A Query Algorithm for Agnostically Learning DNF?

Parikshit Gopalan[*] and **Adam T. Kalai**[†] and **Adam R. Klivans**[‡]

Let $f : \{-1,1\}^n \to \{-1,1\}$ be an *arbitrary* Boolean function and let $\mathcal{C}$ be a concept class where each concept has size at most t. Define

$$\mathsf{opt} = \min_{c \in \mathcal{C}} \Pr_{x \in \{-1,1\}^n}[c(x) \neq f(x)]$$

where x is chosen uniformly at random from $\{-1,1\}^n$. We say that $\mathcal{C}$ is *agnostically learnable with queries with respect to the uniform distribution* if there exists an algorithm that– given black box access to any f– runs in time $\mathrm{poly}(n, t, \epsilon^{-1})$ and outputs a hypothesis h such that

$$\Pr_{x \in \{-1,1\}^n}[h(x) \neq f(x)] \leq \mathsf{opt} + \epsilon.$$

The algorithm may be randomized, in which case it must output such an h with high probability.

The main question is as follows: are polynomial-size DNF formulas agnostically learnable with queries with respect to the uniform distribution? A related question is, are halfspaces agnostically learnable with queries with respect to the uniform distribution?

Motivation: One of the most celebrated results in computational learning theory is Jackson's query algorithm for PAC learning DNF formulas with respect to the uniform distribution [3]. A natural question is whether DNF formulas can be learned (even with queries and with respect to the uniform distribution) in a highly noisy setting, i.e., the well-known agnostic framework of learning [5].

Additionally, it is straightforward to see that an agnostic learning algorithm for DNF formulas would give algorithms for weakly learning polynomial-size depth-3 circuits with respect to the uniform distribution in the standard PAC learning model.

Halfspaces are another simple and important concept class of functions still not known to be agnostically learnable with respect to the uniform distribution, even if the learner can make queries (although some relevant work exists for the uniform distribution that we mention below).

Current status: Very recently, Gopalan et al. [2] have shown that the weaker concept class of *decision trees* are agnostically learnable with queries with respect to the uniform distribution. Their algorithm implicitly solves a high-dimensional convex program using the well-known Kushilevitz/Mansour [6] algorithm for finding large Fourier coefficients as a subroutine.

Applying a result due to Mansour on the sparsity of DNF formulas [7], the Gopalan et al. query algorithm will agnostically learn DNF formulas with respect to the uniform distribution in time $n^{O(\log(1/\epsilon)\log\log n)}$. If the Friedgut-Kalai "Entropy/Influence" conjecture [1] is true (or better bounds are proved on the sparsity of DNF formulas) then the running time can be improved even further (see also Gil Kalai's post on Terry Tao's weblog [8]).

Gopalan et al. do show, however, that that their algorithm will *not* agnostically learn DNF formulas in polynomial time in all the relevant parameters.

Given Jackson's algorithm and Gopalan et al.'s recent work for agnostically learning decision trees, we feel that the case of DNF formulas is particularly compelling.

Regarding halfspaces, Kalai et al. [4] showed how to agnostically learn halfspaces with respect to the uniform distribution *without queries* in time $n^{O(1/\epsilon^4)}$. Further, they showed that any algorithm running in time $n^{O(1/\epsilon^{2-\gamma})}$ for any $\gamma > 0$ would give the fastest known algorithm for the notoriously difficult "learning parity with noise problem." As such, we do not think that much further progress will be made on this problem unless the learner is allowed to make queries.

[*]University of Washington; parik@cs.washington.edu
[†]Georgia Tech.; adamology@gmail.com
[‡]University of Texas at Austin; klivans@cs.utexas.edu

References

[1] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society*, 124:2993–3002, 1996.

[2] P. Gopalan, A. Kalai, and A. Klivans. Agnostically learning decision trees. In *Proceedings of the* 40^{th} *ACM Symp. on Theory of Computing*, 2008.

[3] J. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55:414–440, 1997.

[4] A. Kalai, A. Klivans, Y. Mansour, and R. Servedio. Agnostically learning halfspaces. In *Proceedings of the* 46^{th} *IEEE Symp. on Foundations of Computer Science*, 2005.

[5] M. Kearns, R. Schapire, and L. Sellie. Toward Efficient

Agnostic Learning. *Machine Learning*, 17(2/3):115–141, 1994.

[6] E. Kushilevitz and Y. Mansour. Learning decision trees using the Fourier spectrum. *SIAM J. on Computing*, 22(6):1331–1348, 1993.

[7] Y. Mansour. An $o(n^{\log \log n})$ learning algorithm for DNF under the uniform distribution. *Journal of Computer and System Sciences*, 50:543–550, 1995.

[8] T. Tao. The entropy/influence conjecture (blog post by gil kalai). Available at `http://terrytao.wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjecture/`.