Proceedings of Machine Learning Research vol 291:1–5, 2025 38th Annual Conference on Learning Theory

A Distributional-Lifting Theorem for PAC Learning

Guy Blanc GBLANC@STANFORD.EDU
Stanford

Jane Lange JLANGE@MIT.EDU
MIT

Carmen Strassle STRASSLE@STANFORD.EDU
Stanford

Li-Yang Tan LYTAN@STANFORD.EDU
Stanford

Editors: Nika Haghtalab and Ankur Moitra

Abstract

The apparent difficulty of efficient distribution-free PAC learning has led to a large body of work on distribution-specific learning. Distributional assumptions facilitate the design of efficient algorithms but also limit their reach and relevance. Towards addressing this, we prove a *distributional-lifting theorem*: This upgrades a learner that succeeds with respect to a limited distribution family $\mathcal{D}$ to one that succeeds with respect to *any* distribution $D^\star$, with an efficiency overhead that scales with the complexity of expressing $D^\star$ as a mixture of distributions in $\mathcal{D}$.

Recent work of Blanc, Lange, Malik, and Tan considered the special case of lifting uniform-distribution learners and designed a lifter that uses a conditional sample oracle for $D^\star$, a strong form of access not afforded by the standard PAC model. Their approach, which draws on ideas from semi-supervised learning, first learns $D^\star$ and then uses this information to lift.

We show that their approach is information-theoretically intractable with access only to random examples, thereby giving formal justification for their use of the conditional sample oracle. We then take a different approach that sidesteps the need to learn $D^\star$, yielding a lifter that works in the standard PAC model and enjoys additional advantages: it works for all base distribution families, preserves the noise tolerance of learners, has better sample complexity, and is simpler. [1]

Keywords: PAC learning, distribution-specific learning, distributional decomposition

Acknowledgments

Guy, Carmen, and Li-Yang are supported by NSF awards 1942123, 2211237, 2224246, a Sloan Research Fellowship, and a Google Research Scholar Award. Guy is also supported by a Jane Street Graduate Research Fellowship; Carmen by an NSF GRFP; Jane by NSF Awards 2006664 and 310818 and an NSF GRFP.

1. Extended abstract. Full version can be found at [arXiv:2506.16651, v1].

References

Maryam Aliakbarpour, Eric Blais, and Ronitt Rubinfeld. Learning and testing junta distributions. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 19–46, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR. URL https://proceedings.mlr.press/v49/aliakbarpour16.html.

Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

Peter Auer, Stephen Kwek, Wolfgang Maass, and Manfred Warmuth. Learning of depth two neural networks with constant fan-in at the hidden nodes. In *Proceedings of the 9th Annual Conference on Computational Learning Theory (COLT)*, pages 333–343, 1996.

Shai Ben-David, Nader Bshouty, and Eyal Kushilevitz. A composition theorem for learning algorithms with applications to geometric concept classes. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)*, pages 324–333, 1997.

Gyora M. Benedek and Alon Itai. Learnability with respect to fixed distributions. *Theoretical Computer Science*, 86(2):377–389, 1991. ISSN 0304-3975.

Rishiraj Bhattacharyya and Sourav Chakraborty. Property testing of joint distributions using conditional samples. *ACM Transactions on Computation Theory (TOCT)*, 10(4):1–20, 2018.

Eric Blais, Ryan O'Donnell, and Karl Wimmer. Polynomial regression under arbitrary product distributions. *Machine learning*, 80(2):273–294, 2010.

Guy Blanc and Gregory Valiant. Adaptive and oblivious statistical adversaries are equivalent, 2024. URL https://arxiv.org/abs/2410.13548.

Guy Blanc, Jane Lange, Ali Malik, and Li-Yang Tan. Popular decision tree algorithms are provably noise tolerant. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, 2022.

Guy Blanc, Jane Lange, Ali Malik, and Li-Yang Tan. Lifting uniform learners via distributional decomposition. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1755–1767, 2023.

Avrim Blum and Ronald Rivest. Training a 3-node neural network is np-complete. *Advances in Neural Information Processing Systems*, 1, 1988.

Nader Bshouty and Hanna Mazzawi. Exact learning composed classes with a small number of mistakes. In *Proceedings of the 19th Annual Conference on Learning Theory (COLT)*, pages 199–213, 2006.

Nader H Bshouty. A new composition theorem for learning algorithms. In *Proceedings of the 30th annual ACM Symposium on Theory of Computing (STOC)*, pages 583–589, 1998.

Nader H Bshouty, Nadav Eiron, and Eyal Kushilevitz. PAC learning with nasty noise. *Theoretical Computer Science*, 288(2):255–275, 2002.

Clément L Canonne, Dana Ron, and Rocco A Servedio. Testing probability distributions using conditional samples. *SIAM Journal on Computing*, 44(3):540–616, 2015.

Clément L Canonne, Xi Chen, Gautam Kamath, Amit Levi, and Erik Waingarten. Random restrictions of high dimensional distributions and uniformity testing with subcube conditioning. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 321–336. SIAM, 2021.

Sitan Chen and Ankur Moitra. Beyond the low-degree algorithm: mixtures of subcubes and their applications. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing (STOC)*, pages 869–880, 2019.

Xi Chen, Rajesh Jayaram, Amit Levi, and Erik Waingarten. Learning and testing junta distributions with sub cube conditioning. In *Conference on Learning Theory*, pages 1060–1113. PMLR, 2021.

Mary Cryan. *Learning and approximation Algorithms for Problems motivated by evolutionary trees*. PhD thesis, Department of Computer Science, 1999.

Mary Cryan, Leslie Ann Goldberg, and Paul W Goldberg. Evolutionary trees can be learned in polynomial time in the two-state general markov model. *SIAM Journal on Computing*, 31(2): 375–397, 2001.

Ilias Diakonikolas and Daniel M Kane. *Algorithmic high-dimensional robust statistics*. Cambridge university press, 2023.

Jon Feldman, Ryan O'Donnell, and Rocco A Servedio. Learning mixtures of product distributions over discrete domains. *SIAM Journal on Computing*, 37(5):1536–1564, 2008.

Yoav Freund and Yishay Mansour. Estimating a mixture of two product distributions. In *Proceedings of the twelfth annual conference on Computational learning theory*, pages 53–62, 1999.

Parikshit Gopalan, Adam Kalai, and Adam Klivans. Agnostically learning decision trees. In *Proceedings of the 40th ACM Symposium on Theory of Computing (STOC)*, pages 527–536, 2008.

Christina Göpfert, Shai Ben-David, Olivier Bousquet, Sylvain Gelly, Ilya Tolstikhin, and Ruth Urner. When can unlabeled data improve the learning rate? In *Proceedings of the 32nd Conference on Learning Theory (COLT)*, pages 1500–1518. PMLR, 2019.

Thomas Hancock. *On the difficulty of finding small consistent decision trees*. 1989. Harvard University.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. ISSN 01621459. URL http://www.jstor.org/stable/2282952.

Russell Impagliazzo, Rex Lei, Toniann Pitassi, and Jessica Sorrell. Reproducibility in learning. In *Proceedings of the 54th Annual ACM Symposium on Theory of Computing (STOC)*, pages 818–831, 2022.

Jeffrey C Jackson. An efficient membership-query algorithm for learning dnf with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55(3):414–440, 1997. ISSN 0022-0000. doi: https://doi.org/10.1006/jcss.1997.1533. URL https://www.sciencedirect.com/science/article/pii/S0022000097915336.

Adam Kalai, Adam Klivans, Yishay Mansour, and Rocco A. Servedio. Agnostically learning halfspaces. *SIAM Journal on Computing*, 37(6):1777–1805, 2008.

Alkis Kalavasis, Amin Karbasi, Grigoris Velegkas, and Felix Zhou. On the computational landscape of replicable learning. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

Daniel Kane. The average sensitivity of an intersection of halfspaces. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 437–440, 2014.

Michael Kearns and Leslie Valiant. Cryptographic limitations on learning boolean formulae and finite automata. In *Proceedings of the 21st ACM Symposium on Theory of Computing (STOC)*, volume 21, pages 433–444, 1989.

Michael Kearns, Ming Li, Leonard Pitt, and Leslie Valiant. On the learnability of boolean formulae. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC)*, pages 285–295, 1987.

Michael Kearns, Robert Schapire, and Linda Sellie. Toward efficient agnostic learning. *Machine Learning*, 17(2/3):115–141, 1994.

Adam R Klivans and Alexander A Sherstov. Unconditional lower bounds for learning intersections of halfspaces. *Machine Learning*, 69(2):97–114, 2007.

Adam R. Klivans and Alexander A. Sherstov. Cryptographic hardness for learning intersections of halfspaces. *Journal of Computer and System Sciences*, 75(1):2–12, 2009. ISSN 0022-0000. doi: https://doi.org/10.1016/j.jcss.2008.07.008. URL https://www.sciencedirect.com/science/article/pii/S0022000008000706. Learning Theory 2006.

Adam R Klivans, Ryan O'Donnell, and Rocco A Servedio. Learning intersections and thresholds of halfspaces. *Journal of Computer and System Sciences*, 68(4):808–840, 2004.

Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the fourier spectrum. *SIAM Journal on Computing*, 22(6):1331–1348, December 1993.

Ming Li and Umest Vazirani. On the learnability of finite automata. In *Proceedings of the 1st Annual Workshop on Computational Learning Theory*, pages 359–370, 1988.

Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

B. K. Natarajan. Probably approximate learning over classes of distributions. *SIAM Journal on Computing*, 21(3):438–449, 1992.

Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008.

Leonard Pitt and Leslie G Valiant. Computational limitations on learning from examples. *Journal of the ACM (JACM)*, 35(4):965–984, 1988.

Leonard Pitt and Manfred Warmuth. The minimum consistent dfa problem cannot be approximated within and polynomial. In *Proceedings of the 21st Annual Symposium on Theory of Computing (STOC)*, pages 421–432, 1989.

Leonard Pitt and Manfred K Warmuth. Reductions among prediction problems: on the difficulty of predicting automata. In *Proceedings of the 3rd Annual Conference on Structure in Complexity Theory*, pages 60–69, 1988.

Leonard Pitt and Manfred K Warmuth. Prediction-preserving reducibility. *Journal of Computer and System Sciences*, 41(3):430–467, 1990.

Alexander A Razborov and Alexander A Sherstov. The sign-rank of AC^0. *SIAM Journal on Computing*, 39(5):1833–1855, 2010.

Alexander A Sherstov. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 523–532, 2010.

Alexander A Sherstov. The intersection of two halfspaces has high threshold degree. *SIAM Journal on Computing*, 42(6):2329–2374, 2013.

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

Karsten Verbeurgt. Learning DNF under the uniform distribution in quasi-polynomial time. In *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, pages 314–326, 1990.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.