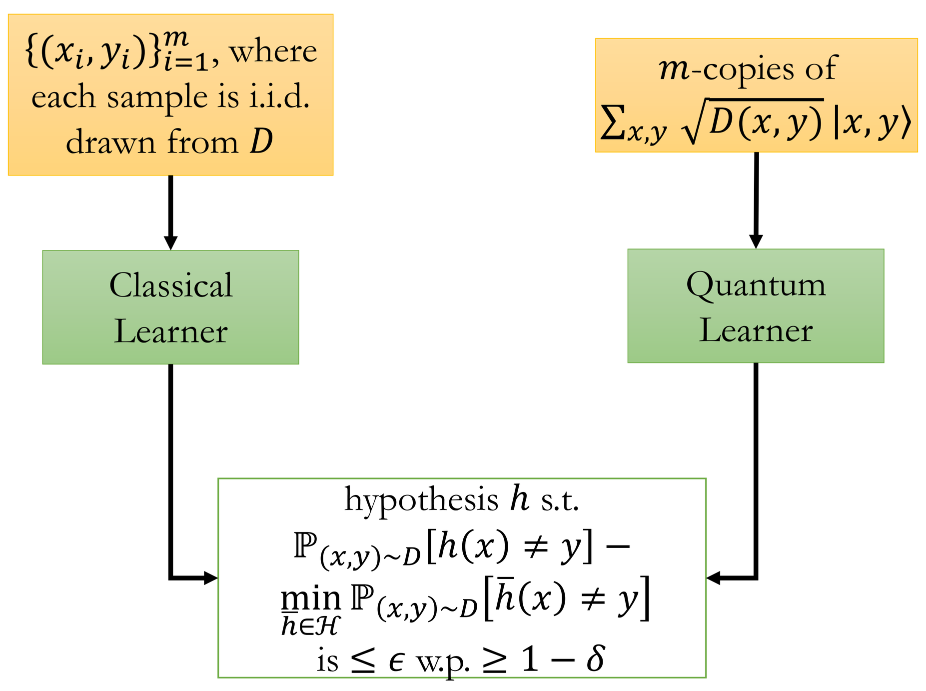
Quantum Learning Theory Beyond Batch Binary Classification

Preetham Mohan* and Ambuj Tewari†‡

{*Department of Mathematics, †Department of Statistics, ‡Department of Electrical Engineering and Computer Science},
University of Michigan – Ann Arbor

Abstract: [AdW18] showed that the sample complexity of quantum batch learning of boolean functions, in the realizable and agnostic settings, has the same form and order as the corresponding classical sample complexities. Here, we show that this, ostensibly surprising, message extends to batch multiclass learning, online boolean learning, and online multiclass learning. For our online learning results, we first introduce the classical adversary-provides-a-distribution model (an adaptive adversary variant of the classical model of [DT22]). Then, we introduce the first (to the best of our knowledge) model of online learning with quantum examples.

Batch Learning



Performance Measure: Sample complexity, i.e., the smallest m that works for all D

Instance/input space, instance/input	$\mathcal{X}, x \in \mathcal{X}$
Label space, label	$\mathcal{Y}, y \in \mathcal{Y}$
Hypothesis class, hypothesis	$\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}, h \in \mathcal{H}$
Boolean setting	$\mathcal{Y} = \{0,1\}$
Multiclass setting	$\lvert\mathcal{Y}\rvert = k > 2$
Realizable	$y = h^*(x)$, for some $h^* \in \mathcal{H}$
Quantum (realizable) sample	$\sum_x \sqrt{D(x)}\,\lvert x, h^*(x)\rangle$, where D is on the input space
Agnostic (*poster default*)	y need not come from some $h^* \in \mathcal{H}$
Quantum (agnostic) sample	$\sum_{x,y} \sqrt{D(x,y)}\,\lvert x,y\rangle$, where D is on the input-label space
$\Leftrightarrow$	Models are equivalent up to constants under the appropriate performance measure
Loss class	$\mathcal{L} \circ \mathcal{H} = \{\mathbf{1}[h(x) \neq y] : h \in \mathcal{H}, x \in \mathcal{X}, y \in \mathcal{Y}\}$, i.e., $\mathcal{L} \circ \mathcal{H}$ captures the 0-1 loss behavior of $\mathcal{H}$

Quantum Batch Multiclass Learning

Result: Quantum batch multiclass learning $\Leftrightarrow$ Classical batch multiclass learning (when k is finite)

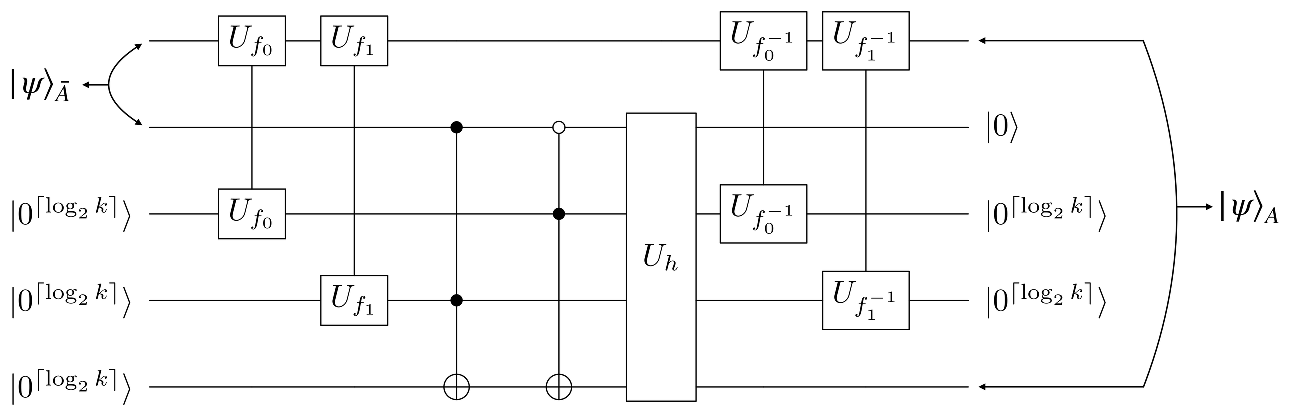
Upper bounds: Measure-and-learn-classically

Lower bounds: Reduction to the quantum batch boolean setting

- A learner A for a multiclass hypothesis class $\mathcal{H}$ implies a learner $\bar{A}$ for a boolean hypothesis class $\mathcal{H}_d$, where VCdim($\mathcal{H}_d$) = Ndim($\mathcal{H}$) = d.

Key: There exists a quantum circuit which can perform the following transformation of $\bar{A}$'s boolean quantum samples to A's multiclass quantum samples:

$|\Psi\rangle_{\bar{A}} = \sum_{x\in[d]} \sqrt{D(x)}|x,y\rangle \rightarrow \sum_{x\in[d]} \sqrt{D(x)}|x, f_y(x)\rangle = |\Psi\rangle_A$, where $y \in \{0,1\}$, and $f_0, f_1 : [d] \rightarrow [k]$.



Crucially, the circuit's existence hinges on the reversibility of the transformation $y \leftrightarrow f_y$, which is guaranteed precisely due to the definition of N-shattering.

- In particular, U_h computes a "genuine" boolean function h, which takes as input $f_0(x)$, $f_1(x)$, $f_y(x)$ and outputs y. If $f_0(x)$ and $f_1(x)$ were not distinct for *all* inputs x, this would be impossible.

		Classical	Quantum
Boolean	Realizable	$\Theta\left(\frac{\text{VCdim}(\mathcal{H})+\log(\frac{1}{\delta})}{\epsilon}\right)$	$\Theta\left(\frac{\text{VCdim}(\mathcal{H})+\log(\frac{1}{\delta})}{\epsilon}\right)$
	Agnostic	$\Theta\left(\frac{\text{VCdim}(\mathcal{H})+\log(\frac{1}{\delta})}{\epsilon^2}\right)$	$\Theta\left(\frac{\text{VCdim}(\mathcal{H})+\log(\frac{1}{\delta})}{\epsilon^2}\right)$
Multiclass	Realizable	$\Omega\left(\frac{\text{Ndim}(\mathcal{H})+\log(\frac{1}{\delta})}{\epsilon}\right)$ $\mathcal{O}\left(\frac{\text{Ndim}(\mathcal{H})\log(k)\log(\frac{1}{\epsilon})+\log(\frac{1}{\delta})}{\epsilon}\right)$	$\Omega\left(\frac{\text{Ndim}(\mathcal{H})+\log(\frac{1}{\delta})}{\epsilon}\right)$ $\mathcal{O}\left(\frac{\text{Ndim}(\mathcal{H})\log(k)\log(\frac{1}{\epsilon})+\log(\frac{1}{\delta})}{\epsilon}\right)$
	Agnostic	$\Omega\left(\frac{\text{Ndim}(\mathcal{H})+\log(\frac{1}{\delta})}{\epsilon^2}\right)$, $\mathcal{O}\left(\frac{\text{Ndim}(\mathcal{H})\log(k)+\log(\frac{1}{\delta})}{\epsilon^2}\right)$	$\Omega\left(\frac{\text{Ndim}(\mathcal{H})+\log(\frac{1}{\delta})}{\epsilon^2}\right)$ $\mathcal{O}\left(\frac{\text{Ndim}(\mathcal{H})\log(k)+\log(\frac{1}{\delta})}{\epsilon^2}\right)$

Combinatorial Parameters

Combinatorial parameters measure the expressiveness of a hypothesis class $\mathcal{H}$ when adapted to a particular model.

- VC dimension (VCdim) – size d of the largest subset of $\mathcal{X}$ that can be labeled in all 2^d possible ways by a boolean hypothesis class $\mathcal{H}$.
- Natarajan dimension (Ndim) – size d of the largest subset S_d of $\mathcal{X}$ that can be labeled by a multiclass hypothesis class $\mathcal{H}$ in each of the 2^d ways prescribed by two distinct functions $f_0, f_1 : S_d \rightarrow [k]$.
- Littlestone dimension (Ldim)/multiclass Littlestone dimension (mcLdim) – maximum depth d of a complete binary tree, with nodes consisting of elements from $\mathcal{X}$ and edges consisting of distinct elements from $\mathcal{Y}$, such that each path is correctly labeled (in its entirety) by some hypothesis in $\mathcal{H}$.

Summary of Online Learning Results

Overall Message: There is limited power in quantum examples to speed up learning especially when the adversary is allowed to play arbitrary distributions (including very degenerate ones like point masses).

		Canonical Classical (Input-based)	Classical (Distribution-based)	Quantum
Boolean	Realizable	$\Theta(\text{Ldim}(\mathcal{H}))$	$\Theta(\text{Ldim}(\mathcal{H}))$	$\Theta(\text{Ldim}(\mathcal{H}))$
	Agnostic	$\Theta(\sqrt{\text{Ldim}(\mathcal{H})T})$	$\Theta(\sqrt{\text{Ldim}(\mathcal{H})T})$	$\Theta(\sqrt{\text{Ldim}(\mathcal{H})T})$
Multiclass	Realizable	$\Theta(\text{mcLdim}(\mathcal{H}))$	$\Theta(\text{mcLdim}(\mathcal{H}))$	$\Theta(\text{mcLdim}(\mathcal{H}))$
	Agnostic	$\Omega(\sqrt{\text{mcLdim}(\mathcal{H})T})$ $\tilde{\mathcal{O}}(\sqrt{\text{mcLdim}(\mathcal{H})T})$	$\Omega(\sqrt{\text{mcLdim}(\mathcal{H})T})$ $\mathcal{O}(\sqrt{\text{mcLdim}(\mathcal{H})T\log(Tk)})$	$\Omega(\sqrt{\text{mcLdim}(\mathcal{H})T})$ $\mathcal{O}(\sqrt{\text{mcLdim}(\mathcal{H})T\log(Tk)})$

Open Problems

1. What is the tight quantum sample complexity bound for batch multiclass learning, in both the realizable and agnostic settings, when the label space is unbounded (i.e., when the number of classes $k \rightarrow \infty$)?
2. What is the tight expected regret bound for quantum online multiclass agnostic learning when the label space is unbounded (i.e., when $k \rightarrow \infty$)?
3. What happens when we impose restrictions on D_t to force it away from a point mass? Would the expected regret bound for the canonical classical online model, classical adversary-provides-a-distribution model, and the quantum online model all diverge from one another?

Canonical Classical Online Learning

Model: The adversary and the learner play the following T-round "game", where at each round t:

Adversary provides x_t → Learner predicts $h_t(x_t)$ → Adversary reveals y_t → Learner suffers a loss of 1 (a 'mistake'), if $h_t(x_t) \neq y_t$

Performance Measure: Expected Regret, i.e., the smallest

$\mathbb{E}\left[\sum_{t=1}^T \mathbf{1}[h_t(x_t) \neq y_t] - \min_{\bar{h}\in\mathcal{H}} \sum_{t=1}^T \mathbf{1}[\bar{h}(x_t) \neq y_t]\right]$ under an arbitrary (worst-case) adversarial sequence of inputs $\{x_t\}_{t=1}^T$

Can we obtain a quantum generalization?

- Issue: Quantum adversary cannot temporally split its provision of x_t and y_t
  - Easy Fix: Reordering provides an equivalent model

Learner provides h_t → Adversary reveals (x_t, y_t) → Learner suffers a loss of 1 (a 'mistake'), if $h_t(x_t) \neq y_t$

- **Main Issue:** Incompatibility between a *single* classical example and a quantum example that, in general, will sit in superposition
  - Towards a Fix: Introducing the intermediary Classical Adversary-provides-a-distribution Model

Classical Adversary-provides-a-distribution Model

Model: The adversary and the learner play the following T-round "game", where at each round t:

Learner provides h_t → Adversary chooses D_t and reveals $(x, y) \sim D_t$ → Learner suffers, but does not see, a loss of $\mathbb{P}_{(x,y)\sim D_t}[h_t(x) \neq y]$

Performance Measure: Expected Regret, i.e., the smallest

$\mathbb{E}\left[\sum_{t=1}^T \mathbb{P}_{(x,y)\sim D_t}[h_t(x) \neq y] - \min_{\bar{h}\in\mathcal{H}} \sum_{t=1}^T \mathbb{P}_{(x,y)\sim D_t}[\bar{h}(x) \neq y]\right]$ under an arbitrary (worst-case) adversarial sequence of distributions $\{D_t\}_{t=1}^T$

Result: Adversary-provides-a-distribution $\Leftrightarrow$ Canonical online learning (in all-but-one cases).

Upper bounds: If the best canonical classical online learner plays this new game, in all-but-one cases, its expected regret only suffers at most a constant factor increase in overhead.

- The exception is the multiclass agnostic case with unbounded label space (i.e., when $k \rightarrow \infty$). The discrepancy arises due to the provable gap [HMR+23] in the rate of canonical online learnability and the rate of online uniform convergence of the loss class in the multiclass case.
  - However, it is unknown whether there exists an algorithm for adversary-provides-a-distribution that would help establish the equivalence.

Lower bounds: An adversary that solely plays point mass distributions $\{D_t\}_{t=1}^T$ provides a reduction to canonical classical online learning.

Quantum Online Learning

Model: The quantum adversary and the quantum learner play a T-round "game", where at each round t:

Learner provides h_t → Adversary provides $\sum_{x,y} \sqrt{D_t(x,y)}\,|x,y\rangle$ → Learner suffers, but does not see, a loss of $\mathbb{P}_{(x,y)\sim D_t}[h_t(x) \neq y]$

Performance Measure: Expected Regret, i.e., the smallest

$\mathbb{E}\left[\sum_{t=1}^T \mathbb{P}_{(x,y)\sim D_t}[h_t(x) \neq y] - \min_{\bar{h}\in\mathcal{H}} \sum_{t=1}^T \mathbb{P}_{(x,y)\sim D_t}[\bar{h}(x) \neq y]\right]$ under an arbitrary (worst-case) adversarial sequence of quantum examples

Result: Quantum online learning $\Leftrightarrow$ Classical adversary-provides-a-distribution $\Leftrightarrow$ Canonical classical online learning (in all cases, except the multiclass agnostic setting when $k \rightarrow \infty$).

Upper bounds: A naïve measure-and-learn-classically "quantum" learner

- In the multiclass agnostic case when $k \rightarrow \infty$, although the upper bound for adversary-provides-a-distribution naturally extends to quantum online learning, it is unknown whether the models are equivalent (this would involve investigating into a "genuine" quantum online learning algorithm).

Lower bounds: Reduction to classical online learning

- At each round, a classical learner can state prepare $|x_t, y_t\rangle$ from the revealed sample (x_t, y_t), pass it as input to a quantum learner and use the quantum learner's output hypotheses. Therefore, classical online lower bounds $\Longrightarrow$ quantum online lower bounds.

References

[AdW18] Srinivasan Arunachalam and Ronald de Wolf (JMLR 2018). Optimal quantum sample complexity of learning algorithms.

[DT22] Philip Dawid and Ambuj Tewari (HDSR 2022). On learnability under general stochastic processes.

[HMR+23] Steve Hanneke, Shay Moran, Vinod Raman, Unique Subedi, and Ambuj Tewari (COLT 2023). Multiclass online learning and uniform convergence.

(Our paper) [MT23] Preetham Mohan and Ambuj Tewari (arXiv preprint). Quantum learning theory beyond batch binary classification.

Scan QR for arXiv link
arXiv:2302.07409