We thank all four reviewers for their encouraging remarks and thorough treatment of our work. We respond individually to Reviewers 1 and 2, and jointly to Reviewers 3 and 4.

Reviewer 1 Response: We thank Reviewer 1 for their in-depth comments and questions on the supplementary materials. We address questions below. Smaller comment/questions not addressed will still be updated in the final version.

"Is there really a need to present Alg 1 and Thm 3.5?:" Alg 1, Thm 3.5, and Thm 3.6 are present in order to show a tight characterization of the query complexity of Comparison-Pool-PAC learning. While it is true that Theorem 4.11 also implies an upper bound on the weaker Comparison-Pool-PAC model, it leaves an unnecessary $\log(1/\varepsilon)$ gap between upper and lower bounds which Theorem 3.6 closes.

"In Alg 1, Threshold(S) is only informally defined, and an elaboration is needed. I think basically, the algorithm can successfully approximately recover b if it can find two neighboring + and - examples?" The procedure Threshold(S) produces some threshold consistent with labeling S by binary search (note that this procedure may mislabel small portions of S near the true threshold). We will clarify this in the text. In Theorem 3.6, we show this procedure produces a good threshold with probability at least $2/3$ based on neighboring $+/-$ examples. This is later amplified by Chernoff.

"In the second half of Corollary 4.7, is the goal only to identify the labels of all examples S drawn? Also, what is the active learning algorithm used here?" Yes, the second half Corollary 4.7 only finds the labels of S. The algorithm is [KLMZ17](Theorem 3.2) (Alg box pg. 16). In the text, we restate this theorem as Theorem 2.1, and reference it in the proof of Corollary 4.7. We will add further description of [KLMZ17]'s algorithm and why it makes no errors when assuming the wrong inference dimension (this is because inference dimension controls only the algorithms coverage).

"In Theorem 4.9, is H here $H_{d,\eta}$? ...What is $h \times [-\gamma, \gamma]$?" Yes, in Theorem 4.9 H should be $H_{d,\eta}$, and $h \times [-\gamma, \gamma]$ is $\{x : h(x) \in [-\gamma, \gamma]\}$. We will update these.

"[is it] possible that there exist two hypotheses h_1, h_2 that have _very small minimal-ratio_, and they agree with the queries in $Q(S'_h - \{x\})$, but disagree on x?" Yes, this is possible for arbitrary $h_1, h_2 \in H_d$. However, this does not affect our argument, since we have reduced (with very high probability) to the case that h has large minimal ratio with respect to S'_h. Note that this does not cause any errors since it is not an assumption, but rather is based off of a verifiable structural property of S (no large subset is too close or too far from any $h \in H_d$).

Reviewer 2 Response: We thank Reviewer 2 for their comments. While we agree with much of their assessment, we would like to address two aspects of the review with which we disagree.

Computational Efficiency: The Reviewer's only stated weakness is that *"The paper does not consider computational efficiency and noise tolerance."* While it is true that our work focuses mainly on characterizing the query complexity of realizable-case learning, the former part of this statement is false: we *do provide computationally efficient algorithms* for both the RPU and PAC models, and *explicitly state this* in the paper (lines 214, 295). In more detail, our main contribution, the $\tilde{O}(d\log(1/\varepsilon)^2)$ Comparison-Pool-RPU learning upper bound, is computationally efficient, which also implies a computationally efficient algorithm for the strictly weaker PAC-model. That said, we realize that this fact is somewhat hidden in the paper, and thank the reviewer for bringing it to our attention. We will add a discussion of computational efficiency to the introduction to fix this.

Novelty and Focus: We would also like to clarify what we view as a misunderstanding of the focus and novelty of our work. Reviewer 2 comments mostly on our Comparison-Pool-Pac upper bound (Theorem 3.3) based upon [BL13], and at one point states *"the paper is built upon [BL13]."* In fact, [BL13] is used only once in the entire paper in order to help characterize the query complexity of Comparison-Pool-PAC learning. *We would like to highlight that we do not view this as the works' main contribution or novelty.* Rather, as we state in the paper on lines 217, 251, and 277, the main novelty and focus of our work lies in the analysis of query and computationally efficient RPU-learning, and especially in the introduction and development of average inference dimension–a novel tool for analyzing distribution dependent RPU-learning crucial to these results. It is worth noting that Reviewer 1 asks why we even include the [BL13] based result given our stronger results on RPU-learning (the only reason is a $\log(1/\varepsilon)$ gap in query complexity between the two algorithms).

Reviewers 3 and 4 Response: We thank Reviewers 3 and 4 for their encouraging comments, and believe both reviews appropriately frame and summarize our work. In response to Reviewer 4's suggestion, should the paper be accepted we would be happy to use part of the additional camera-ready page for covering further research perspectives.

References

[BL13] Long P. Balcan, M. Active and passive learning of linear separators under log-concave distributions. In *Proceedings of the 26th Conference on Learning Theory*, 2013.

[KLMZ17] Lovett S. Moran S. Zhang J. Kane, D. Active classification with comparison queries. In *IEEE 58th Annual Symposium on Foundations of Computer Science*, 2017.