Cover Letter for Paper 5810: 'When Expressivity Meets Trainability: Fewer than n Neurons Can Work'

Anonymous Author(s)
Affiliation
Address
email

Since our paper is part of the NeurIPS 2021 consistency experiment, we received two sets of reviews with Paper number 5810 and 8953, respectively. One committee (for 8953) recommended accept, and another committee (for 5810) recommended reject (mainly due to missing discussion of a relevant work). According to the message from Program Chairs in OpenReview (under the meta-review of paper 5810), the current decision is "conditional accept", and we need to address the concerns in the meta-review of paper 5810.

By now, we have revised the paper according to the meta-review, and we have submitted an anonymous version of our revised paper through the button of 'camera-ready-revision'. In this cover letter, we will explain how our modifications address the concern expressed in the meta-review.

1 Meta-Review 5810 and PC's Comments

Decision by Program Chairs: Conditional accept.

Meta-Review (for 5810). This paper makes a significant contribution and the results were found interesting and technically innovative.

However, the authors failed to discuss relevant literature in their original paper, in particular Daniely [2] and Bubeck et al. [1]. Regarding Bubeck et al [1], this is an easily fixable mistake, and there seem to be no real issue here that cannot be resolved.

However, as far as Daniely's work, the relevance is too large to be brushed off. Now, the authors did make convincing arguments in their rebuttal. But, bottom-line, they make arguments about the validity of the result in Daniely. Now, of course this is completely acceptable, and from my brief view of their comments I am even inclined to believe that they have a point. But such claims and discussions should have been part of the original paper and must go through the scrutiny of peer-review, they shouldn't be assessed in the limited form of the discussion period.

Given that Neurips does not accept revised versions, it is not possible to assess how the paper might look like once these issues will be resolved by the authors, and therefore I cannot recommend acceptance.

This copy's committee reached the following decision: Reject.

Comments by Program Chairs. The paper will be accepted conditioned on the revision appropriately addressing the concern described in the meta-review. The paper must pass a re-examination to ensure that the claims made with respect to the work of Daniely are correct.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

2 Response to Meta-Review 5810

Our response. We thank the AC for the valuable feedback, and we are grateful that the AC and reviewers appreciate our contribution.

In the revised version, we make the following two major changes (highlighted in red in the paper):

- Part I: We add discussions on the two related works Bubeck et al. [1] and Daniely [2] to Section 2 "Related Works".
- Part II: We add a more detailed discussion of Daniely [2] supporting Sec. 2 in Appendix A.

We briefly explain the contents of the two parts.

In Part I, we explain the two major differences with those two works (similar to what we wrote in the rebuttal):

- The width requirements of those works are not just $O(n/d)$, but have extra factors such as $1/\epsilon$ or a super-polynomial factor on n.
- We analyze a different optimization formulation from those two works; this is motivated by the bad performance of the standard SGD when training narrow nets (which those papers did not point out).

In Part II, we elaborate the calculation of the width requirement of Daniely [2]. We add this part because Daniely [2] did not explicitly write out the precise bound, but use $\tilde{O}(\cdot)$ to present a rough bound. In this part, we carefully explain how we derive a more precise bound by tracking the proof in Daniely [2].

References

[1] Sébastien Bubeck, Ronen Eldan, Yin Tat Lee, and Dan Mikulincer. Network size and weights size for memorization with two-layers neural networks. *arXiv preprint arXiv:2006.02855*, 2020.

[2] Amit Daniely. Neural networks learning and memorization with (almost) no over-parameterization. *arXiv preprint arXiv:1911.09873*, 2019.