Ph 220: Quantum Learning Theory – Lecture Note 3: Learning and Predicting Properties of a Quantum State

Hsin-Yuan Huang (Robert)

Caltech

In this lecture, we explore the fundamental problems of learning and predicting properties of an unknown quantum state. We begin by looking at an extremely simple approach for learning the full description of a quantum state. We will then study random unitaries and randomized measurements to develop a better approach for learning and predicting properties of quantum states.

1 A Simple Approach for Learning a Quantum State

Given access to multiple copies of an n-qubit quantum state ρ, our goal is to produce a classical description, $\hat{\rho}$, that is close to the true state ρ.

Q: How do we obtain a full classical description of an n-qubit quantum state ρ?

To represent the $2^n \times 2^n$ density matrix ρ, we can express it in a convenient basis. The set of n-qubit Pauli operators, $\{I, X, Y, Z\}^{\otimes n}$, forms an orthonormal basis for the space of Hermitian matrices under the Hilbert-Schmidt inner product, $\langle A, B\rangle = \mathrm{tr}(A^\dagger B)$. Any state ρ can be uniquely decomposed as:

$$\rho = \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} \alpha_P \frac{P}{2^n}$$

where the Pauli coefficients $\alpha_P = \mathrm{tr}(P\rho)$ are the real expectation values of the corresponding Pauli operators. Since $\mathrm{tr}(\rho) = 1$, the coefficient for the identity Pauli string $P = I^{\otimes n}$ is always $\alpha_I = 1$. The problem of learning the state ρ is therefore equivalent to learning the remaining $4^n - 1$ unknown Pauli coefficients, which are all between -1 and $+1$.

Algorithm

A straightforward way to learn these coefficients is to measure the expectation value of each Pauli operator directly. For each Pauli operator $P \in \{I, X, Y, Z\}^{\otimes n}$ (excluding the identity):

1. Prepare a copy of the state ρ.
2. Measure the observable P. This is done by measuring each qubit i in the basis corresponding to the single-qubit Pauli operator P_i. If $P_i = X$, we measure in the $\{|+\rangle, |-\rangle\}$ basis; if $P_i = Y$, we measure in the $\{|y+\rangle, |y-\rangle\}$ basis; if $P_i = Z$ or I, we measure in the $\{|0\rangle, |1\rangle\}$ basis.
3. The measurement yields an outcome $b_i \in \{+1, -1\}$ for each qubit. The overall outcome is the product $b_1 b_2 \cdots b_n$, which is an eigenvalue of P.

4. Repeat the above steps k times and average the results to obtain an estimate $\hat{\alpha}_P$ for the true expectation value $\alpha_P = \text{tr}(P\rho)$.

After estimating all the coefficients, we construct an estimate of the state: $\hat{\rho} = \sum_P \hat{\alpha}_P \frac{P}{2^n}$, where we set $\hat{\alpha}_I = 1$. Our goal is to determine the total number of copies of ρ (the sample complexity) required to ensure that our estimate $\hat{\rho}$ is close to ρ with high probability.

2 Sample Complexity Analysis

Let's analyze the number of samples needed for this simple tomography scheme. Each measurement of a Pauli operator P yields an outcome in $\{+1, -1\}$. The average of k such outcomes, $\hat{\alpha}_P$, can be related to its expectation α_P using Hoeffding's inequality:

$$\Pr\left[|\hat{\alpha}_P - \alpha_P| \geq \varepsilon\right] \leq 2e^{-k\varepsilon^2/2}$$

To reconstruct the full state, we need to estimate all $4^n - 1$ coefficients accurately at the same time. Using the union bound, the probability that *any* of our estimates is off by more than ε is at most $4^n \cdot 2e^{-k\varepsilon^2/2}$. To make this failure probability smaller than some δ, the number of measurements *per Pauli basis* must be:

$$k = O\left(\frac{\log(4^n/\delta)}{\varepsilon^2}\right) = O\left(\frac{n + \log(1/\delta)}{\varepsilon^2}\right)$$

Next, we relate the coefficient error ε to the error in the final state estimate, measured by the squared Frobenius norm:

$$\|\hat{\rho} - \rho\|_F^2 = \text{tr}\left(\left(\sum_P (\hat{\alpha}_P - \alpha_P)\frac{P}{2^n}\right)^2\right) = \frac{1}{2^n}\sum_P (\hat{\alpha}_P - \alpha_P)^2$$

If we guarantee that $|\hat{\alpha}_P - \alpha_P| \leq \varepsilon$ for all P, the squared Frobenius norm is bounded by $\|\hat{\rho} - \rho\|_F^2 \leq \frac{1}{2^n}\sum_P \varepsilon^2 = 2^n\varepsilon^2$. To achieve a final state error of $\|\hat{\rho} - \rho\|_F \leq \epsilon$, we must set the precision for each coefficient to $\varepsilon \leq \epsilon/\sqrt{2^n}$. Substituting this into our expression for k, the total sample complexity $N = 4^n \cdot k$ becomes:

$$N = 4^n \cdot O\left(\frac{(n + \log(1/\delta))}{(\epsilon/\sqrt{2^n})^2}\right) = O\left(\frac{(n + \log(1/\delta))8^n}{\epsilon^2}\right)$$

This exponential scaling in 8^n makes the protocol prohibitively expensive for even a small number of qubits, motivating the search for more efficient methods.

3 From State Tomography to Predicting Observables

A key operational meaning of "learning a state" is the ability to accurately predict the expectation values of arbitrary observables. Let O be an observable with eigenvalues bounded between -1 and $+1$ (i.e., its operator norm $\|O\|_\infty \leq 1$). The error in predicting $\text{tr}(O\rho)$ using our learned state $\hat{\rho}$ is:

$$|\text{tr}(O\hat{\rho}) - \text{tr}(O\rho)| = |\text{tr}(O(\hat{\rho} - \rho))|$$

By Hölder's inequality for matrices, this error is bounded by the trace norm (Schatten 1-norm) of the state error:

$$|\text{tr}(O(\hat{\rho} - \rho))| \leq \|O\|_\infty \|\hat{\rho} - \rho\|_1 \leq \|\hat{\rho} - \rho\|_1$$

This shows that to bound the prediction error for *any* such observable, we must bound the trace distance between $\hat{\rho}$ and ρ. We can relate the trace norm to the Frobenius norm using the inequality $\|\cdot\|_1 \leq \sqrt{d}\|\cdot\|_F$ for matrices of dimension $d = 2^n$. This gives:

$$\|\hat{\rho} - \rho\|_1 \leq \sqrt{2^n}\|\hat{\rho} - \rho\|_F$$

To guarantee a prediction error of at most ϵ for any bounded observable, we need $\|\hat{\rho} - \rho\|_1 \leq \epsilon$. This requires us to achieve a Frobenius norm error of $\|\hat{\rho} - \rho\|_F \leq \epsilon/\sqrt{2^n}$. Plugging this new, more stringent error target into our sample complexity formula yields:

$$N = O\left(\frac{(n + \log(1/\delta))8^n}{(\epsilon/\sqrt{2^n})^2}\right) = O\left(\frac{(n + \log(1/\delta))8^n \cdot 2^n}{\epsilon^2}\right) = O\left(\frac{(n + \log(1/\delta))16^n}{\epsilon^2}\right)$$

The sample complexity to learn the state well enough for universal prediction is even worse, scaling as 16^n. This underscores the extreme inefficiency of this simple approach for practical tasks.