1   We thank the reviewers for their comments and suggestions. We address the specific the specific concerns below.

2   **Reviewer 1:**   - ll. 38-41: We agree that Corollary 1 is not a direct corollary. It does, however, follow from the
3   techniques used to prove Theorem 1, and the details of this are described in Section 3.1. We will state and prove this
4   more formally in the final version.

5   - ll. 266-267: We were trying to point out that for example for random Boolean functions, the constant hypothesis 1/2
6   achieves smaller square loss than any homogeneous ReLU.

7   **Reviewer 2:**   Thank you for pointing out the typo. We will fix it.

8   **Reviewer 3:**   The details of the SQ lower bound are mentioned in section 3.1. We will be more precise in our theorem
9   statements in the final version.

10  Regarding the $d^{\Omega(\log(1/\epsilon))}$ lower bound, note that this rules out any algorithm that has a running time that is a fixed
11  polynomial in the dimension. This is stronger than, say a $(1/\epsilon)^{\log(1/\epsilon)}$ lower bound. We are not aware of a $d^{O(1/\epsilon)}$ time
12  algorithm for this problem.