Journal of Physics:
Complexity

Celebrating the 5th anniversary of JPhys Complexity

PURPOSE-LED PUBLISHING™

PAPER • OPEN ACCESS

General Vapnik–Chervonenkis dimension bounds for quantum circuit learning

To cite this article: Chih-Chieh Chen *et al* 2022 *J. Phys. Complex.* **3** 045007

View the article online for updates and enhancements.

You may also like

- Spin-reorientation transitions in rare-earth magnets
  Konstantin P Belov, Anatolii K Zvezdin, Antonina M Kadomtseva et al.
- Vapnik-Chervonenkis bounds for generalization
  J M R Parrondo and C Van den Broeck
- Widths of some classes of convex functions and bodies
  V N Konovalov and Vitalii E Maiorov

This content was downloaded from IP address 23.94.129.119 on 18/09/2025 at 08:51

IOP Publishing J. Phys. Complex. 3 (2022) 045007 https://doi.org/10.1088/2632-072X/ac9f9b

Journal of Physics: Complexity

OPEN ACCESS

RECEIVED
19 April 2022

REVISED
9 August 2022

ACCEPTED FOR PUBLICATION
1 November 2022

PUBLISHED
14 November 2022

Original Content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

PAPER

General Vapnik–Chervonenkis dimension bounds for quantum circuit learning

Chih-Chieh Chen[1,*], **Masaru Sogabe**[1], **Kodai Shiba**[1,2], **Katsuyoshi Sakamoto**[2,3] **and Tomah Sogabe**[1,2,3,*]

[1] Grid Inc., 107-0061 Tokyo, Japan
[2] Engineering department, The University of Electro-Communications, 182-8585 Tokyo, Japan
[3] i-PERC, The University of Electro-Communications, 182-8585 Tokyo, Japan
[*] Authors to whom any correspondence should be addressed.

E-mail: chen.chih.chieh@gridsolar.jp and sogabe@uec.ac.jp

Keywords: machine learning, quantum circuit, VC dimension

Abstract

Quantifying the model complexity of quantum circuits provides a guide to avoid overfitting in quantum machine learning. Previously we established a Vapnik–Chervonenkis (VC) dimension upper bound for 'encoding-first' quantum circuits, where the input layer is the first layer of the circuit. In this work, we prove a general VC dimension upper bound for quantum circuit learning including 'data re-uploading' circuits, where the input gates can be single qubit rotations anywhere in the circuit. A linear lower bound is also constructed. The properties of the bounds and approximation-estimation trade-off considerations are discussed.

1. Introduction

Due to the difficulty of simulating quantum systems using classical computers, building computing machines using quantum mechanics is suggested as a way toward computational advantages [1–6]. The computational capability of current Noisy Intermediate-Scale Quantum (NISQ) [7] hardware was experimentally demonstrated [8]. On the other hand, classical machine learning [9–12] for Artificial Intelligence (AI) shows a wide range of applications [13, 14]. It is reasonable to consider NISQ devices for AI applications [15, 16].

Using variational quantum circuits [17–19] as prediction models in supervised learning leads to the quantum circuit learning (QCL) method [16, 19–21]. In this setting, the learning task is similar to classical setting where the training data set and predictions are restricted to classical data. Only the hypothesis set is constructed using variational quantum circuits. Theoretical efforts toward understanding the expressive power of QCL have been conducted by many groups [22–26].

One important question in supervised learning is the learnability of the hypothesis set being used. If the size of training data set is small but the model complexity is high, a learning machine could overfit to the data noise and hence fail to generalize well for future predictions. Uniform non-asymptotic theory of generalization for supervised machine learning started with Vapnik–Chervonenkis (VC) theory [27] and is generally known as statistical learning theory [28–32]. Probably Approximately Correct (PAC) framework proposed by Valiant [33] also includes computational requirements in its original form. For binary classification tasks, VC theory can be used to establish the generalization ability by using the VC dimension of the model class [34].

Previous learnability results for quantum machine learning are based on fat-shattering dimension [35], pseudo-dimension [36], or quantum sample complexity [37]. Many recent learnability results based on various measures and settings could be found in literatures [38–45]. Another VC-dimension upper bound, which is different from our result, is proposed in [45]. Their bound is related to the dimension of vector space sum of images of observable operators, and their bound is restricted to 'encoding-first' circuits. Caro *et al* [43] obtain a Rademacher complexity generalization bound for Lipschitz loss functions, which is asymptotically equivalent to our result. Abbas *et al* [39] and Huang *et al* [38] provide input-dependent results. Du *et al* [41] give generalization result by using covering number bound [46] for Lipschitz loss functions. Bu *et al* [42] give Rademacher complexity bound in terms of $L_{p,q}$ matrix norm of operators.

© 2022 The Author(s). Published by IOP Publishing Ltd

The limitation of expressibility of 'encoding-first' quantum circuit was observed by many groups [24, 43, 47], and the 'data re-uploading' circuit [47] was proposed to resolve the limitation. The learnability of data re-uploading QCL is shown in [43] by using Rademacher complexity. Our previous study [48] shows that the growth of VC dimension saturates for deep QCL in the 'encoding-first' scheme. This is different from classical deep neural networks (number of edges $= |E|$, number of vertices $= |V|$), where the VC dimension grows asymptotically as $O(|E|\log(|E|))$ (for sign activation function) or $O(|V|^2|E|^2)$ (for sigmoid activation function) [29, 31, 49, 50]. In this work, we extend our previous [48] result of VC dimension upper bound to include the data re-uploading scheme [51]. The new results also include more general cases like mixed initial states and some hardware noise channels. A lower bound is also presented.

This paper is organized as follows. Section 2 provides brief explanations for quantum circuit learning method and statistical learning theory. Section 3 contains the main results and their proofs. Further discussions about the results are presented in section 4.

2. Preliminaries

Quantum circuit learning and statistical learning theory are introduced in this section.

2.1. Quantum circuit learning

For a supervised binary classification learning problem, we are given some classical training data set $\{(\vec{x}_1,y_1),(\vec{x}_2,y_2),\ldots,(\vec{x}_N,y_N)\} \subseteq X\times Y$, $Y=\{-1,1\}$ drawn from some unknown joint probability distribution $(\vec{x}_i,y_i)\sim P(\vec{x},y)$ over $X\times Y$. The goal of learning is to obtain a model $h: X\rightarrow Y$ such that the prediction error (out-of-sample error) $E_{out}=\mathbb{P}_{(\vec{x},y)\sim P(\vec{x},y)}[h(\vec{x})\neq y]$ is small.

The QCL considered in this work uses some quantum circuits to construct the hypothesis set H. Figure 1 depicts one example of data re-uploading QCL. For some d-dimensional input vector $\vec{x}=(x_0,\ldots,x_{d-1})\in[-1,1]^d=X$, some encoding maps $\vec{\phi}=(\phi_0,\ldots,\phi_{d-1}):[-1,1]^d\rightarrow[-\pi,\pi]^d$, and some real variational parameters θ, the circuit gives a unitary evolution $U_\theta(\vec{\phi}(\vec{x}))$ acting on all-zero initial state $|0\rangle^{\otimes n}$. n denotes the number of qubits (circuit width). We do not assume any special structure for variational parameters and entanglers, while the encoding method is specified as follows. For one input vector $\vec{x}=(x_0,\ldots,x_{d-1})\in[-1,1]^d=X$, each dimension $x_i\in[-1,1]$ is encoded by one encoding mapping $\phi_i:[-1,1]\rightarrow[-\pi,\pi]$ with one single qubit rotation $R_s\in\{R_Y,R_Z\}$. The gate $R_s(\phi_i(x_i))$ is applied to the quantum circuit to upload the data. Data re-uploading means that the gate $R_s(\phi_i(x_i))$ is applied to the circuit several times for an $i\in\{0,\ldots,d-1\}$. The number n_i denotes the total number of $R_s(\phi_i(x_i))$ gates being applied for an $i\in\{0,\ldots,d-1\}$. The measurement result is used to compute the expectation value for some fixed observable O:

$$\langle O(\theta,\vec{\phi}(\vec{x}))\rangle = \mathrm{Tr}[OU_\theta(\vec{\phi}(\vec{x}))|0\rangle^{\otimes n}\langle 0|^{\otimes n}U_\theta^\dagger(\vec{\phi}(\vec{x}))]. \tag{1}$$

The expectation value is then thresholded to construct a hypothesis set $H=\{\mathrm{sgn}(f_\theta(\vec{\phi}(\vec{x}))+c): f_\theta(\vec{\phi}(\vec{x}))=\langle O(\theta,\vec{\phi}(\vec{x}))\rangle=\mathrm{Tr}[OU_\theta(\vec{\phi}(\vec{x}))|0\rangle^{\otimes n}\langle 0|^{\otimes n}U_\theta^\dagger(\vec{\phi}(\vec{x}))], c\in\mathbb{R}\}$ for binary classification.

2.2. Statistical learning theory

Under suitable measure-theoretical assumptions [52], VC theory provides a general theory of generalization ability for binary classification tasks. We use the definition that the generalization error is $E_{out}-E_{in}$, where E_{out} is the out-of-sample error (prediction error) and $E_{in}=\frac{1}{N}\sum_{i=1}^{N}[\![h(\vec{x}_i)\neq y_i]\!]$ is the in-sample error. $[\![\cdot]\!]$ is the Iverson bracket. Given a statement s, $[\![s]\!]=1$ if the statement s is true, and $[\![s]\!]=0$ if the statement s is false. The VC generalization error bound is [27]:

$$\mathbb{P}[\sup_{h\in H}|E_{out}(h)-E_{in}(h)|>\epsilon]\leqslant 4m_H(2N)e^{(-\frac{1}{8}\epsilon^2 N)}, \tag{2}$$

where the randomness is over i.i.d. samples $\{(\vec{x}_i,y_i)\sim P(\vec{x},y)\ \forall i\in\{1,\ldots,N\}\}$. N is the sample size. The function $m_H(N)=\max_{\vec{x}_1,\ldots,\vec{x}_N\in X}|\{(h(\vec{x}_1),\ldots,h(\vec{x}_N)):h\in H\}|$ could be upper bounded by $m_H(N)\leqslant\sum_{i=0}^{d_{vc}}\binom{N}{i}\leqslant N^{d_{VC}}+1$ for finite VC-dimension $d_{VC}=\max_{N\in\mathbb{N}}\{N: m_H(N)=2^N\}$. VC dimension is the maximum number of points that can be shattered by the hypothesis set. In general, d_{VC} could be infinite for an uncountable hypothesis set. If d_{VC} is finite, then the generalization ability of the learning machine is guaranteed by the VC bound and the hypothesis set is called 'PAC-learnable.' Several features of VC theory are worth noting [28]: (1) VC bound is independent of the input distribution. (2) VC bound is non-asymptotic, so it can be applied when the size of training data set is small. (3) VC bound is uniform over the hypothesis set, which means that it is true for all the models in the set.

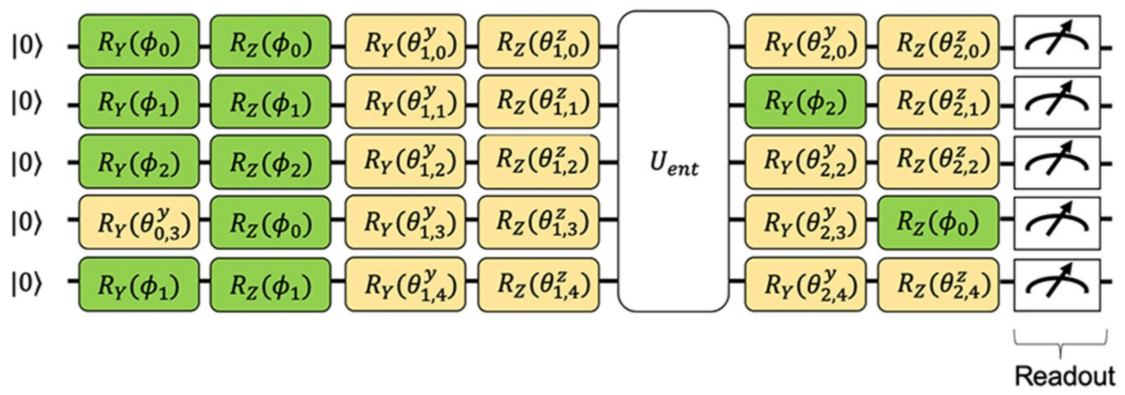
$|0\rangle$
$|0\rangle$
$|0\rangle$
$|0\rangle$
$|0\rangle$
$R_Y(\phi_0)$ $R_Z(\phi_0)$ $R_Y(\theta^y_{1,0})$ $R_Z(\theta^z_{1,0})$ $R_Y(\theta^y_{2,0})$ $R_Z(\theta^z_{2,0})$
$R_Y(\phi_1)$ $R_Z(\phi_1)$ $R_Y(\theta^y_{1,1})$ $R_Z(\theta^z_{1,1})$ $R_Y(\phi_2)$ $R_Z(\theta^z_{2,1})$
$R_Y(\phi_2)$ $R_Z(\phi_2)$ $R_Y(\theta^y_{1,2})$ $R_Z(\theta^z_{1,2})$ $R_Y(\theta^y_{2,2})$ $R_Z(\theta^z_{2,2})$
$R_Y(\theta^y_{0,3})$ $R_Z(\phi_0)$ $R_Y(\theta^y_{1,3})$ $R_Z(\theta^z_{1,3})$ $R_Y(\theta^y_{2,3})$ $R_Z(\phi_0)$
$R_Y(\phi_1)$ $R_Z(\phi_1)$ $R_Y(\theta^y_{1,4})$ $R_Z(\theta^z_{1,4})$ $R_Y(\theta^y_{2,4})$ $R_Z(\theta^z_{2,4})$
U_{ent}
Readout

Figure 1. One example of data re-uploading quantum circuit learning. A green gate $R_\mu(\phi_i(x_i))$ denotes a data encoding gate, where $\vec{x}$ is the input vector. A yellow gate $R_\nu(\theta_{j,k})$ denotes a variational gate, where $\theta_{j,k}$ is a variational parameter. The entangling gate U_{ent} can be an arbitrary long-range entangler. In this example, $n_0 = 4$, $n_1 = 4$, and $n_2 = 3$.

After VC theory, there are latter developments for the generalization ability of learning machines. For real-valued functions, the pseudo-dimension [53] and the fat-shattering dimension [54, 55] could be used for generalization bounds. VC theory is also extended to real-valued functions [28]. Rademacher complexity can be used to obtain generalization bounds for classification and regression [32]. PAC-Bayesian bounds are proposed for Bayesian setting [56–59]. There are also other generalization bounds which are not VC bound but use VC dimension as a measure [60]. Some introductions and comparative study of these measures could be found in [29, 31, 32, 60].

3. Main result

The main results are presented here. The proofs are extensions of the proofs in [48].

Theorem 1 (VC dimension upper bound for quantum circuits). *Assume the input vector* $\vec{x} = (x_0, \ldots, x_{d-1}) \in [-1, 1]^d$*. Each dimension* $x_i \in [-1, 1]$ *is uploaded* n_i *times using single qubit encoding rotations* $R_s(\phi_i(x_i))$ *for some fixed encoding mapping* $\phi_i : [-1, 1] \to [-\pi, \pi]$ *with* $s \in \{Y, Z\}$*. Then the VC dimension of the hypothesis set* $H = \{\mathrm{sgn}(f_\theta(\vec{\phi}(\vec{x})) + c) : f_\theta(\vec{\phi}(\vec{x})) = \langle O(\theta, \vec{\phi}(\vec{x}))\rangle = \mathrm{Tr}[OU_\theta(\vec{\phi}(\vec{x}))|0\rangle^{\otimes n}\langle 0|^{\otimes n}U_\theta^\dagger(\vec{\phi}(\vec{x}))], c \in \mathbb{R}\}$ *for a fixed observable* O *is upper bounded by:*

$$d_{VC} \leqslant \prod_{i=0}^{d-1}(2n_i + 1). \tag{3}$$

Proof. We claim that $f_\theta(\vec{\phi}(\vec{x}))$ is a real trigonometric polynomial of d variables, and the degree of the polynomial for each variable is at most n_i. Then the theorem is proved by Dudley's theorem for VC dimension of thresholded real vector space function classes [29, 31, 50, 61].

The proof of the claim is as follows. The initial density matrix $\rho_0 = |0\rangle^{\otimes n}\langle 0|^{\otimes n}$ has constant matrix elements. From the assumptions, all the variational unitaries and entanglers do not depend on input vector $\vec{x}$. Consider an input dimension $x_i \in [-1, 1]$ and encoding mapping $\phi_i(x_i) \in [-\pi, \pi]$ where $i \in \{0, \ldots, d-1\}$. If this dimension is uploaded by R_Y:

$$R_Y(\phi_i) = \begin{pmatrix} \cos(\frac{\phi_i}{2}) & -\sin(\frac{\phi_i}{2}) \\ \sin(\frac{\phi_i}{2}) & \cos(\frac{\phi_i}{2}) \end{pmatrix} \tag{4}$$

$$= \cos\left(\frac{\phi_i}{2}\right)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin\left(\frac{\phi_i}{2}\right)\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{5}$$

then the action of this gate on kth qubit of n-qubit Hilbert space is:

$$R_Y(\phi_i)|_k = \cos\left(\frac{\phi_i}{2}\right)\mathbb{I}_{2^n} + \sin\left(\frac{\phi_i}{2}\right)\mathbb{I}_{2^k} \otimes \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \otimes \mathbb{I}_{2^{n-k-1}} \tag{6}$$

$$= \cos\left(\frac{\phi_i}{2}\right)\mathbb{I}_{2^n} + \sin\left(\frac{\phi_i}{2}\right)\mathbb{A}, \tag{7}$$

where $\mathbb{I}_M$ denotes $M \times M$ identity matrix and $\mathbb{A}$ is some constant matrix. The action of this gate on a density matrix ρ is then:

$$R_Y(\phi_i)|_k \rho R_Y(\phi_i)|_k^\dagger \tag{8}$$

$$= \left(\cos\left(\frac{\phi_i}{2}\right)\mathbb{I}_{2^n} + \sin(\frac{\phi_i}{2})\mathbb{A}\right)\rho\left(\cos\left(\frac{\phi_i}{2}\right)\mathbb{I}_{2^n} + \sin\left(\frac{\phi_i}{2}\right)\mathbb{A}^\dagger\right) \tag{9}$$

$$= \frac{1}{2}[(1+\cos(\phi_i))\rho + \sin(\phi_i)(\mathbb{A}\rho + \rho\mathbb{A}^\dagger) \tag{10}$$

$$+ (1-\cos(\phi_i))\mathbb{A}\rho\mathbb{A}^\dagger]. \tag{11}$$

If the matrix elements of ρ are trigonometric polynomials of $\vec{\phi}$, then the matrix elements of the updated density matrix $R_Y(\phi_i)|_k \rho R_Y(\phi_i)|_k^\dagger$ are trigonometric polynomials where the degree for the variable ϕ_i is increased by at most one. Similar argument works if the dimension is uploaded by $R_Z(\phi_i)$. Hence, $f_\theta(\vec{\phi}(\vec{x}))$ is a trigonometric polynomial with the claimed degree upper bound. Let $f_\theta(\vec{\phi}(\vec{x})) = \sum_k a_k(\theta) f_k(\vec{\phi}(\vec{x}))$, where $\{f_k(\vec{\phi})\}$ is the real trigonometric polynomial basis and $\{a_k(\theta)\}$ are the Fourier coefficients. Since $f_\theta(\vec{\phi}(\vec{x}))$ is a real-valued function, the coefficients $a_k(\theta) = \langle f_k | f_\theta \rangle \in \mathbb{R}\ \forall k$. The claim is proved. □

Theorem 2 (VC dimension lower bound for quantum circuits). *For d-dimensional input space, there is a quantum circuit hypothesis set with VC dimension $d_{VC} \geqslant 2d+1$ using d qubits, d encoding gates, and d variational gates.*

Proof. For any input vector $\vec{x} = (x_0, \ldots, x_{d-1})$, use the circuit $\bigotimes_{i=0}^{d-1}(R_y(\theta_i)R_y(\pi x_i)|0\rangle_i)$ such that $\langle Z_i \rangle = \cos(\theta_i + \pi x_i) = \cos(\theta_i)\cos(\pi x_i) - \sin(\theta_i)\sin(\pi x_i)$. Then the hypothesis set $H = \{\mathrm{sgn}(\sum_{i=0}^{d-1} c_i[\cos(\theta_i)\cos(\pi x_i) - \sin(\theta_i)\sin(\pi x_i)] + b) : c_i \in \mathbb{R}, b \in \mathbb{R}, \theta_i \in (-\pi, \pi]\}$ has VC dimension $d_{VC} \geqslant 2d+1$ by Dudley's theorem. □

4. Discussions

In this section, we provide some short discussions regarding the obtained theorem.

4.1. Applicability of the bounds

Regarding the upper bound, there is no requirement on the structure of variational (trainable) gates and entangling gates of the circuit, except that they should not contain any input data x_i. There is no requirement on the encoding gates $R_s(\phi_i(x_i))$, except that they should not contain any variational parameter.

Notice that in practice, one usually applies some classical post processing techniques to the output expectation values [16]. The VC dimension bound should be adjusted accordingly.

We provide some extensions.

Corollary 1 (Linear combinations of expectations). *If the hypothesis set is the real linear combination of several observables for a fixed circuit such that $H = \{\mathrm{sgn}(f_\theta(\vec{\phi}(\vec{x})) + c_0) : f_\theta(\vec{\phi}(\vec{x})) = \sum_i c_i \langle O_i(\theta, \vec{\phi}(\vec{x}))\rangle = \sum_i c_i \mathrm{Tr}[O_i U_\theta(\vec{\phi}(\vec{x}))|0\rangle^{\otimes n}\langle 0|^{\otimes n} U_\theta^\dagger(\vec{\phi}(\vec{x}))], c_i \in \mathbb{R}\}$, then the bound in Theorem 1 is still true.*

Proof. Apply Theorem 1 to each O_i. The corollary is then a direct consequence of Dudley's theorem. □

Corollary 2 (Mixed states). *If the initial state is some mixed state ρ which does not depend on the input vector $\vec{x}$ such that $H_{QCL} = \{\mathrm{sgn}(f_\theta(\vec{\phi}(\vec{x})) + c_0) : f_\theta(\vec{\phi}(\vec{x})) = \sum_i c_i \langle O_i(\theta, \vec{\phi}(\vec{x}))\rangle = \sum_i c_i \mathrm{Tr}[O_i U_\theta(\vec{\phi}(\vec{x}))\rho U_\theta^\dagger(\vec{\phi}(\vec{x}))]$, $c_i \in \mathbb{R}\}$, then the bound in Theorem 1 is still true.*

Proof. The proof in Theorem 1 remains true if ρ_0 is an input-independent mixed state density matrix. □

Corollary 3 (Kraus operations). *[62] If any completely positive trace-preserving map $\rho \mapsto \sum_k E_k \rho E_k^\dagger$ is applied to the system, where the E_k's are independent of the input $\vec{x}$, then the bound in Theorem 1 is still true.*

Proof. If the matrix elements of ρ are trigonometric polynomials of $\vec{\phi}$, then the matrix elements of the updated density matrix $\rho' = \sum_k E_k \rho E_k^\dagger$ are trigonometric polynomials of degrees no more than that of ρ. □

Corollary 3 includes several types of hardware noise channels [4, 6]. It does not include the situations where the density matrix has to be normalized by $\rho \mapsto E_k \rho E_k^\dagger / \mathrm{Tr}(E_k \rho E_k^\dagger)$.

For the special case where $n_i = 1$, the upper bound and the lower bound together give $3^d \geqslant d_{VC} \geqslant 2d+1$. For $d = 1$, the bound is saturated with exact number $d_{VC} = 3$. For larger d, there is a gap between the exponential upper bound and the linear lower bound to be explored.

4.2. Reduction to the previous results

We show how to obtain the special case in our previous work [48] for the ansatz in [20]:

$$d_{VC} \leqslant \left(2\frac{n}{d}+1\right)^{2d}. \tag{12}$$

This bound can be obtained from the general bound in Theorem 1 as follows. The encoding used in [20] can be understood as performing feature maps $x_i \mapsto x_i^2$ to increase the feature dimension from d to $2\,d$. The encoding maps $\phi_i(x_i) = \arcsin(x_i)$ and $\phi_i'(x_i^2) = \arccos(x_i^2)$ are used, and are uploaded by $R_Y(\phi_i(x_i)) = R_Y(\arcsin(x_i))$ and $R_Z(\phi_i'(x_i^2)) = R_Z(\arccos(x_i^2))$. Each dimension is uploaded $n_i = \frac{n}{d}$ times, and hence we get the bound $(2n_i+1)^{2d} = (2\frac{n}{d}+1)^{2d}$. The lightcone bound can be calculated by counting n_i covered by the lightcone for a specific ansatz.

4.3. Independent of the number of variational gates

Notice that the upper bound is based on counting the number of basis functions, hence the bound does not depend on the number of variational parameters. This suggests that the bound is asymptotically tight (constant) with respect to number of trainable parameters, but cannot be tight in general (the constant could be too large). For example, if the number of variational parameter is zero, then the VC dimension is zero. It is desirable to also have a scaling with respect to the number of variational gates, like the cases in [40, 44].

4.4. Approximation-estimation trade-off considerations

To achieve low prediction error in supervised learning, the approximation-estimation trade-off (also known as bias-variance trade-off) should be considered [31, 32]. The generalization error bound discussed in this work is only for estimation error.

Barron [63] gives the approximation error bound for single layer classical neural network hypothesis set $H_{NN} = \{f(\vec{x}) = \sum_{k=1}^{n} c_k\phi(\vec{a}_k\cdot\vec{x}+b_k)+c_0 : \vec{a}_k \in \mathbb{R}^d, b_k, c_k \in \mathbb{R}\}$ where ϕ is a sigmoid function and n is the number of nodes. Barron also analyzed the approximation-estimation trade-off of neural networks [64]. It is shown that neural networks have approximation advantage over linear combinations of fixed basis functions in the sense that the approximation has faster convergence rate for high-dimensional inputs.

One attempt to overcome the limitation of fixed basis functions of QCL was actually proposed in [47]: combining neural networks with QCL to construct, for example, the hypothesis set $H_{affineQCL} = \{(f_\theta(\vec{\phi}(\vec{x}))+c_0) : f_\theta(\vec{\phi}(\vec{x})) = \sum_i c_i\langle O_i(\theta,\vec{\phi}(W\cdot\vec{x}+\vec{b}))\rangle = \sum_i c_i\mathrm{Tr}[O_iU_\theta(\vec{\phi}(W\cdot\vec{x}+\vec{b}))\rho U_\theta^\dagger(\vec{\phi}(W\cdot\vec{x}+\vec{b}))],$ $c_i\in\mathbb{R}, W\in\mathbb{R}^{d\times d}, \vec{b}\in\mathbb{R}^d\}$, where the affine transformation $W\cdot\vec{x}+\vec{b}$ is composited with QCL. However, a simple special case $\{\sin(Wx) : W\in\mathbb{R}\}$ has infinite VC dimension, and hence is not PAC-learnable [28, 29, 48]. This is because W provides possibly high-frequency oscillations to shatter arbitrarily many data points. One possible way to resolve this problem could be using a sigmoid activation function ϕ for encoding. For example, the input x_i could be uploaded by the $R_s(\pi\phi(W_ix_i+b_i))$ gate. This could be a future direction.

5. Conclusion

In this work, we give a general VC dimension upper bound and a lower bound for quantum circuit learning, and hence establish the PAC learnability of this hypothesis set. While this result provides a basis for quantum circuit supervised learning, many questions remain. For example, we did not address the issues of sampling error of quantum machines (due to finite readout samples), which could affect the generalization ability. We did not have a bound which scales with respect to the number of trainable parameters. The approximation-estimation trade-off should also be addressed. We do not have experimental result. Numerical simulations for overfitting of data re-uploading QCL could be found in [65]. Entangling dropout is suggested as a regularization technique to avoid overfitting for data re-uploading QCL [65]. It would be desirable to see comparison between theory and experiments for large-scale circuits. These questions are left for future investigations.

Data availability statement

No new data were created or analyzed in this study.

Acknowledgments

We thank Naoki Yamamoto for valuable discussions. We thank Matthias C Caro for providing many useful references.

ORCID iDs

Chih-Chieh Chen https://orcid.org/0000-0003-3092-4346
Tomah Sogabe https://orcid.org/0000-0001-9258-6130

References

[1] Feynman R P 1982 *Int. J. Theor. Phys.* **21** 467–88
[2] Deutsch D 1985 *Proc. R. Soc.* A **400** 97–117
[3] Barenco A, Bennett C H, Cleve R, DiVincenzo D P, Margolus N, Shor P, Sleator T, Smolin J A and Weinfurter H 1995 *Phys. Rev.* A **52** 3457–67
[4] Preskill J 1998 *Lecture Notes for Physics 229: Quantum Information and Computation* (Pasadena, CA: Caltech) (available at: http://theory.caltech.edu/~preskill/ph229/)
[5] Mermin N D 2007 *Quantum Computer Science: An Introduction* (Cambridge: Cambridge University Press)
[6] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge: Cambridge University Press)
[7] Preskill J 2018 *Quantum* **2** 79
[8] Arute F *et al* 2019 *Nature* **574** 505–10
[9] Bishop C M 2006 *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Berlin: Springer)
[10] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press)
[11] Hastie T, Tibshirani R and Friedman J 2009 *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (New York: Springer Science & Business Media) (https://doi.org/10.1007/978-0-387-84858-7)
[12] Bengio Y, Lecun Y and Hinton G 2021 *Commun. ACM* **64** 58–65
[13] Russell S and Norvig P 2020 *Artificial Intelligence: A Modern Approach* 4th edn (Englewood Cliffs, NJ: Prentice Hall Press)
[14] Silver D *et al* 2017 *Nature* **550** 354–9
[15] Dunjko V and Briegel H J 2018 *Rep. Prog. Phys.* **81** 074001
[16] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M and Gambetta J M 2019 *Nature* **567** 209–12
[17] Peruzzo A, McClean J, Shadbolt P, Yung M H, Zhou X Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 *Nat. Commun.* **5** 4213
[18] McClean J R, Romero J, Babbush R and Aspuru-Guzik A 2016 *New J. Phys.* **18** 023023
[19] Cerezo M *et al* 2021 *Nat. Rev. Phys.* **3** 625–44
[20] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 *Phys. Rev.* A **98** 032309
[21] Schuld M, Bergholm V, Gogolin C, Izaac J and Killoran N 2019 *Phys. Rev.* A **99** 032331
[22] Sim S, Johnson P D and Aspuru-Guzik A 2019 *Adv. Quantum Technol.* **2** 1900070
[23] Nakaji K and Yamamoto N 2021 *Quantum* **5** 434
[24] Schuld M, Sweke R and Meyer J J 2021 *Phys. Rev.* A **103** 032430
[25] Hubregtsen T, Pichlmeier J, Stecher P and Bertels K 2021 *Quantum Mach. Intell.* **3** 9
[26] Goto T, Tran Q H and Nakajima K 2021 *Phys. Rev. Lett.* **127** 090506
[27] Vapnik V N and Chervonenkis A Y 1971 *Theory Probab. Appl.* **16** 264–80
[28] Vapnik V N 1999 *The Nature of Statistical Learning Theory* 2nd edn (New York: Springer)
[29] Anthony M and Bartlett P L 2009 *Neural Network Learning: Theoretical Foundations* 1st edn (New York: Cambridge University Press)
[30] Abu-Mostafa Y S, Magdon-Ismail M and Lin H T 2012 *Learning From Data* (New York: AMLBook)
[31] Shalev-Shwartz S and Ben-David S 2014 *Understanding Machine Learning: From Theory to Algorithms* (New York: Cambridge University Press)
[32] Mohri M, Rostamizadeh A and Talwalkar A 2018 *Foundations of Machine Learning* (Cambridge, MA: MIT Press)
[33] Valiant L G 1984 *Commun. ACM* **27** 1134–42
[34] Blumer A, Ehrenfeucht A, Haussler D and Warmuth M K 1989 *J. ACM* **36** 929–65
[35] Aaronson S 2007 *Proc. of the Royal Society A: Mathematical, Physical and Engineering Sciences* **463** 3089–114
[36] Caro M C and Datta I 2020 *Quantum Machine Intelligence* **2** 14
[37] Arunachalam S and de Wolf R 2018 *J. Mach. Learn. Res.* **19** 1–36
[38] Huang H Y, Broughton M, Mohseni M, Babbush R, Boixo S, Neven H and McClean J R 2021 *Nat. Commun.* **12** 2631
[39] Abbas A, Sutter D, Zoufal C, Lucchi A, Figalli A and Woerner S 2021 *Nat. Comput. Sci.* **1** 403–9
[40] Du Y, Tu Z, Yuan X and Tao D 2022 *Phys. Rev. Lett.* **128** 080506
[41] Du Y, Hsieh M H, Liu T, You S and Tao D 2021 *PRX Quantum* **2** 040337
[42] Bu K, Koh D E, Li L, Luo Q and Zhang Y 2021 *Phys. Rev.* A **105** 062431
[43] Caro M C, Gil-Fuster E, Meyer J J, Eisert J and Sweke R 2021 *Quantum* **5** 582
[44] Caro M C, Huang H Y, Cerezo M, Sharma K, Sornborger A, Cincio L and Coles P J 2021 *Nat. Commun.* **13** 4919
[45] Gyurik C, van Vreumingen D and Dunjko V 2021 arXiv:2105.05566
[46] Barthel T and Lu J 2018 *Phys. Rev. Lett.* **121** 080406
[47] Pérez-Salinas A, Cervera-Lierta A, Gil-Fuster E and Latorre J I 2020 *Quantum* **4** 226
[48] Chen C C, Watabe M, Shiba K, Sogabe M, Sakamoto K and Sogabe T 2021 *ACM Trans. Quantum Comput.* **2** 1–24
[49] Kearns M J and Vazirani U V 1994 *An Introduction to Computational Learning Theory* (Cambridge: MIT Press)
[50] Sontag E 1998 *NATO ASI Series. Series F: Computer and System Sciences* pp 69–95
[51] Chen C C, Sogabe M, Shiba K, Sakamoto K and Sogabe T 2022 A general vc dimension upper bound for quantum circuit learning *Technical Report* (Information Processing Society of Japan) (available at: http://id.nii.ac.jp/1001/00217544/)
[52] Dudley R M 2014 Uniform central limit theorems *Cambridge Studies in Advanced Mathematics* 2nd edn (Cambridge: Cambridge University Press)
[53] Pollard D 1984 *Convergence of Stochastic Processes* (Berlin: Springer)
[54] Kearns M J and Schapire R E 1994 *J. Comput. Syst. Sci.* **48** 464–97
[55] Bartlett P L, Long P M and Williamson R C 1996 *J. Comput. Syst. Sci.* **52** 434–52
[56] McAllester D A 1998 Some PAC-Bayesian theorems *Proc. 11th Annual Conf. on Computational Learning Theory (COLT' 98)* (New York: Association for Computing Machinery) pp 230–4

[57] McAllester D A 1999 *Mach. Learn.* **37** 355–63
[58] McAllester D A 2003 *Mach. Learn.* **51** 5–21
[59] McAllester D 2013 arXiv:1307.2118
[60] Jiang Y, Neyshabur B, Mobahi H, Krishnan D and Bengio S 2019 arXiv:1912.02178
[61] Dudley R M 1978 *Ann. Probab.* **6** 899–929
[62] Kraus K 1983 *States, Effects and Operations Fundamental Notions of Quantum Theory* vol 190 (Berlin: Springer)
[63] Barron A 1993 *IEEE Trans. Inf. Theory* **39** 930–45
[64] Barron A R 1994 *Mach. Learn.* **14** 115–33
[65] Kobayashi M, Nakaji K and Yamamoto N 2022 arXiv:2205.11446