Quantum learning theory and sample complexity

Compare quantum learning models by their examples, queries, accuracy and confidence guarantees.

Research 22 min lesson Dr. Charith Jayasekara

By the end, you'll be able to

- Calculate prediction error for a finite threshold class.
- Distinguish sample complexity from computation time.
- Identify when coherent examples change the access model.

Ask what the learner is allowed to see

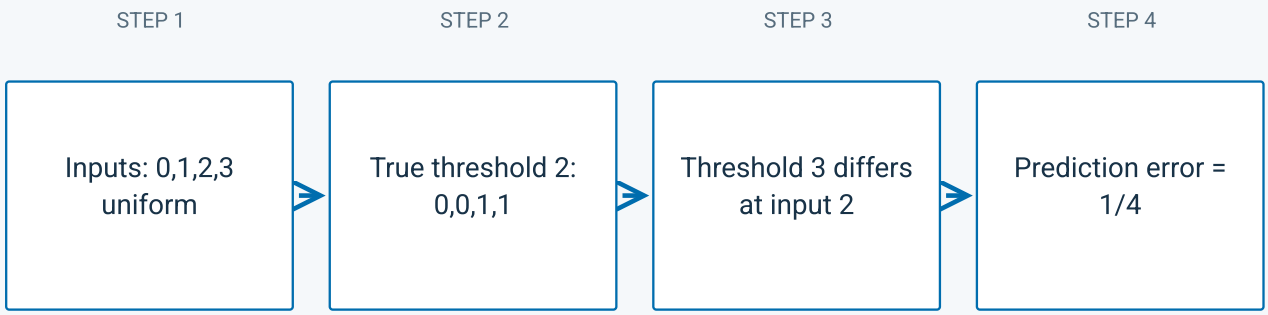
Learning theory studies what can be inferred from examples or queries, with guarantees about accuracy and confidence. Quantum learning theory changes the learner's resources: perhaps examples arrive as coherent states, perhaps the learner can query a target function in superposition, or perhaps the object being learned is itself quantum. These are different models, and they can have very different advantages and limitations.

The word "sample" needs particular care. A classical labelled example is an ordinary pair of an input and its label. A quantum example can be a coherent superposition weighted by an input distribution. Creating that state from an existing classical dataset is not free. A theorem granting quantum examples studies the power of that resource, not necessarily the cost of acquiring it in an application.

Review quantum information, measurement and oracle assumptions. The basic statistical distinction is between fitting observed data and making accurate predictions on unseen examples from the specified distribution.

WORKED DIAGRAM · AA20

Observed data can leave several hypotheses



Under the uniform distribution, threshold 3 has error 1/4 relative to threshold 2. If only endpoints 0 and 3 are observed, several thresholds fit; computation alone does not supply the missing evidence.

Define accuracy and confidence separately

In probably approximately correct learning, choose an error tolerance epsilon and failure probability delta. The learner should, with probability at least one minus delta, output a hypothesis whose prediction error is at most epsilon under the input distribution. Epsilon concerns the quality of the returned model. Delta concerns how often the whole learning procedure may fail to achieve that quality.

Agnostic learning allows labels that do not exactly match any hypothesis in the chosen class. The objective is then to compete with the best hypothesis in that class, within an additional error tolerance. This is a different promise from a perfectly realisable target. Noise, misspecification and distribution shift should not be hidden by presenting only the easier setting.

Work a small hypothesis example

Let inputs be the numbers zero, one, two and three, equally likely. Consider threshold hypotheses that label an input one when it is at least a chosen threshold. Suppose the true threshold is two, so labels are zero, zero, one, one. If the learner outputs threshold three, it disagrees only at input two and has error one quarter under the uniform distribution.

If the distribution instead places almost all its weight on input zero, that same wrong threshold can have very small prediction error. Learning guarantees are distribution-dependent statements about mistakes on future examples, not simply the fraction of labels in a truth table that were reconstructed. Exact identification and approximate prediction are different tasks.

Now suppose the learner sees examples only at zero and three. Both thresholds one, two and three fit those observations. More computation cannot identify the missing boundary without additional information or a modelling assumption. Quantum processing may change how information is queried, but it does not erase the need to specify what evidence is available.

Prediction error is probability mass on disagreements, rather than a raw count of unknown labels.

$$\mathrm{err}_D(h) = \Pr_{X \sim D}[h(X) \neq h_*(X)], \qquad h_t(x) = [x \geq t], \qquad \mathrm{err}_{\mathrm{uniform}}(h_3) = \Pr[X = 2] = \frac{1}{4} \quad (h_* = h_2).$$

D is the future-input distribution and h_* the true label rule. Threshold three differs from threshold two only at input two. A different distribution changes that error even though the hypotheses themselves are unchanged.

Query complexity and sample complexity differ

A membership-query learner chooses an input and asks for its label. A passive learner receives examples drawn from a distribution. A coherent query learner may access the label function in superposition. These access models can produce different complexity separations. A strong advantage in one model should not be advertised as a universal reduction in the number of ordinary training records.

Arunachalam and de Wolf's research on quantum sample complexity shows that, for the distribution-independent PAC and agnostic models they analyse, quantum and classical sample complexities match up to constant factors. This leaves room for differences in computational time or in other learning models, but rules out a blanket exponential sample reduction in those settings. Read the primary sample-complexity paper.

The broader model landscape is described in their author-written account of quantum learning theory. It distinguishes exact query learning, PAC learning and agnostic learning rather than merging them under one machine-learning label. Read the learning-theory account.

Learning a quantum state is another task

If the unknown object is a density matrix, full tomography asks for a classical description sufficient to reconstruct it to a chosen error. Other tasks ask only for predictions of selected observables. The latter can need much less output information than a full matrix. State the target measurements, allowed copies and error metric before comparing sample requirements.

A quantum state cannot generally be read as a list of amplitudes from one copy. Measurements produce limited outcomes and can disturb the specimen. Repeated identically prepared copies are therefore a resource. A simulator that exposes all amplitudes provides information unavailable to a learner receiving one unknown physical state.

Follow a fair learning analysis

Define the hypothesis class, data distribution or distribution family, target loss and access model. Separate training examples from evaluation examples. State epsilon, delta and any realisability or noise assumptions. Count both examples and computational operations, including the cost of producing coherent examples or queries. Finally compare learners with matched access and output requirements.

For empirical work, keep model selection separate from final testing. A quantum circuit can overfit a small dataset just as a classical model can. Repeatedly tuning against the test set invalidates an apparent generalisation estimate. Statistical soundness is a property of the experiment design, not of whether the model uses qubits.

What a useful advantage claim looks like

A precise claim might say that a specified concept class is learned with fewer coherent membership queries under a particular promise. Another might show improved computational time when quantum examples are already supplied. A practical claim would include encoding, state preparation, measurement and classical optimisation costs. These are meaningful but distinct results.

Avoid the idea that superposition lets a learner inspect every training example simultaneously and retain every answer. A coherent query transforms a state; its information must still be extracted through an algorithm and measurements. The right question is which property becomes observable with the available resources.

Practice with answers

Exercise 1. Under the uniform four-input example, what error does threshold one make when the true threshold is two?

<details>
<summary>Check the worked answer</summary>

It mislabels only input one, so its error is one quarter. Under another distribution, the error is the probability weight on that input, not necessarily one quarter.
</details>

Exercise 2. Does equal quantum and classical PAC sample complexity imply equal computational runtime for every learning problem?

Check the worked answer

No. Sample complexity counts the information examples needed, while runtime counts computation performed with those examples. The theorem also concerns specified models; other access assumptions or output tasks can have different behaviour.

Check your understanding

PAUSE & THINK

Does equal sample complexity imply equal runtime for every learner?

- [ ] Yes
- [ ] No; information requirements and computational effort are different
- [ ] Only with quantum examples

Answer: No; information requirements and computational effort are different. Sample complexity counts examples. Algorithms can differ in processing time, and other access models can have different sample guarantees.

Sources

- Read the primary sample-complexity paper
- Read the learning-theory account