples in the training sets S_i , α_{err} is the penalty for making errors, and α_{inc} is the penalty for giving inconclusive outcomes. $P_{\text{suc}}(\psi)/P_{\text{err}}(\psi)/P_{\text{inc}}(\psi)$ are the probabilities of giving a correct/erroneous/inconclusive measurement outcome for the specific input quantum data ρ_i . This loss function measures the performance of our quantum circuit as a minimal-error discriminator (when $\alpha_{\text{err}} < \alpha_{\text{inc}}$) or as an unambiguous discriminator (when $\alpha_{\text{err}} > \alpha_{\text{inc}}$).

To train this circuit, we used the Adam optimization algorithm (Kingma & Ba (2014)), and the gradients were calculated by the forward difference formula.

For our specific problem of classifying ρ_1 and ρ_2 as defined in Eq. 2.4 and Eq. 2.5, we defined an extra set of success/erroneous/inconclusive rates in Eq. 2.7 to summarize and compare the performance of different trainings:

$$\begin{aligned} P_s &= \frac{1}{3}P_s(\rho_1)_{\text{avg}} + \frac{2}{3}P_s(\rho_2)_{\text{avg}} \\ &= \frac{1}{3}P_s(\psi_1)_{\text{avg}} + \frac{1}{3}P_s(\psi_2)_{\text{avg}} + \frac{1}{3}P_s(\psi_3)_{\text{avg}} \end{aligned} \quad (2.7)$$

where s stands for *suc* (successful), *err* (erroneous) or *inc* (inconclusive). The subscript *avg* means that the probabilities are calculated as the value averaged in all samples available in either the training set, or the test set (but not both). The choice of weights ($\frac{1}{3}$ and $\frac{2}{3}$) in the Eq. 2.7 was made to be consistent with the paper Mohseni et al. (2004).

3 THEORETICAL ANALYSIS

Here we describe a theoretical result to which we will compare our numerical results. First, we mention a general result. Assume we have a family of quantum data $\rho(a)$, each one is parameterized by a and occur with a probability $P(a)$. Assume in addition that we have a quantum measurement described by a POVM with elements $\{\Pi_i\}_{i \in \mathbb{N}}$, where i labels different measurement outcomes. Then, the probability of detecting measurement outcome i , averaged over any possible quantum data $\rho(a)$, is,

$$\begin{aligned} \int \text{Tr}(\Pi_i \rho(a)) P(a) da &= \text{Tr} \left[\int \Pi_i \rho(a) P(a) da \right] = \text{Tr} \left[\Pi_i \int \rho(a) P(a) da \right] \\ &= \text{Tr}[\Pi_i \rho], \end{aligned} \quad (3.1)$$

where $\rho \equiv \int \rho(a) P(a) da$, and the integration of the matrix is done in an element-wise fashion. Therefore, if $\text{Tr}(\Pi_i \rho) = 0$ for some i , then $\int_D \text{Tr}(\Pi_i \rho(a)) P(a) = 0$ for any subset D with non-zero measure in the whole parameter space of a . This is due to the fact that $\text{Tr}[\Pi_i \rho(a)] P(a) \geq 0$ for any parameter a .

Applied to the problem of unambiguous discrimination, it is obvious that the problem of unambiguously discriminating $\rho_1 = \int_a \rho_1(a) P_1(a) da$ and $\rho_2 = \int_a \rho_2(a) P_2(a) db$, is equivalent to the problem of unambiguously discriminating the family $\rho_1(a), \forall a$, from the family $\rho_2(b), \forall b$, where $P_1(a)/P_2(b)$ is the probability of occurrence of $\rho_1(a)/\rho_2(b)$. That is, if $\{\Pi_1, \Pi_2, \Pi_?\}$ is a POVM that unambiguously classifies all members of the two families $\rho_1(a)$ and $\rho_2(b)$, for all possible parameters, i.e. $\Pi_?$ corresponds to the inconclusive outcome with,

$$\begin{aligned} \text{Tr}(\Pi_2 \rho_1(a)) &= 0, \forall a, \\ \text{Tr}(\Pi_1 \rho_2(b)) &= 0, \forall b, \end{aligned}$$

then $\text{Tr}(\Pi_1 \rho_2) = \text{Tr}(\Pi_2 \rho_1) = 0$, and vice versa. Using this formalism, we theoretically analyzed the different cases we described in Table 1 based on the works of Raynal et al. (2003) and Barnett & Croke (2009), and the results are displayed in Table 2. Note that these are average case success probabilities.

	Family 1 (ρ_1)	Family 2 (ρ_2)	P_{suc}
Case 1	$a \in [0, 1]$	$b \in [0, 1]$	0.67
Case 2	$a \approx 0.25$	$b \approx \frac{1}{\sqrt{2}}$	0.64
Case 3	$a \approx 0.25$	$b \in [0, 1]$	0.76
Case 4	$a \in [0, 1]$	$b \approx \frac{1}{\sqrt{2}}$	0.55

Table 2: A summary of maximal success rate when the error rate is exactly 0 for different test cases we have classified in this work. They were calculated using Eq. 2.7 and obtained theoretically using methods mentioned in Section 3.

4 NUMERICAL RESULTS ON SIMULATED QUANTUM COMPUTERS

In this work, we aim to train a universal discriminator to discriminate different families of quantum data. Here, we presented the results of training the universal discriminator to discriminate different distributions summarized in Table 1 on a simulated quantum computer. To balance between eliminating the error rate (P_{err}) while minimizing the inconclusive rate (P_{inc}), we used a specific training strategy described in the following. We first prioritized a smaller inconclusive rate by starting with a zero penalty for erroneous outcomes ($\alpha_{\text{inc}} > \alpha_{\text{err}} = 0$), and then increased the α_{err} in a step-wise manner until a certain objective error rate is achieved. Similar optimization procedures have been used in the context of variational auto-encoders both in classical machine learning (Sønderby et al. (2016)), and in quantum machine learning applications (Rocchetto et al. (2017)). Using this scheme, we trained our circuit to unambiguously discriminate the two families of quantum states and observed a convergence towards the theoretical success rates obtained in Section 3 with increasing amount of data used to train the discriminator. Notably, we did not observe any signs of overfitting despite the varying size of the training dataset (Figure 1 a).

4.1 TRADE-OFF BETWEEN THE ERROR RATE AND THE INCONCLUSIVE RATE

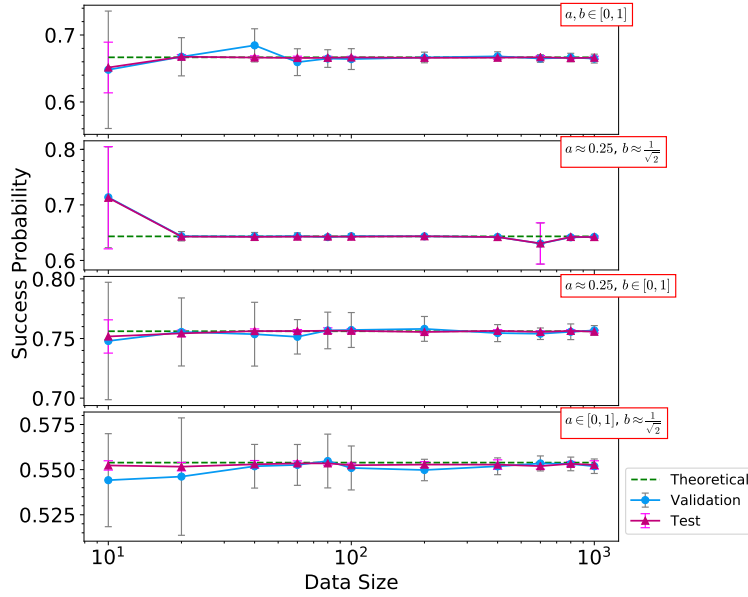
In another test, we showed that our model is able to obtain a much higher success rate (P_{suc}) if we allow a slightly higher error rate compared with previous results during the trainings. This hints at a trade-off between the error rate (P_{err}) and the inconclusive rate (P_{inc}) which can be utilized in real-world applications.

Specifically, for the dataset “Case 4” in Table 1 we fixed the two penalties, α_{err} and α_{inc} , during the training and observed a gradual transition from unambiguous-like classification (characterized by a near-zero error probability) to minimal-error-like classification (characterized by the near-zero inconclusiveness) when we used varying penalties (Figure 2 a-c) throughout the different trainings. Allowing a small error rate resulted in a much higher success rate, which has not been predicted theoretically. We note that introducing the penalty terms α_{err} and α_{inc} made the training process also more stable (Figure 2 a). Therefore, the hyperparameters α_{err} and α_{inc} acted as a form of regularization and could be adjusted to give a higher success probability or a lower inconclusiveness rate for the final model (Figure 2).

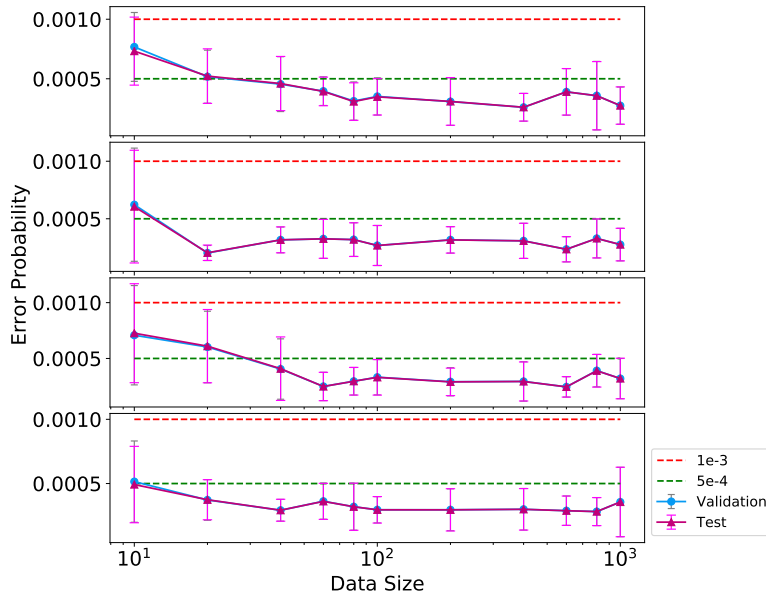
For all the datasets in Table 1, we achieved a much higher success rate than the theoretical case of an exactly zero error rate, with the required error rate tolerance smaller than 0.01 (Figure 3).

5 LEARNING CONVERGENCE FROM ENSEMBLE MEASUREMENTS

Although we trained our circuit using probabilities which are not available on an actual quantum device, we can estimate these probabilities from repeated measurements on this quantum device. We found that when using probabilities estimated from repeated measurements of quantum states, the training process still converged in all cases. In practice, the noise introduced by the inferred probabilities could be effectively countered by increasing the number of repeated measurements, using a lowered error rate, and adjusting the step size for gradient calculation using the forward difference formula. Overall, a combination of a step size of 0.01, 10^5 repeated measurements, and a learning rate of 10^{-3} well approximated the result obtained using exact probabilities. Therefore, our study here is feasible for actual quantum devices. We leave open the actual implementation as future projects. Further discussions are deferred in Appendix B.



(a) The success rates converged to the theoretical result (the green dashed line in the figure) as the training data size was increased. The performance on the testing (purple line) and the validation dataset also converged with the growing data size. The abnormal deviation at the point with data size 600 for the case $a \approx 0.25$ and $b \approx 1/\sqrt{2}$ was due to unsuccessful optimizations.



(b) Most of the error rate for the corresponding trainings are well below 0.0005, indicating an unambiguous discrimination was achieved.

Figure 1: Unambiguous classification of non-orthogonal quantum data sampled from different probability distributions. The data were averaged over 10 repeated trials starting with random initializations and the bars indicate the standard deviations. The training, validation and test data were sampled from the corresponding distributions. The data size indicates the size of the training and the validation set. The test data was fixed at a size of 10^4 for each family and each distribution.

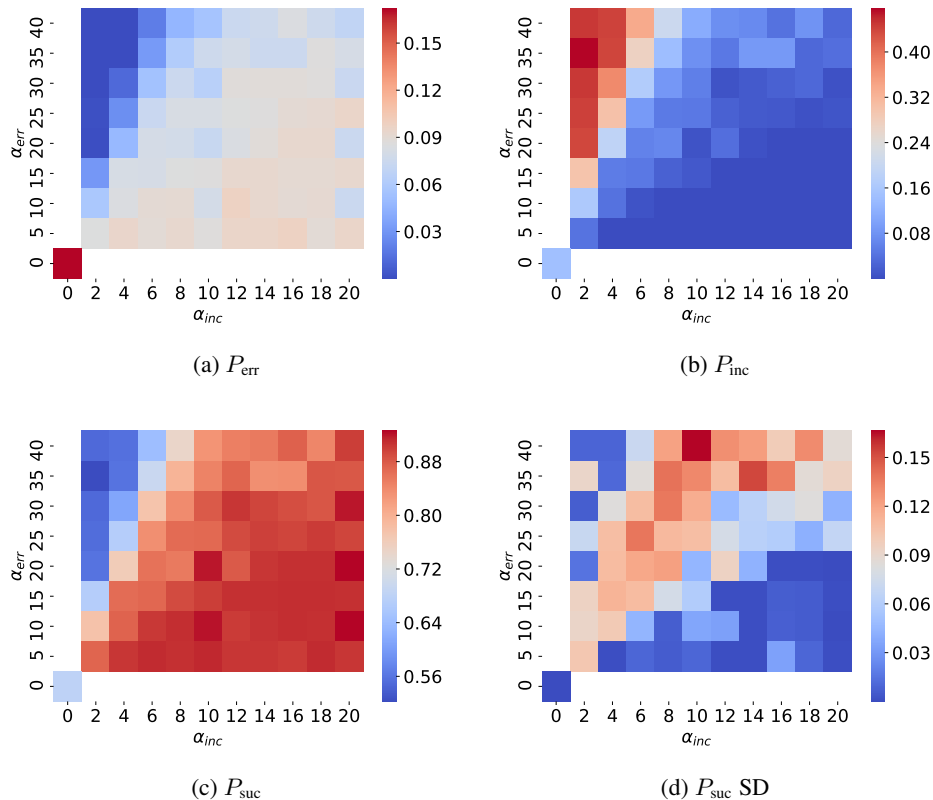
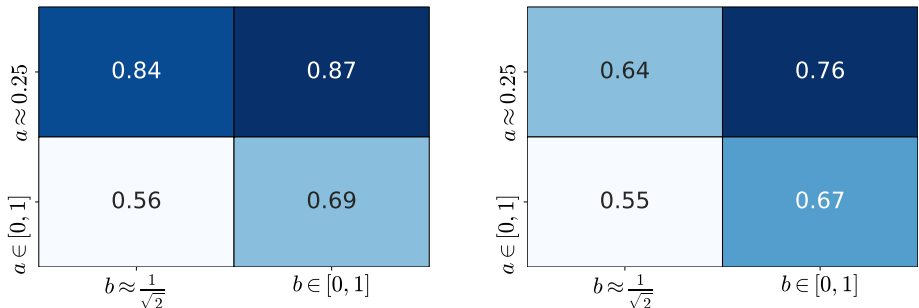


Figure 2: **With different penalties we observed a trade-off between the error rate and the inconclusive rate.** Compared with the point $\alpha_{err} = \alpha_{inc} = 0$ (bottom left corner), the added penalties improved the success probability or the inconclusiveness respectively. **(a)-(c)**: The gradual transition from the unambiguous classification (near-zero error rate, top left corner) to a minimal error classification (near-zero inconclusiveness, bottom right corner) with changes in the error penalty α_{err} and the inconclusiveness penalty α_{inc} . We observed that the gain in the success rate was around 0.32 when we made a sacrifice of only 0.1 in the error rate. The data was tested on $a \in [0, 1]$, and averaged over 50 repeated trails with random initializations. **(d)**: Standard deviation for P_{suc} . With an increasing standard deviation (closer to the diagonal line), the result became increasingly unstable when the two penalties (α_{err} and α_{inc}) are closer in value. The standard deviations for P_{err} and P_{inc} showed the same pattern as for P_{suc} (not shown).



(a) Success rate for trained circuit for classification of data (b) The theoretical success rate when exact 0 error rate is achieved.

Figure 3: **Unambiguous discrimination of data sampled from different probability distributions with higher success rate.** (a) Trained quantum circuits were capable of classifying quantum data which was sampled from a variety of different mixed probability distributions for $\rho_1(a)$ and $\rho_2(b)$. The classification was done in an unambiguous manner (with error rate < 0.01). (b) For comparison, we included here the theoretical result mentioned in Table 2.

6 CONCLUSIONS

We have developed a universal quantum circuit learning approach for the classification of quantum data. In particular, we have designed a theoretically motivated loss function and used the stochastic optimization algorithm Adam in a quantum-classical hybrid scheme to train a circuit to perform quantum state discrimination. This training process generalized well for the discrimination task on new data, i.e., states from the parameter range which have not been seen during the training process. This in particular distinguishes our work from previous results on quantum circuit learning, in particular the very recent study in Fanizza et al. (2018), which only optimizes circuits for specific inputs. Note that this prior work hence does not consider the generalization ability and hence does not treat the actual learning problem, which aims at optimization as well as generalization.

In our work, we observed a trade-off between the error rates and the inconclusive rates when we penalized them differently in the loss function. Although this experiment was done on simulated quantum computers where exact measurement probabilities are available, we showed that this optimization could be experimentally performed with repeated measurements of the quantum states. Finally, we note that the recent quantum methods for estimating the analytical gradient via variations in the unitaries (Mitarai et al. (2018)) can be directly applied to training our circuits and therefore one can perform the optimization efficiently on near-term quantum devices.

With the progress on technologies for preservation and transportation of quantum states, we can expect many applications of a trained discriminative quantum circuits introduced here. Quantum state discrimination by itself plays a key role in quantum information processing protocols and is used in quantum cryptography (Bennett (1992b)), quantum cloning (Duan & Guo (1998)), quantum state separation, and entanglement concentration (Chefles (2000)). Our work can provide improvements on these traditional areas by producing a classifier that is resilient to the statistical noise found in the actual communication. For example, we can consider an improved version of the B92 quantum key distribution protocol (Bennett (1992a)) by including the noise-induced randomness in its two quantum keys and classify them with our discriminative circuit. Furthermore, we can consider training a discriminative quantum circuit used to construct quantum repeaters and state purification units within quantum communication networks. The training can take quantum data that have noise specific to the communication networks and therefore produces a discriminator that can recognize and filter those noise to provide better performance. Our discriminator can also be used to verify the output of other generative models, such as the quantum version of Boltzmann machines (Amin et al. (2018)), or generative artificial neural networks (Goodfellow et al. (2014); Lloyd & Weedbrook (2018)).

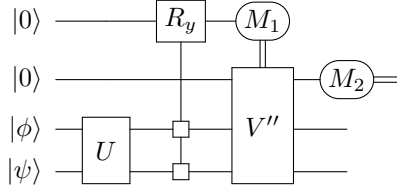
REFERENCES

- Mohammad H. Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchytskyy, and Roger Melko. Quantum boltzmann machine. *Physical Review X*, 8(2), may 2018. doi: 10.1103/physrevx.8.021050. URL <https://doi.org/10.1103/physrevx.8.021050>.
- Leonardo Banchi, Nicola Pancotti, and Sougato Bose. Quantum gate learning in qubit networks: Toffoli gate without time-dependent control. *Npj Quantum Information*, 2:16019, 2016.
- Stephen M. Barnett and Sarah Croke. Quantum state discrimination. *Advances in Optics and Photonics*, 1(2):238, apr 2009. ISSN 1943-8206. doi: 10.1364/AOP.1.000238. URL <https://www.osapublishing.org/aop/abstract.cfm?uri=aop-1-2-238>.
- Shabir Barzanjeh, Saikat Guha, Christian Weedbrook, David Vitali, Jeffrey H. Shapiro, and Stefano Pirandola. Microwave quantum illumination. *Physical Review Letters*, 114(8), feb 2015. doi: 10.1103/physrevlett.114.080503. URL <https://doi.org/10.1103/physrevlett.114.080503>.
- Charles H. Bennett. Quantum cryptography using any two nonorthogonal states. *Physical Review Letters*, 68(21):3121–3124, may 1992a. doi: 10.1103/physrevlett.68.3121. URL <https://doi.org/10.1103/physrevlett.68.3121>.
- Charles H Bennett. Quantum cryptography using any two nonorthogonal states. *Physical review letters*, 68(21):3121, 1992b.
- János A Bergou. Quantum state discrimination and selected applications. *Journal of Physics: Conference Series*, 84:012001, oct 2007. ISSN 1742-6596. doi: 10.1088/1742-6596/84/1/012001. URL <http://stacks.iop.org/1742-6596/84/i=1/a=012001?key=crossref.675d05b3d19f4f2bf39356d3f466d6b1>.
- Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, sep 2017. doi: 10.1038/nature23474. URL <https://doi.org/10.1038/nature23474>.
- Anthony Chefles. Quantum state discrimination. *Contemporary Physics*, 41(6):401–424, 2000.
- Carlo Ciliberto, Mark Herbster, Alessandro Davide Ialongo, Massimiliano Pontil, Andrea Rocchetto, Simone Severini, and Leonard Wossnig. Quantum machine learning: a classical perspective. *Proc. R. Soc. A*, 474(2209):20170551, 2018.
- Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks, 2018.
- C.L. Degen, F. Reinhard, and P. Cappellaro. Quantum sensing. *Reviews of Modern Physics*, 89(3), jul 2017. doi: 10.1103/revmodphys.89.035002. URL <https://doi.org/10.1103/revmodphys.89.035002>.
- Lu-Ming Duan and Guang-Can Guo. Probabilistic cloning and identification of linearly independent quantum states. *Physical review letters*, 80(22):4999, 1998.
- Marco Fanizza, Andrea Mari, and Vittorio Giovannetti. Optimal universal learning machines for quantum state discrimination. *arXiv preprint arXiv:1805.03477*, 2018.
- Edward Farhi and Hartmut Neven. Classification with quantum neural networks on near term processors. *arXiv preprint arXiv:1802.06002*, 2018.
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.
- Edward Grant, Marcello Benedetti, Shuxiang Cao, Andrew Hallam, Joshua Lockhart, Vid Stojevic, Andrew G Green, and Simone Severini. Hierarchical quantum classifiers. *arXiv preprint arXiv:1804.03680*, 2018.

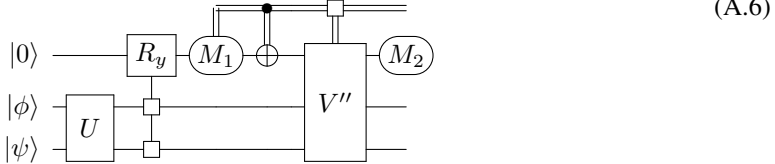
- Luca Innocenti, Leonardo Banchi, Alessandro Ferraro, Sougato Bose, and Mauro Paternostro. Supervised learning of time-independent hamiltonians for gate design. *arXiv preprint arXiv:1803.07119*, 2018.
- Raban Iten, Roger Colbeck, Ivan Kukuljan, Jonathan Home, and Matthias Christandl. Quantum Circuits for Isometries. *Physical Review A - Atomic, Molecular, and Optical Physics*, jan 2015. doi: 10.1103/PhysRevA.93.032318. URL <http://arxiv.org/abs/1501.06911><http://dx.doi.org/10.1103/PhysRevA.93.032318>.
- Raban Iten, Roger Colbeck, and Matthias Christandl. Quantum Circuits for Quantum Channels. *Physical Review A - Atomic, Molecular, and Optical Physics*, 93(3):052316, sep 2016. ISSN 24699934. doi: 10.1103/PhysRevA.95.052316. URL <http://arxiv.org/abs/1609.08103><http://link.aps.org/doi/10.1103/PhysRevA.95.052316><http://dx.doi.org/10.1103/PhysRevA.95.052316>.
- Torsten Karzig, Christina Knapp, Roman M. Lutchyn, Parsa Bonderson, Matthew B. Hastings, Chetan Nayak, Jason Alicea, Karsten Flensberg, Stephan Plugge, Yuval Oreg, Charles M. Marcus, and Michael H. Freedman. Scalable designs for quasiparticle-poisoning-protected topological quantum computation with majorana zero modes. *Physical Review B*, 95(23), jun 2017. doi: 10.1103/physrevb.95.235305. URL <https://doi.org/10.1103/physrevb.95.235305>.
- H. J. Kimble. The quantum internet. *Nature*, 453(7198):1023–1030, jun 2008. doi: 10.1038/nature07127. URL <https://doi.org/10.1038/nature07127>.
- Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. dec 2014. URL <http://arxiv.org/abs/1412.6980>.
- Ying Li and Simon C Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Physical Review X*, 7(2):021050, 2017.
- Seth Lloyd and Christian Weedbrook. Quantum generative adversarial learning. *Physical Review Letters*, 121(4), jul 2018. doi: 10.1103/physrevlett.121.040502. URL <https://doi.org/10.1103/physrevlett.121.040502>.
- Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *arXiv preprint arXiv:1803.00745*, 2018.
- Masoud Mohseni, Aephraim M. Steinberg, and János A. Bergou. Optical Realization of Optimal Unambiguous Discrimination for Pure and Mixed Quantum States. *Physical Review Letters*, 93(20):200403, nov 2004. ISSN 0031-9007. doi: 10.1103/PhysRevLett.93.200403. URL <https://link.aps.org/doi/10.1103/PhysRevLett.93.200403>.
- Xiao-Liang Qi and Shou-Cheng Zhang. Topological insulators and superconductors. *Reviews of Modern Physics*, 83(4):1057–1110, oct 2011. doi: 10.1103/revmodphys.83.1057. URL <https://doi.org/10.1103/revmodphys.83.1057>.
- Philippe Raynal, Norbert Lütkenhaus, and Steven J. van Enk. Reduction theorems for optimal unambiguous state discrimination of density matrices. *Phys. Rev. A*, 68:022308, Aug 2003. doi: 10.1103/PhysRevA.68.022308. URL <https://link.aps.org/doi/10.1103/PhysRevA.68.022308>.
- Ji-Gang Ren, Ping Xu, Hai-Lin Yong, Liang Zhang, Sheng-Kai Liao, Juan Yin, Wei-Yue Liu, Wen-Qi Cai, Meng Yang, Li Li, Kui-Xing Yang, Xuan Han, Yong-Qiang Yao, Ji Li, Hai-Yan Wu, Song Wan, Lei Liu, Ding-Quan Liu, Yao-Wu Kuang, Zhi-Ping He, Peng Shang, Cheng Guo, Ru-Hua Zheng, Kai Tian, Zhen-Cai Zhu, Nai-Le Liu, Chao-Yang Lu, Rong Shu, Yu-Ao Chen, Cheng-Zhi Peng, Jian-Yu Wang, and Jian-Wei Pan. Ground-to-satellite quantum teleportation. *Nature*, 549(7670):70–73, aug 2017. doi: 10.1038/nature23675. URL <https://doi.org/10.1038/nature23675>.
- Andrea Rocchetto, Edward Grant, Sergii Strelchuk, Giuseppe Carleo, and Simone Severini. Learning hard quantum distributions with variational autoencoders. 2017. doi: 10.1038/s41534-018-0077-z.

- Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. Quantum autoencoders for efficient compression of quantum data. *Quantum Science and Technology*, 2(4):045001, 2017.
- Gernot Schaller and Ralf Schützhold. Quantum algorithm for optical-template recognition with noise filtering. *Physical Review A*, 74(1), jul 2006. doi: 10.1103/physreva.74.012303. URL <https://doi.org/10.1103/physreva.74.012303>.
- Maria Schuld, Alex Bocharov, Krysta Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *arXiv preprint arXiv:1804.00633*, 2018.
- Vivek V. Shende, Igor L. Markov, and Stephen S. Bullock. Smaller two-qubit circuits for quantum communication and computation. *Proceedings - Design, Automation and Test in Europe Conference and Exhibition*, 2:980–985, 2004. ISSN 1530-1591. doi: 10.1109/DATE.2004.1269020.
- Vivek V. Shende, Stephen S. Bullock, and Igor L. Markov. Synthesis of Quantum Logic Circuits. *Ieee Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 18, 2006. ISSN 0278-0070. doi: 10.1109/TCAD.2005.855930. URL <http://arxiv.org/abs/quant-ph/0406176>.
- Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. Ladder variational autoencoders, 2016.
- Guillaume Verdon, Michael Broughton, and Jacob Biamonte. A quantum algorithm to train neural networks using low-depth circuits. *arXiv preprint arXiv:1712.05304*, 2017.
- Kwok Ho Wan, Oscar Dahlsten, Hlér Kristjánsson, Robert Gardner, and MS Kim. Quantum generalisation of feedforward neural networks. *npj Quantum Information*, 3(1):36, 2017.

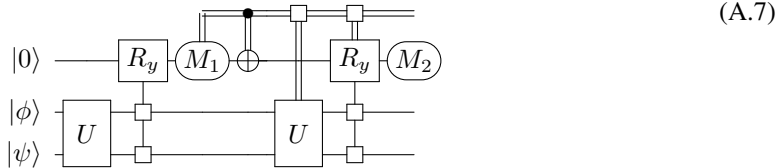
The uniformly controlled V' and U can be merged and put after the measurement of M_1 as:



The first line of the circuit could be merged with the second line as follows:



And then we can apply the cosine-sine decomposition to V'' . Throwing away the last gate on the third and the fourth qubits, we obtain:



The uniformly-controlled rotations and the remaining two-qubit unitary gates could be easily parametrised by CNOTs and single qubit rotations. For example, see [Shende et al. \(2006\)](#) and [Shende et al. \(2004\)](#).

B LEARNING CONVERGENCE FROM ENSEMBLE MEASUREMENTS

Here we simulate the process that a classical-quantum hybrid scheme would implement utilizing a quantum device and analyse its performance. These numerical simulations can in principle be validated in a physical experiment, where the measurement outcomes are used to infer the different probabilities for the cost function. To have a good estimation of the probabilities, and hence the cost function, one has to make repeated measurements to train the model, and we note that in particular better methods to evaluate the analytical gradient are available on a shallow quantum device ([Mitarai et al. \(2018\)](#)). We first give a brief discussion of the estimated number of repeated measurements which are required to approximate the gradient. This follows the treatment of [Farhi & Neven \(2018\)](#)[Section 3]. Since the gradients are calculated using the forward difference formula:

$$\frac{df}{dx}(x) = \frac{f(x + \varepsilon) - f(x)}{\varepsilon} + O(\varepsilon) \tag{B.1}$$

The error in the calculation of f must be at most of the order of $O(\varepsilon^2)$, in order to prevent dominating the total error. To achieve this ideally with a 99% probability, one requires the number of repeated measurement to be of the order $\frac{1}{(\varepsilon^2)^2} = \frac{1}{\varepsilon^4}$.¹ For example, when $\varepsilon = 10^{-3}$, the ideal number of repetitions is given by 10^{12} .

In practice, we do not use $\frac{1}{\varepsilon^4}$ measurements, since the Adam optimization algorithm is designed with the noise of the cost function taken into account. To give an estimate of the number of repeated measurements which are required for the convergence of the optimization process, we performed two numerical experiments. We first looked at the case when the number of repeated measurements

¹ This assumes that the cost function follows a normal distribution with variance of the order $\frac{1}{\sqrt{N}}$, where N is the number of measurements made in each run in order to calculate the cost function.

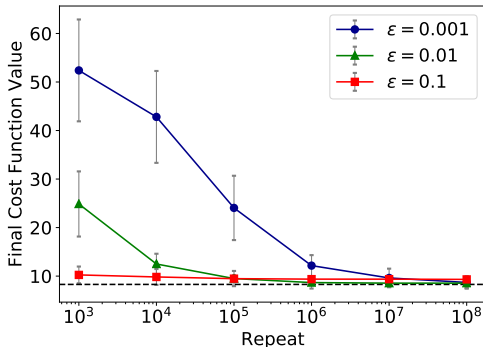
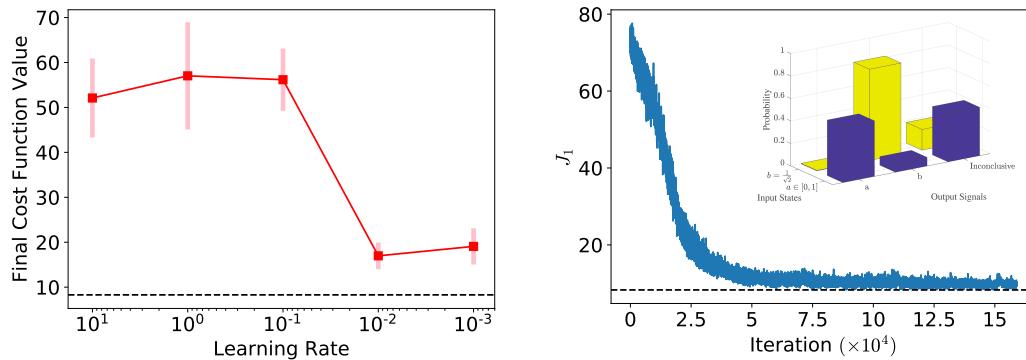


Figure 4: The cost function after 5000 iterations. The result obtained using exact probabilities is shown by the horizontal dashed line. For a smaller step size (ε) for gradient calculations, we found that more repetitions were required to give a consistent result. However, a combination of $\varepsilon = 10^{-2}$ and 10^5 repetitions gave a result which well approximated the result obtained using exact probabilities. Here *repeat* is the number of repeated measurements that were made each time to calculate the cost function. The cost function values were averaged over 50 repeated runs of the training process, and the bars indicate the standard deviations.

was large ($\geq 10^3$) and $\varepsilon = 10^{-2}$. We found that 10^5 repeated measurements for each iteration were a robust configuration for a successful convergence. Second, we used a small number of repeated measurements but varied the learning rate and increased the maximal number of iterations for Adam. Setting $\varepsilon = 10^{-2}$ and taking only 100 repeated measurements, we observed that the optimizations were successful with a large number of iterations. In both experiments, the penalties were set to $\alpha_{inc} = 5$ and $\alpha_{err} = 40$.

Large number of repetitions. Our results showed that for a fixed maximum number of iterations (5000) for Adam, a combination of $\varepsilon = 10^{-2}$ and 10^5 repeated measurements gave robust results, i.e., the final cost function was close to the value obtained with the exact probabilities (with an error within 3%) and was stable (with a relative standard deviation of 13%). A more detailed description of the trade-off between repeated measurements and the stability of the cost function is shown in Fig. 4.

Small learning rates and high number of iterations. Our numerical experiments further showed that in the case of using a small number of repeated measurements, lowered learning rates could effectively counter the noisy brought by the insufficient sampling. Although in this case, the optimization required a large number of iterations to finish. For example, with only 100 repeated measurements, the variance of cost function J_1 after 20000 iterations decreased as we lowered the learning rate (Fig.5(a)). We could visually observe the optimization process where the cost function J_1 slowly approached the optimal value in Fig 5(b). Here, gradient step was taken as $\varepsilon = 10^{-2}$.



(a) Lowering learning rate to counter the effect of insufficient sampling. Both the value and the standard deviation of cost function at the 20000'th iteration were brought down by lowered learning rate. Here the number of repeated measurements was only 100, much smaller than $\frac{1}{\epsilon^4} = 10^8$.

(b) The noisy cost function J_1 estimated by 100 repeated measurements slowly moved to its optimal value when the learning rate was set as 0.001. The horizontal dashed line showed the minimal value 8.3 for the cost function. The inset illustrates that trained circuit could discriminate the two quantum states. Although the error rate in the inset is not 0, we believe it could be achieved by further tuning the penalties, which was set as ($\alpha_{err} = 40, \alpha_{inc} = 5$) for this training.

Figure 5: Small learning rates with high number of iterations.