We thank the referees for a thorough reading of the manuscript and helpful comments.

Reviewer 1(Weaknesses): Our main goal is to develop a lower-bound for the target generalization error achievable by any algorithm. We agree that our model/distance may not capture all practical scenarios but as our simulations demonstrate it does seem to correlate well with algorithmic performance. **(1) Re overparam.n:** Thanks for the very interesting question. Overparameterization does not pose a fundamental problem when using our notion of distance in practice. There can indeed be many different W that generate the same output on the training data due to overparam. However, in practice the W found by GD is one that generalizes well. And while all W with the same training output are not close, all W that generalize well must be close. This is because for such a W we must have $c \cdot \sigma_{\min}^2(V)\|\Sigma_S(W - W_S)\|_F^2 \leq \mathbb{E}\big[\|V\phi(Wx) - V\phi(W_S x)\|^2\big] < \delta$. So even though GD may not find W_S exactly due to overparameterization (since it typically find a generalizable model) it must be close to W_S. In fact one can make this rigorous using recent generalization theory (e.g. arxiv 1901.08584 and 1906.05392). Will further elaborate. **(2) Re not continuous:** Our more elaborate bound in the proof of the main theorem (Sec. 6.3 457-459) are indeed continuous at these transition points. To make the result more interpretable we simplified the expressions/loosened the bounds which is the source of discontinuity. We will further clarify. **(3) Re empirical results:** Thanks for the suggestion. We will move experiments to supp. and instead add more proof insights/mention some simple scenarios where our bounds are tight.

Reviewer 2: Re Weaknesses: Our goal is to understand the fundamental limits of what is possible, hence the focus on a lower bound. A lower bound is of significant practical interest in a variety of applications (see DARPA LwLL program TA2) as it can help predict how much transferability from source to target is possible prior to committing extensive resources to train complex transfer learning algorithms. Unfortunately, there is no good way to test the sharpness of lower bounds numerically. We do expect our bounds to be tight up to numerical constants as they resemble non-transfer learning bounds that are known to be sharp. In fact, very recently colleagues have informed us that they have developed algorithms that achieve our lower bound up to a fixed constant (under a more restrictive covariate shift assumption). To alleviate the reviewer's concern, in addition to mentioning this result (not yet publicly available) we will also provide some simple instances/scenarios that demonstrate the sharpness of our bounds. We also hope to develop matching algorithms in the general case in our future work. **Re question in correctness:** This is based on slightly modifying corollary 4.2.13 in "High dimensional probability" by R. Vershynin. We will clarify. **Re refs in Clarity:** Thanks, we will add more citations in the introduction section of the paper as well as discuss their pros and cons w.r.t. Our paper. **Re Relation to prior work:** Thanks for suggesting this paper we will cite/add a discussion about it.**Re Additional feedback): (1) re tightness:** Thanks, will add a discussion regarding the upper bound for the risk as well as the tightness of our result. **(2) re more complex models:** We view the models discussed in the paper as a first step towards studying more complicated neural network models. We do think it already captures some realistic phenomena as demonstrated in our numerical experiments. That said, we are working on generalizing our result to the case that both the hidden layer and output layers can both vary. **(3) Re are corollaries:** In Thm. 1, the result for the third model (input-to-hidden fixed), is a corollary of the linear case. However, this is not the case re second model (hidden-to-output fixed). In order to derive a lower bound in this case we need to find a metric for the risk and simplify the generalization error and a major part of the proof is devoted to this purpose. Will further clarify.

Reviewer 3: Thanks a lot for the positive feedback/assessment.

Reviewer 4: Re weaknesses: (1) re definitions: Lower bounds (unlike upper bounds for algorithms) do not always have easily interpretable quantities. That said, we have made an attempt to break our lower bound down into interpretable and intuitive terms. All the terms are well defined by precise mathematical expressions and we have named them accordingly to give some intuition what they would capture in the lower bound. We are happy to add more explanations for these terms in the final version for further clarification. As for the parameter A it should be replaced by V. Sorry for the typo. We caught this typo right after the submission and in fact have highlighted the correct version in the first paragraph of the supplementary. **(2) re strict constraints:** Indeed, our goal is to develop a lower bound that can be applied to more complex models (which is a very challenging problem) and our goal is to provide an important first step in this paper towards this goal. We note that even in the non-transfer learning scenario the theoretical study of more complex models remains elusive. We would also like to note that the lower bound is in fact not hard to apply to real datasets. All the parameters of the lower bound can be experimentally calculated/estimated from real data and one can apply the lower bound in the practical setting by having access to a large enough number of samples as done in our numerical experiments. In fact, we plan to participate in a challenge for such lower bounds (DARPA LwLL). **(Correctness):** We did not understood the reviewer's concern. Note that in three of the experiments the noise level is around the same and therefore the difference can only be attributed to the transfer distance. In a real experiment it is not possible to keep the noise level exactly the same. We will however add some synthetic experiments to demonstrate this further and to alleviate the reviewer's concern. We will also plot the final lower-bounds of target generalization error in our experiment and compare it to the final target generalization error. **(Clarity):** Thanks for finding the typo. We will fix it. **(Relation to prior work):** We will add further discussions on other related papers and their pros and cons.