Quantum Learning Algorithms for Decision Trees with Optimal Bounds

Sagnik Chatterjee*

Indraprastha Institute of Information Technology (IIIT-D), Delhi, India

1 Summary

The 21st century has seen the adoption of machine learning and artificial intelligence on an unprecedented and unexpected level. With widespread use, however, comes the threat of rampant misuse. Large-scale predictive and generative models are essentially black boxes impermeable to scrutiny, which can lead to serious concerns over security, lack of transparency, and fairness in the generated or predicted data. Interpretable or *human-explainable* machine learning models are critical to AI trust moving forward [Miller, 2019, Rudin, 2019, Rudin et al., 2022].

Decision trees are a popular class of non-parametric supervised machine learning algorithms used for both classification and regression tasks that are canonical examples in explainable machine learning. The recent survey by Rudin et al. [2022] lists decision tree learning as one of ten grand challenges in interpretable machine learning. Many decision tree learning heuristics exist, like ID3, C4.5, and CART, which have enjoyed decades of empirical success. These algorithms, however, have little in the way of theoretical guarantees, which, unfortunately, puts them in the same boat as deep neural networks.

Recently, it was shown that there are specific barriers to learning decision trees and their variants [Koch et al., 2023, Bshouty, 2023] under certain restrictions on the decision tree learning algorithms. In the forthcoming sections, I will elaborate on a plan to circumvent these hardness results by leveraging a more general class of decision tree learning algorithms, which have proved notoriously hard to lower bound. I will further explain how to obtain improved theoretical guarantees for decision tree learning and testing using quantum algorithms. Together with existing upper bounds, this will give us a clearer view of obtaining theoretically robust algorithms for decision trees.

2 Background

Bshouty [1993] showed that decision trees are universal for Boolean functions, i.e., any Boolean function is learnable as a decision tree, which makes decision tree learning a central question in algorithmic learning theory. Since then, there has been a large body of work[1] centered around providing theoretical guarantees for learning decision trees under various generalizations and restrictions of the original PAC model introduced

*sagnikc@iiitd.ac.in

[1]See Ehrenfeucht and Haussler [1989], Kushilevitz and Mansour [1991], Linial et al. [1993], Mehta and Raghavan [2002], Gopalan et al. [2008], Kalai and Kanade [2009], Feldman [2009], Blanc et al. [2020] for details.

by Valiant [1984]. There are two types of decision tree learning algorithms: proper learning algorithms, which output decision trees, and improper learning algorithms, where the output hypothesis is not required to be a decision tree. Strong computational hardness results exist for proper learning of decision trees [Koch et al., 2023, Bshouty, 2023], while there are known efficient improper learning algorithms for decision trees [Kalai and Kanade, 2009, Feldman, 2009]. A closely related problem to learning decision trees is *testing* decision trees. Instead of explicitly constructing decision trees, the task in testing is simply to decide if a decision tree representation of a given size exists.

The issue with the above classical polytime agnostic learning and testing algorithms is that they make *membership queries* in the agnostic PAC setting [1993, 1998, 2008, 2009, 2009]. A membership query (**MQ**) oracle allows a learning algorithm to learn the label of any desired instance in the domain, even among the ones absent in the training set. It is known that the PAC+MQ model is strictly stronger than the PAC with random examples [1984, 1988, 1993, 2006]. Despite these interesting theoretical results, MQ oracles are extremely difficult to implement in practice [Awasthi et al., 2013]. Baum and Lang [1992] observed that in experiments on handwritten characters and digits, the learning algorithms generated query points that often had no structure to a human observer. Using such oracles immediately detracts from our goal of producing human-interpretable models as well.

In practice, machine learning algorithms use data in the training set to learn a hypothesis. This setup can be modeled as having query access to a random example oracle where we sample training points according to a uniform distribution. Having access to a uniform superposition over the training set is an extremely natural notion in quantum computing. Many algorithms [1998, 2007, 2020, 2020] have been designed for the realizable Quantum PAC model where we have access to uniform superposition over the training sets. My previous work [Chatterjee et al., 2022a] gives a quantum improper learning algorithm for decision stumps using random examples. In the agnostic setting, however, nothing is known about learning (classical or quantum), with just access to random examples. My previous work [Chatterjee et al., 2022b] makes partial progress in this direction by giving a polytime quantum improper learning algorithm for polynomial-sized decision trees using quantum example (**QEX**) queries[2] which are *weaker* compared to membership queries [1998], but stronger than random example queries.

Table 1: Comparing different algorithms for learning size-t decision trees for n-bit functions in the agnostic setting. Here membership queries are denoted by MQ. Super-polynomial means that the running time is not polynomial in n or t.

Running Time	**Proper Learning**	**Improper Learning (with MQ)**	**Improper Learning (with random examples)**
Lower Bounds	Super-polynomial [2023]	Polynomial [2008]	Open (**This proposal**)
Upper Bounds	Super-polynomial [1989]	Polynomial [2008]	Open (**This proposal**) Polynomial with QEX queries (**Our result**) [2022b]

[2]These produce a superposition over all examples in the domain instead of simply the training set as in random examples.

3 Plan

I plan to design improper quantum learning algorithms for decision trees that obtain optimal tight bounds without the use of membership queries in the realizable and agnostic setting. My proposal builds on my previous works [Chatterjee, Bhatia, Chani, and Bera, 2022a, Chatterjee, SAPV, and Bera, 2022b], where I show how to carefully design and obtain upper bounds for quantum improper learning algorithms in the realizable and agnostic PAC settings which have no dependence on membership queries and still retain polynomial running time. Our algorithms also obtain a quadratic speedup with respect to the VC dimension over their classical counterparts. To address the lower bounds, we consider the following theorems:

Theorem 1 ([Goldreich et al., 1998, Blanc et al., 2020]). *Proper learning of decision trees is equivalent to testing.*

Theorem 2 ([Kothari and Livni, 2018]). *Refutation is a generalization of Testing.*

Theorem 3 ([Vadhan, 2017, Kothari and Livni, 2018]). *Improper Learning is equivalent to Refutation.*

The relationships captured in theorems 1 to 3 are outlined below.

$$\text{ProperLearning} \equiv \text{Testing} \supseteq \text{Refutation} \equiv \text{ImproperLearning}$$

We state our main conjecture and plan of attack below:

Conjecture 1. *For some constants $c, k > 0$, there exists quantum improper algorithms which learn size-t decision trees on n-bit functions using random examples in the agnostic setting in time $\Theta\left(n^c \cdot t^k\right)$.*

- **Lower Bounds:** Kearns and Ron [1998] gave a framework for obtaining lower bounds for decision tree classical testing algorithms using both random examples membership queries. I will generalize this framework to obtain lower bounds for quantum testing algorithms for decision trees without membership queries, using techniques demonstrated in our work [Chatterjee et al., 2022b]. I will then use the generalization of testing to refutation [Vadhan, 2017, Kothari and Livni, 2018] to obtain refutation oracles for decision trees using only access to random examples. This will require modifying the existing proof frameworks to obtain lower bounds for quantum refutation algorithms in both realizable and agnostic settings.
- **Upper Bounds:** I plan to obtain quantum weak learners for decision trees from random examples by using refutation algorithms in the corresponding setting (realizable or agnostic), and use our quantum boosting algorithms [Chatterjee et al., 2022a,b] to obtain quantum improper learning algorithms from random examples with tight bounds.

4 Conclusion

As machine learning and artificial intelligence become more pervasive in modern society, we must resort to interpretable learning models with concrete theoretical guarantees. The ultimate goal is to create models with theoretical bounds which can be understood and trusted by the end user, particularly in areas with significant real-world consequences. My work will contribute towards resolving some long-standing open problems in computational learning theory related to decision trees and simultaneously construct actual algorithms which abide by these guarantees.

References

Dana Angluin. Queries and concept learning. *Machine learning*, 2(4):319–342, 1988. 2

Srinivasan Arunachalam and Reevu Maity. Quantum boosting. In *37th Int. Conf. Mach. Learn. ICML 2020*, volume PartF16814, pages 354–364. PMLR, nov 2020. ISBN 9781713821120. 2

Alp Atıcı and Rocco A Servedio. Quantum algorithms for learning and testing juntas. *Quantum Information Processing*, 6(5):323–348, 2007. 2

Pranjal Awasthi, Vitaly Feldman, and Varun Kanade. Learning using local membership queries. In *Conference on Learning Theory*, pages 398–431. PMLR, 2013. 2

Eric B Baum and Kenneth Lang. Query learning can work poorly when a human oracle is used. In *International joint conference on neural networks*, volume 8, page 8. Beijing China, 1992. 2

Guy Blanc, Jane Lange, and Li-Yang Tan. Provable guarantees for decision tree induction: the agnostic setting. In *International Conference on Machine Learning*, pages 941–949. PMLR, 2020. 1, 1

Nader H. Bshouty. Superpolynomial lower bounds for learning monotone classes, 2023. 1, 2

Nader H Bshouty and Jeffrey C Jackson. Learning dnf over the uniform distribution using a quantum example oracle. *SIAM Journal on Computing*, 28(3):1136–1153, 1998. 2

N.H. Bshouty. Exact learning via the monotone theory. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 302–311, 1993. doi: 10.1109/SFCS.1993.366857. 2

Sagnik Chatterjee, Rohan Bhatia, Parmeet Singh Chani, and Debajyoti Bera. Quantum Boosting using Domain-Partitioning Hypotheses, August 2022a. URL http://arxiv.org/abs/2110.12793. arXiv:2110.12793 [quant-ph]. 2, 3, 3

Sagnik Chatterjee, Tharrmashastha SAPV, and Debajyoti Bera. Efficient quantum agnostic improper learning of decision trees, 2022b. 2, 1, 3, 3

Andrzej Ehrenfeucht and David Haussler. Learning decision trees from random examples. *Information and Computation*, 82(3):231–246, 1989. 1, 1

Vitaly Feldman. Optimal hardness results for maximizing agreements with monomials. In *21st Annual IEEE Conference on Computational Complexity (CCC'06)*, pages 9–pp. IEEE, 2006. 2

Vitaly Feldman. Distribution-Specific Agnostic Boosting, September 2009. 1, 2

Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, jul 1998. ISSN 0004-5411. doi: 10.1145/285055.285060. URL https://doi.org/10.1145/285055.285060. 1

Parikshit Gopalan, Adam Tauman Kalai, and Adam R Klivans. Agnostically learning decision trees. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 527–536, 2008. 1, 2, 1

Adam Izdebski and Ronald de Wolf. Improved Quantum Boosting, sep 2020. 2

Adam Tauman Kalai and Varun Kanade. Potential-Based Agnostic Boosting. *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, pages 880–888, December 2009. 1, 2

Michael Kearns and Dana Ron. Testing problems with sub-learning sample complexity. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, COLT' 98, page 268–279, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 1581130570. doi: 10.1145/279943.279996. URL https://doi.org/10.1145/279943.279996. 2, 3

Caleb Koch, Carmen Strassle, and Li-Yang Tan. Superpolynomial lower bounds for decision tree learning and testing. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1962–1994. SIAM, 2023. 1, 2, 1

Pravesh Kothari and Roi Livni. Improper learning by refuting. In *Information Technology Convergence and Services*, 2018. 2, 3, 3

Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the fourier spectrum. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pages 455–464, 1991. 1

Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, jul 1993. ISSN 0004-5411. doi: 10.1145/174130.174138. URL https://doi.org/10.1145/174130.174138. 1, 2

Dinesh Mehta and Vijay Raghavan. Decision tree approximations of boolean functions. *Theoretical Computer Science*, 270(1):609–623, 2002. ISSN 0304-3975. doi: https://doi.org/10.1016/S0304-3975(01)00011-1. URL https://www.sciencedirect.com/science/article/pii/S0304397501000111. 1

Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2018.07.007. URL https://www.sciencedirect.com/science/article/pii/S0004370218305988. 1

Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5): 206–215, 2019. 1

Cynthia Rudin, Chaofan Chen, Zhi Chen, Haiyang Huang, Lesia Semenova, and Chudi Zhong. Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistic Surveys*, 16:1–85, 2022. 1

Salil Vadhan. On learning vs. refutation. In Satyen Kale and Ohad Shamir, editors, *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 1835–1848. PMLR, 07–10 Jul 2017. URL https://proceedings.mlr.press/v65/vadhan17a.html. 3, 3

Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11): 1134–1142, 1984. 2