

# Improved Algorithms for Efficient Active Learning Halfspaces with Massart and Tsybakov Noise

**Chicheng Zhang**

*The University of Arizona*

CHICHENGZ@CS.ARIZONA.EDU

**Yinan Li**

*The University of Arizona*

YINANLI@EMAIL.ARIZONA.EDU

**Editors:** Mikhail Belkin and Samory Kpotufe

## Abstract

We give a computationally-efficient PAC active learning algorithm for  $d$ -dimensional homogeneous halfspaces that can tolerate Massart noise (Massart and Nédélec, 2006) and Tsybakov noise (Tsybakov, 2004). Specialized to the  $\eta$ -Massart noise setting, our algorithm achieves an information-theoretically near-optimal label complexity of  $\tilde{O}\left(\frac{d}{(1-2\eta)^2} \text{polylog}\left(\frac{1}{\epsilon}\right)\right)$  under a wide range of unlabeled data distributions (specifically, the family of “structured distributions” defined in Diakonikolas et al. (2020a)). Under the more challenging Tsybakov noise condition, we identify two subfamilies of noise conditions, under which our efficient algorithm provides label complexity guarantees strictly lower than passive learning algorithms.

**Keywords:** Active learning, halfspaces, noise-tolerant learning

## 1. Introduction

Halfspaces are arguably one of the most fundamental concept classes studied in machine learning and data analysis. Computational hardness results (Feldman et al., 2006; Guruswami and Raghavendra, 2009; Daniely, 2016; Klivans and Kothari, 2014; Diakonikolas et al., 2020b) motivate the study of learning halfspaces under more benign noise conditions, prominent examples include Massart noise (Massart and Nédélec, 2006) and Tsybakov noise (Tsybakov, 2004). In this work, we advance the state of the art on efficient PAC active halfspace learning (Balcan et al., 2009) under a set of structural assumptions on the unlabeled data distribution (Diakonikolas et al., 2020a), by providing an efficient algorithm such that, with appropriate setting of its parameters:

1. Under the  $\eta$ -Massart noise condition, it has an information-theoretically near-optimal label complexity of  $O\left(\frac{d}{(1-2\eta)^2} \text{polylog}\left(\frac{1}{\epsilon}\right)\right)$ . This substantially weakens existing distributional requirements to achieve such near-optimal label complexity results (Yan and Zhang, 2017; Awasthi et al., 2015).
2. Under the  $(A, \alpha)$ -Tsybakov noise condition with  $\alpha \in (\frac{1}{2}, 1]$ , it has a label complexity  $\tilde{O}\left(d\left(\frac{1}{\epsilon}\right)^{\frac{2-2\alpha}{2\alpha-1}}\right)$ . Furthermore, in the special case of  $(B, \alpha)$ -geometric Tsybakov noise condition (see full version), it has a lower label complexity of  $O\left(d\left(\frac{1}{\epsilon}\right)^{\frac{2}{\alpha}-2}\right)$  for  $\alpha \in (0, 1]$ ; for certain ranges of  $\alpha$ , our algorithm has a better label complexity than passive learning (see e.g., Hanneke, 2014, Section 3).

. Extended abstract. Full version appears as [arXiv:2102.05312, v2].

## References

- Pranjal Awasthi, Maria-Florina Balcan, Nika Haghtalab, and Ruth Uerner. Efficient learning of linear separators under bounded noise. In *Proceedings of the 28th Annual Conference on Learning Theory*, pages 167–190, 2015.
- Maria-Florina Balcan, Alina Beygelzimer, and John Langford. Agnostic active learning. *Journal of Computer and System Sciences*, 75(1):78–89, 2009.
- Amit Daniely. Complexity theoretic limitations on learning halfspaces. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 105–117, 2016.
- Ilias Diakonikolas, Daniel M Kane, Vasilis Kontonis, Christos Tzamos, and Nikos Zarifis. A polynomial time algorithm for learning halfspaces with tsybakov noise. *arXiv preprint arXiv:2010.01705*, 2020a.
- Ilias Diakonikolas, Daniel M Kane, and Nikos Zarifis. Near-optimal sq lower bounds for agnostically learning halfspaces and relus under gaussian marginals. *arXiv preprint arXiv:2006.16200*, 2020b.
- Vitaly Feldman, Parikshit Gopalan, Subhash Khot, and Ashok Kumar Ponnuswami. New results for learning noisy parities and halfspaces. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 563–574, 2006.
- Venkatesan Guruswami and Prasad Raghavendra. Hardness of learning halfspaces with noise. *SIAM Journal on Computing*, 39(2):742–765, 2009.
- Steve Hanneke. Theory of disagreement-based active learning. *Foundations and Trends® in Machine Learning*, 7(2-3):131–309, 2014.
- Adam Klivans and Pravesh Kothari. Embedding hard learning problems into gaussian space. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.
- Pascal Massart and Élodie Nédélec. Risk bounds for statistical learning. *The Annals of Statistics*, pages 2326–2366, 2006.
- Alexander B. Tsybakov. Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, 32(1):135–166, 2004.
- Songbai Yan and Chicheng Zhang. Revisiting perceptron: Efficient and label-optimal learning of halfspaces, 2017.