CS5234 Mini Project: Testing Juntas

Yeo Qi Xun (A0228331M) and Pitchappan P RM (A0236575W)

October 23, 2021

1 Introduction

The curse of dimensionality has been a fundamental problem which bottlenecked the progress in computer science research. Although this problem has been alleviated to a certain extent by dimensionality reduction techniques such as Principal Component Analysis (PCA), it is still useful to be able to determine if a certain attribute is relevant to the output obtained from the algorithm. Herein lies the motivation for the testing of attributes to see if they are relevant to the output obtained. Before we try to run these algorithms to try to obtain the attributes that are important, it is often useful to test what is the number of attributes that are capable of influencing the output value. This gives rise to research in junta testing where a k-junta is defined as a function of with k relevant attributes though it may have n attributes ($n >> k$). While learning a k-junta requires at least ω (k log n) samples, k-juntas can be tested with query complexity independent of n.

The core paper around which this initial exploration of the Juntas problem has been centered around Eric Blais's "Testing Juntas Nearly Optimally", 2009 [Bla09].

2 Related Work

The Junta problem was theorized by Avrim Blum in 1994 with the main idea that for a data set and a goal, some samples may be more relevant than others, and some features may be more relevant than others [BL97]. There was a followup to the paper in 1996 by Avrim Blum and Pat Langley which surveyed the latest methods of that time that focused on selecting the most relevant features. It did mention that the obvious bound for finding k-juntas is $n^(k + O(1))$, which is brute forcing through the whole data set and a primary challenge would be to do better [BL97].

In 2003, Mossel, O'Donnell, and Servedio had a breakthrough in the Junta problem. This paper named the problem the 'Junta problem' (it was unnamed previously) and showed that the problem could be solved in $n^{(\alpha k+O(1))}$ where $\alpha = \omega/(\omega + 1)$ where ω is the Strassen constant. They did so by splitting the solution of juntas into 2 cases, one where degree of the Boolean function was small and another where the Boolean function has a small non-zero Fourier coefficient [MOS03].

While learning a k-junta requires at least Omega(k log n) samples, k-juntas can be tested with query complexity independent of n. Eric Blais bought down the complexity of testing if a function was k-junta or far from k-junta to complexity $O(k^{3/2})/epsilon$ in [Bla08] by using a non-adaptive, two-sided upper bound and then reduced it to $O(k/\varepsilon + klogk)$ in [Bla09] using a adaptive, one-sided upper bound.

3 A detailed summary of the research paper.

The main idea behind this paper is further improve on the author's previous paper regarding the upper bound of k-junta testing. This is achieved by replacing the Independence test with finding a relevant part of the partition via binary search instead of doing the binary search on the coordinates themselves.

Independence testing was already previously used in obtaining the $\tilde{O}(k^{3/2}/\epsilon)$ bound for the non-adaptive, two-sided k-junta testing. Nonetheless it is a key component of the algorithm as we are trying to find individually what are the coordinates relevant to the given function f. This concept is used to verify if a subset of coordinates do not influence f. Finding relevant part of the partition extends that idea reducing the partition space from n to poly(k/ϵ) then subsequently doing binary search on the parts in the new partition space.

3.1 Theorems

In this paper, there is one main lemma and one main theorem that takes from the lemma.

The main lemma states that if I is a random partition of n that contains s parts where each coordinate is assigned via an i.i.d to one of the s parts, with probability of $\frac{5}{6}$ that a function f that is ϵ-far from being a k-junta is also $\frac{\epsilon}{2}$-far from being a k-junta with respect to partition I.

The number of queries required to ϵ-test k-juntas is bounded above by $O(k/\epsilon + k\log k)$. This stems from the $O(k/\epsilon)$ queries from the execution of the independence testing. Subsequently, if f is deemed to be dependent or influenced by the coordinates being tested, there will be at most (k+1) queries of binary search to find the relevant part of the partition which takes O(log(k/ϵ)) queries leading to a total of O($(k+1)\log(k/\epsilon)$) = O($k\log k$) queries.

3.2 Proofs

For the positive case where f is a k-junta, the k-Junta Test always gives the correct answer. If there are only k attributes where influence > 0, then at most k parts in the random partition will have influence > 0.

For the negative case, the main idea is to show that expectation of influence is at most $\epsilon/2$ since the probability of every set J formed by taking the union of k parts in a random partition of the coordinates should be at most $1 - \epsilon$ .The author utilised the Efron-Stein decomposition to redefine the influence of sets S for 3 different settings of S.

Firstly, the author explores the setting where $S > 2k$ or high-dimension where if J is unlikely to contain all the elements in S. The sum of every set J formed by taking the union of at most k parts in a random partition of the coordinates should be at most $\epsilon/4$ with probability $\frac{17}{18}$.

Secondly, for the setting where $S \leq 2k$ and influence is larger than θ (large low-order influence), H_f is defined as the coordinates with large influence on f. It is likely that there are few coordinates with large influence and they are likely to all get split up in the random partition. The sum of every set J formed by taking the union of at most k parts in a random partition of the coordinates should be at most $1 - 2\epsilon$ with probability $\frac{17}{18}$.

Lastly, for the setting where $S \leq 2k$ and influence is smaller than θ (small low-order influence), this setting will contribute little to the expectation of influence. Consequently, the expectation of a coordinate j lands in a random partition I_i of the coordinates should be at most $\epsilon/4k$. We can find that it holds with probability $\frac{1}{18s}$ using Hoeffding's bound. Using the union bound, we are able to get the sum of every set J formed by taking the union of at most k parts in a random partition of the coordinates should be at most $\epsilon/4$ with probability $\frac{17}{18}$.

Thus, by summing all the probabilities, we can prove the main lemma.

The main theorem naturally follows since we define s such that probability that the algorithm does not identify (k + 1) relevant parts in $12(k+1)/\epsilon$ rounds is at most 1/6 using Markov's inequality (expectation of the number of rounds is $2(k+1)/\epsilon$ from (k+1) parts divided by expectation of influence)). The probability that a wrong answer is returned is 1/6 for not identifying the relevant parts and 1/6 for function f that is ϵ-far from being a k-junta not being $\frac{\epsilon}{2}$-far from being a k-junta with respect to partition I. Thus, overall probability that the algorithm fails to reject f is at most 1/3.

4 Discussion

While the testing k-juntas problem has been around in the computational learning theory community for over 20 years now, many questions still remain unanswered. Tolerant testing for k-juntas is still a predominant problem, where the tester is trying to distinguish between a slight noisy version of a K-Junta and another very noisy k-junta. Examining k-juntas testing in an environment of bounded adaptivity as introduced in [CG17] is another path of exploration. Work by Levi and Waingarten [LW18] establishes a polynomial separation between standard and tolerant testing. Checking if this separation extends to adaptive algorithms is another open question.

References

[BL97] Avrim L. Blum and Pat Langley. "Selection of relevant features and examples in machine learning". In: *Artificial Intelligence* 97.1 (1997). Relevance, pp. 245–271. ISSN: 0004-3702. DOI: https://doi.org/10.1016/S0004-3702(97)00063-5. URL: https://www.sciencedirect.com/science/article/pii/S0004370297000635.

[MOS03] Elchanan Mossel, Ryan O'Donnell, and Rocco P. Servedio. "Learning Juntas". In: *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*. STOC '03. San Diego, CA, USA: Association for Computing Machinery, 2003, pp. 206–212. ISBN: 1581136749. DOI: 10.1145/780542.780574. URL: https://doi.org/10.1145/780542.780574.

[Bla08] Eric Blais. "Improved Bounds for Testing Juntas". In: *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*. Ed. by Ashish Goel et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 317–330. ISBN: 978-3-540-85363-3.

[Bla09] Eric Blais. "Testing Juntas Nearly Optimally". In: *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*. STOC '09. Bethesda, MD, USA: Association for Computing Machinery, 2009, pp. 151–158. ISBN: 9781605585062. DOI: 10.1145/1536414.1536437. URL: https://doi.org/10.1145/1536414.1536437.

[CG17] Clément L. Canonne and Tom Gur. "An Adaptivity Hierarchy Theorem for Property Testing". In: *CoRR* abs/1702.05678 (2017). arXiv: 1702.05678. URL: http://arxiv.org/abs/1702.05678.

[LW18] Amit Levi and Erik Waingarten. "Lower Bounds for Tolerant Junta and Unateness Testing via Rejection Sampling of Graphs". In: *CoRR* abs/1805.01074 (2018). arXiv: 1805.01074. URL: http://arxiv.org/abs/1805.01074.