Fundamental aspects of solving quantum problems with machine learning*

Hsin-Yuan Huang, Richard Kueng, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, Jarrod R. McClean, and John Preskill

I. OVERVIEW

Machine learning (ML) has received great attention in the quantum community, both for the hope of using classical ML to solve challenging quantum many-body problems in physics/chemistry [11, 12, 24], and for the aspiration of designing new quantum ML that harnesses significant advantages over classical ML [8, 14, 21]. Despite considerable interest, the prospect remains somewhat unclear: How can ML be more useful than traditional algorithms for solving difficult quantum problems? And what are the advantages of quantum ML in general? Here, we focus on an important special case which subsumes many major quantum problems: training an ML model to predict $x \mapsto f(x) = \mathrm{tr}(O\,\mathcal{E}(|x\rangle\langle x|))$. Here, x is a classical input, $\mathcal{E}$ is an unknown completely positive trace-preserving (CPTP) map, and O is an observable. Concrete examples are predicting outcomes of physical experiments or ground state properties of various quantum systems. Two fundamental questions are central for understanding the power of classical ML for quantum problems:

i. *Information-theoretic question:* Would the sample complexity (data size needed) for training classical ML model to accurately predict $\mathrm{tr}(O\,\mathcal{E}(|x\rangle\langle x|))$ be significantly more than the optimal quantum ML model?

ii. *Computational question:* Suppose that we are able to obtain classical data by performing actual experiments. Can a classical ML model use this data to efficiently predict $f(x) = \mathrm{tr}(O\,\mathcal{E}(|x\rangle\langle x|))$ even if $\mathcal{E}(|x\rangle\langle x|))$ is hard to simulate?

We address question i. by proving that there always exists a classical ML model whose sample complexity matches the best quantum ML model (up to a small polynomial factor) for achieving small prediction error in average. In other words: if quantum ML can handle a certain quantum problem, then classical ML can approximately do so as well. Regarding question ii., we first put the problem into a complexity-theoretic context. Efficient classical ML augmented with (potentially hard to compute) training data is strictly more powerful than traditional polynomial-time algorithms (captured by the complexity class BPP). Subsequently, we establish sufficient conditions for efficiently learning quantum problems with classical ML models. As a concrete example, we prove that after training from existing data, classical ML can efficiently predict ground state representations of (any smooth family of) gapped Hamiltonians in any constant spatial dimension. The ground state representations accurately approximate the expectation values of any local observable.

II. INFORMATION-THEORETIC ASPECT

Assume that $\mathcal{E}$ is an unknown CPTP map that belongs to a (known) function class $\mathcal{F}$ ($\mathcal{F}$ may be uncountably large). Apart from $\mathcal{E} \in \mathcal{F}$, the process $\mathcal{E}$ can be arbitrary – a common assumption in statistical learning theory [4, 7, 10, 23, 25]. For the sake of concreteness, let us assume that $\mathcal{E}$ is a CPTP map from a Hilbert space of n qubits to a Hilbert space of m qubits. We will also assume that the inputs are bit-strings[1] of size n: $x \in \{0,1\}^n$. These can be encoded in computational basis states and we shall distinguish two types of ML settings:

- *Classical* (C) ML models are given access to training data $\{(x_i, o_i)\}_{i=1}^{N_\mathrm{C}}$, where each input $x_i \in \{0,1\}^n$ is sampled according to a distribution $\mathcal{D}$. The label o_i is a single-shot measurement outcome with expectation $f(x_i) = \mathrm{tr}(O\mathcal{E}(|x_i\rangle\langle x_i|))$. Classical data of size N_C is used to create a prediction model $f_C(x)$ that approximates $f(x)$.
- *Quantum* (Q) ML models are strictly more powerful. We envision them as external, universal quantum computers with an unlimited amount of quantum memory. The quantum ML models may query the CPTP map $\mathcal{E}$ by applying $\mathcal{E}$ to any part of this quantum memory in addition to performing arbitrary quantum processing that depends on (and has arbitrary access to) the input distribution $\mathcal{D}$, the observable O, and the function class $\mathcal{F}$. After querying $\mathcal{E}$ for N_Q times, the quantum ML model will output a function $f_Q(x)$ that approximates $f(x)$.

The quantum ML setting is strictly more powerful than the classical ML setting. To see this, note that classical ML is restricted to classical information – in the form of observable measurement outcomes. By contrast, quantum ML can work directly in the quantum realm. Given this discrepancy, a large quantum advantage seems plausible. The following theorem, however, dispels this prospect in the context of approximately accurate predictions.

Theorem 1. *Fix an n-bit input distribution $\mathcal{D}$, an m-qubit observable O ($\|O\| \leq 1$) and a family $\mathcal{F}$ of CPTP maps with compatible dimensions (n input qubits and m output qubits). Suppose that the best quantum ML model uses N_Q queries of the unknown process $\mathcal{E} \in \mathcal{F}$ to learn a function $f_Q(x)$ that achieves small average prediction error*

$$\mathbb{E}_{x\sim\mathcal{D}} \left|f_Q(x) - \mathrm{tr}(O\mathcal{E}(|x\rangle\langle x|))\right|^2 \leq \epsilon. \tag{1}$$

Then, classical training data of size $N_\mathrm{C} \leq \mathcal{O}(mN_Q/\epsilon)$ suffices to train a classical ML model that is guaranteed to achieve

$$\mathbb{E}_{x\sim\mathcal{D}} \left|f_C(x) - \mathrm{tr}(O\mathcal{E}(|x\rangle\langle x|))\right|^2 \leq \mathcal{O}(\epsilon). \tag{2}$$

---

* This submission merges *Provable machine learning algorithms for quantum many-body problems* [16], which contains most of the theoretical contributions, and some part of *Power of data in quantum machine learning* [15], which contains more empirical studies. To facilitate assessment, we present all relevant theoretical and empirical contributions in a self-contained manuscript.

[1] This is not a severe restriction, since one can truncate floating-point representations of continuous parameters to a finite number of digits.

The proof is based on three steps. We first lower-bound the sample complexity of the best quantum ML model using tools from quantum communication complexity (encode the family $\mathcal{F}$ and the quantum ML model into a communication protocol and apply a mutual information analysis). Subsequently, we identify a concrete classical ML model that minimizes the training error for a related class of functions (tight statistical analysis yields upper bound on sample complexity). Finally, we combine these results by means of a fundamental relation between covering and packing nets. We also prove that this result is tight by constructing a learning problem with $N_C \geq \Omega(mN_Q/\epsilon)$ for any classical ML.

The contrapositive of Theorem 1 is almost as interesting as the statement itself. Suppose that there exists a quantum problem that really is hard for classical ML – in the sense that the sample complexity is prohibitively large. Then, quantum ML is incapable of offering substantial improvements. From an information-theoretic perspective, all quantum problems that can approximately be solved with quantum ML can also be approximately solved with classical ML.

It is instructive to relate the above setting to quantum probably approximately correct (PAC) learning [3–5, 22, 26], which established rigorous results on the lack of quantum advantage in sample complexity for learning classical Boolean function $h : \{0,1\}^n \to \{0,1\}$. In quantum PAC learning, the quantum ML models (quantum learners) are restricted to obtaining quantum samples of the form $\sum_x \sqrt{D(x)}\,|x\rangle\,|h(x)\rangle$, where $D(x)$ is the probability for sampling x in the input distribution $\mathcal{D}$ and $h(x)$ is the classical Boolean function. Furthermore, quantum PAC learning addresses the worst-case input distribution $\mathcal{D}_\star$. This leaves open the potential for large quantum advantage in more general settings, such as when the quantum ML can access $\mathcal{E}$ arbitrarily instead of just obtaining the quantum sample or when we focus on certain distribution $\mathcal{D}$ instead of the worst-case $\mathcal{D}_\star$. Our result closes these open questions by giving the maximum separation in a general setting where the existence of exponential quantum advantage might be expected.

We conclude by emphasizing that only achieving an average prediction error – see Eq. (2) – is crucial in the context of Theorem 1. The situation looks drastically different if we insist on predicting $f(x)$ accurately for *all* $x \in \{0,1\}^n$. In this case, we show that exponential separations in sample complexity (N_C vs. N_Q) do occur for certain learning problems.

III. COMPUTATIONAL ASPECT

Another important figure of merit is efficiency in terms of runtime. The availability of training data can empower classical ML algorithms in a way similar to how advice provides additional power to classical computers. To capture this phenomenon rigorously, we introduce the complexity class BPP/samp. A language L of bit strings is in BPP/samp if and only if there exists probabilistic Turing machines D and M such that the following hold. For any input size n, the "sampler" D generates samples x with $|x| = n$ in polynomial time. The "learner" M takes an input x of size n along with $\mathcal{T} = \{(x_i, y_i)\}_{i=1}^{\mathrm{poly}(n)}$ of polynomial size, where x_i is sampled using Turing machine D for input size n, and y_i conveys language membership for x_i: $y_i = 1$ if $x_i \in L$ and $y_i = 0$ if $x_i \notin L$. And, we require

- The probabilistic Turing machine M to process all inputs x in polynomial time (polynomial runtime).
- For all $x \in L$, M outputs 1 with probability greater than or equal to 2/3 (prob. completeness).
- For all $x \notin L$, M outputs 1 with probability less than or equal to 1/3 (prob. soundness).

It is not hard to show that BPP is contained in BPP/samp. This inclusion is in fact strict (there are languages in BPP/samp but not in BPP). This suggests that some challenging quantum problems may be solved using classical ML models that learn from data obtained in physical experiments. However, this prospect is not without limitations. We have also show that BPP/samp is contained in P/poly.

To support the aforementioned complexity-theoretic insights, we present a sufficient condition for solving quantum problems with efficient classical ML. Fix $f(x) = \mathrm{tr}(O\,\mathcal{E}(|x\rangle\langle x|))$ and consider a training set $\mathcal{T} = \{(x_i, y_i = f(x_i))\}_{i=1}^{N_C}$. Let $k(\cdot,\cdot)$ be an efficiently computable kernel function whose associated kernel matrix $K_{ij} = k(x_i, x_j)$ is normalized ($\mathrm{tr}(K) = N_C$) and obeys the condition $s(\mathcal{T}; k) = \sum_{i=1}^{N_C}\sum_{j=1}^{N_C}(K^{-1})_{ij} f(x_i) f(x_j) \ll N_C$. Then, one could efficiently train a classical ML model to produce a function $f_C(x)$ that is guaranteed to obey

$$\mathop{\mathbb{E}}_{x\sim\mathcal{D}} |f_C(x) - \mathrm{tr}(O\,\mathcal{E}(|x\rangle\langle x|))|^2 \leq \mathcal{O}\big(\sqrt{s(\mathcal{T}; k)/N_C}\big) \ll 1, \quad \text{with high probability.} \tag{3}$$

In words, efficient classical ML can achieve very small prediction errors – even if $\mathcal{E}(|x\rangle\langle x|)$ is hard to simulate classically! The error bound (3) is not limited to classical ML, but also holds for quantum ML with an associated kernel [9, 14].

IV. CONCRETE EXAMPLE: PREDICTING GROUND STATE REPRESENTATIONS

Let us consider m-qubit local Hamiltonians $H(x)$ parameterized by some classical description $x \in [-1,1]^n$. We also assume that we do not know the function $x \mapsto H(x)$ perfectly. The goal is to produce a model $\sigma(x)$ that accurately approximates the ground state $\rho(x)$ of Hamiltonian $H(x)$; more precisely, observable properties thereof. In this case, $\mathcal{E}$ is a cooling process that prepares the ground state: $\mathcal{E}(|x\rangle\langle x|) = \rho(x)$ and we want to predict $f(x) = \mathrm{tr}\,(O\rho(x))$. We give a classical ML algorithm that comes with rigorous guarantees. The starting point is the classical shadow formalism [17]. It allows for converting randomized quantum measurement outcomes into efficient classical representations $S(\rho)$ of the underlying quantum state ρ. We use this formalism to turn ground state measurement outcomes into suitable classical training data $\mathcal{T} = \{(x_i, S(\rho(x_i))\}_{i=1}^{N_C}$ with $x_i \sim \mathrm{Unif}[-1,1]^n$. Subsequent prediction is based on an ℓ_2-Dirichlet kernel:

$$\sigma(x) = \sum\nolimits_{i=1}^{N_C} \kappa(x, x_i) S(\rho(x_i)) \quad \text{with} \quad \kappa(x,y) = \sum\nolimits_{k\in\mathbb{Z}^n} \mathbf{1}\left\{\|k\|_2 \leq \Lambda\right\} \exp\left(\mathrm{i}\pi\langle k, x-y\rangle\right). \tag{4}$$

We then use quasi-adiabatic continuation [6, 13, 20] to ensure generalization beyond training data. This guarantee is strongest if the parametrization $x \mapsto H(x)$ is sufficiently smooth and all Hamiltonians $H(x)$ have constant spectral gap. In this case, we can efficiently predict the expectation of any local observable for the ground state of new Hamiltonians $H(x)$ to small, constant error averaging over input x. The general statement reads as follows.

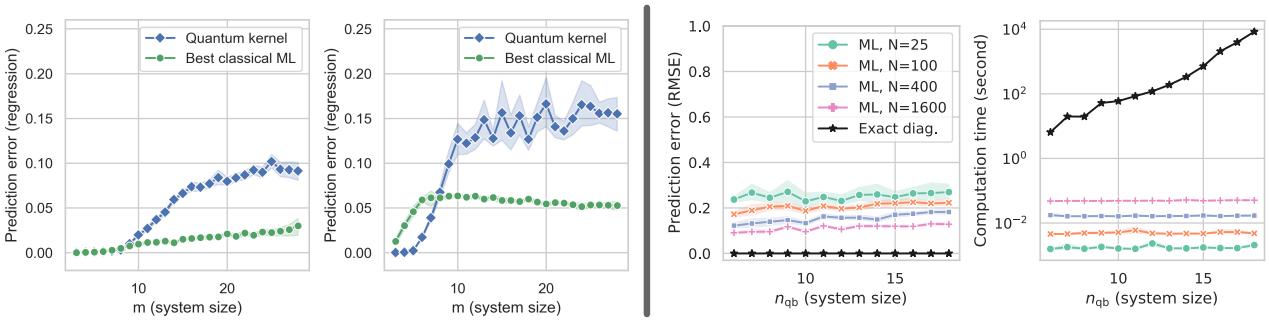

Figure 1: *Empirical performance of classical ML for quantum problems.* (Left) Predicting measurement outcomes after quantum evolutions using a training set of size $N = 600$ with comparison to the quantum kernel method [14]. (Right) Predicting ground state properties of 2D Heisenberg models with random interactions for different system size m and training data N.

Theorem 2. *Consider a smoothly parametrized class $\{H(x) : x \in [-1,1]^n\}$ of local Hamiltonians in a finite spatial dimension. Assume that each $H(x)$ has constant spectral gap. Let $O = \sum_{j=1}^{L} O_j$ be a sum of L local observables and set $B = \sum_j \|O_j\|$. We consider the data to be $\mathcal{T} = \{(x_i, S(\rho(x_i))\}_{i=1}^{N_C}$ with $x_i \sim \mathrm{Unif}[-1,1]^n$. Then, for $\epsilon > 0$,*

$$N_C = B^2 n^{\mathcal{O}(B^2/\epsilon)} \quad \text{ensures that the resulting model } \sigma(x) \text{ obeys} \quad \mathbb{E}_{x\sim[-1,1]^n} |\operatorname{tr}(O\sigma(x)) - \operatorname{tr}(O\rho(x))|^2 \leq \epsilon \tag{5}$$

with high probability. Moreover, the runtime for computing $\operatorname{tr}(O\sigma(x))$ is bounded by $\mathcal{O}((m+L)B^2 n^{\mathcal{O}(B^2/\epsilon)})$.

The classical ML actually works for more general classes of Hamiltonians. Rel. (5) holds as long as the smoothness condition $\mathbb{E}_{x\sim[-1,1]^n} \|\nabla_x \operatorname{tr}(O\rho(x))\|_2^2 \leq \mathcal{O}(B^2)$ is satisfied. Neither geometric locality nor a constant spectral gap is genuinely needed. At this level of generality, the sample complexity is tight up to logarithmic factors. We prove the following (almost) matching lower bound.

Theorem 3. *Consider a quantum ML model that learns from N_Q quantum data $\{x_i, \rho(x_i)\}_{i=1}^{N_Q}$ to generate $\sigma(x)$. Suppose the quantum ML model is strong enough to achieve small prediction error $\mathbb{E}_{x\sim[-1,1]^n} |\operatorname{tr}(O\sigma(x)) - \operatorname{tr}(O\rho(x))|^2 \leq \epsilon$. Then,*

$$N_Q \geq B^2 n^{\Omega(B^2/\epsilon)}/\log(B) \quad \text{and the quantum runtime is lower bounded by } B^2 n^{\Omega(B^2/\epsilon)}/\log(B). \tag{6}$$

Theorem 3 also lower-bounds the optimal sample complexity of classical ML models. Invoking (the contrapositive of) Theorem 1 then implies that even a more powerful quantum ML model with explicit access to the cooling process $\mathcal{E}$ for preparing ground states cannot perform substantially better. Hence, there can not be significant quantum advantages in the general setting. We also note that existing classical algorithms for solving ground state [1, 2, 18], which do not learn from data, only work for a spatial dimension less than two, while Theorem 2 applies to any finite spatial dimension.

V. NUMERICAL EXPERIMENTS

a. Predicting outcomes after a quantum evolution: Suppose that the unknown CPTP map implements a unitary evolution $\rho_{\mathrm{init}}(x) \mapsto \mathcal{E}(\rho_{\mathrm{init}}(x)) = U\rho_{\mathrm{init}}(x)U^\dagger$ and the state $\rho_{\mathrm{init}}(x)$ is parametrized by a classical input $x \in [-1,1]^n$. The goal is to predict the resulting expectation value of a local observable: $f(x) = \operatorname{tr}\left(OU\rho_{\mathrm{init}}(x)U^\dagger\right)$. We compare classical ML procedures with a promising quantum ML model [14]. Based on the quantum kernel $k(x_i, x_j) = \operatorname{tr}(\rho_{\mathrm{init}}(x_i)\rho_{\mathrm{init}}(x_j))$, this quantum model is known to yield rigorous advantages for certain problems based on discrete logarithm [19]. Fig. 1(a) depicts a numerical comparison up to 28 qubits. The two plots correspond to two different encoding procedures $x \mapsto \rho_{\mathrm{init}}(x)$. We see that the best classical ML achieves small prediction error and even outperforms the quantum ML model. This discrepancy is particularly pronounced for large system sizes. We refer to manuscript [15] for additional theoretical underpinning. The manuscript also supplies necessary conditions for quantum advantages in quantum problems.

b. Predicting ground state properties of the 2D Heisenberg model: Suppose that the classical input $x = \vec{J}$ describes the coupling constants in the 2D Hamiltonian $H(\vec{J}) = \sum_{\langle ij\rangle} J_{ij}(X_iX_j + Y_iY_j + Z_iZ_j)$ on n qubits. We consider the training of classical ML models to predict the expectation of local energy $X_iX_j + Y_iY_j + Z_iZ_j$ in the ground state of $H(\vec{J})$. Fig. 1(b) compares the results to exact diagonalization in terms of both prediction error and computation time. Perhaps surprisingly, the runtime is nearly independent of the system size m, and only depends on the training data size N. There are also substantial runtime savings comparing to exact diagonalization.

VI. CONCLUSION

We have collected rigorous and empirical evidence that classical ML for quantum problems can be more powerful than one might expect. From an optimistic perspective, these results provide the foundation for training classical ML models to predict properties in unexplored quantum systems. This could be relevant for designing novel materials, catalysts, and pharmaceuticals, where existing data are already available to train the ML models. From a pessimistic perspective, these findings narrow the scope for quantum advantages in ML problems. One should not take quantum advantages for granted – even if the underlying problem has a quantum origin. The search for interesting quantum problems whose solutions come with genuine quantum advantages might be even more important (and less trivial) than previously anticipated.

[1] N. Abrahamsen. A polynomial-time algorithm for ground states of spin trees, 2020.

[2] I. Arad, Z. Landau, U. Vazirani, and T. Vidick. Rigorous rg algorithms and area laws for low energy eigenstates in 1d. *Commun. Math. Phys.*, 356(1):65–105, 2017.

[3] S. Arunachalam and R. de Wolf. Optimal quantum sample complexity of learning algorithms. *arXiv preprint arXiv:1607.00932*, 2016.

[4] S. Arunachalam and R. de Wolf. Guest column: A survey of quantum learning theory. *ACM SIGACT News*, 48(2):41–67, 2017.

[5] S. Arunachalam, A. B. Grilo, and H. Yuen. Quantum statistical query learning. *arXiv preprint arXiv:2002.08240*, 2020.

[6] S. Bachmann, S. Michalakis, B. Nachtergaele, and R. Sims. Automorphic equivalence within gapped phases of quantum lattice systems. *Commun. Math. Phys.*, 309(3):835–871, 2012.

[7] P. L. Bartlett and S. Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *J Mach Learn Res*, 3(Nov):463–482, 2002.

[8] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

[9] C. Blank, D. K. Park, J.-K. K. Rhee, and F. Petruccione. Quantum classifier with tailored quantum kernel. *npj Quantum Inf.*, 6(1):1–7, 2020.

[10] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the vapnik-chervonenkis dimension. *J. ACM*, 36(4):929–965, 1989.

[11] G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[12] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. *arXiv preprint arXiv:1704.01212*, 2017.

[13] M. B. Hastings and X.-G. Wen. Quasiadiabatic continuation of quantum states: The stability of topological ground-state degeneracy and emergent gauge invariance. *Phys. Rev. B*, 72(4):045141, 2005.

[14] 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*, 567(7747):209–212, 2019.

[15] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean. Power of data in quantum machine learning. *arXiv preprint arXiv:2011.01938*, 2020.

[16] H.-Y. Huang, R. Kueng, and J. Preskill. Provable machine learning algorithms for quantum many-body problems. *to appear soon.*

[17] H.-Y. Huang, R. Kueng, and J. Preskill. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.*, 16:1050—1057, 2020.

[18] Z. Landau, U. Vazirani, and T. Vidick. A polynomial time algorithm for the ground state of one-dimensional gapped local hamiltonians. *Nat. Phys.*, 11(7):566–569, 2015.

[19] Y. Liu, S. Arunachalam, and K. Temme. A rigorous and robust quantum speed-up in supervised machine learning, 2020.

[20] T. J. Osborne. Simulating adiabatic evolution of gapped spin systems. *Phys. Rev. A*, 75(3):032321, 2007.

[21] M. Schuld and N. Killoran. Quantum machine learning in feature hilbert spaces. *Phys. Rev. Lett.*, 122(4):040504, 2019.

[22] R. A. Servedio and S. J. Gortler. Equivalences and separations between quantum and classical learnability. *SIAM J. Comput.*, 33(5):1067–1092, 2004.

[23] L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.

[24] E. P. Van Nieuwenburg, Y.-H. Liu, and S. D. Huber. Learning phase transitions by confusion. *Nat. Phys.*, 13(5):435–439, 2017.

[25] V. Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[26] C. Zhang. An improved lower bound on query complexity for quantum pac learning. *Inf. Process. Lett.*, 111(1):40–45, 2010.