1 We thank the reviewers for their insightful comments which we will integrate in the final version of the paper.

2 We appreciate reviewers' consensus that this work provides a novel theoretical study. Given this novelty, we are also
3 well aware that this work opens many questions and answering all of them requires future work. Being the first work
4 that aims at understanding the influence of transformations on kNN, we see this work as a starting point of a very
5 interesting research area.

6 **Common Concern (Reviewer 1, Reviewer 4): Significance of Empirical Result.** One common, major concern from
7 R1 and R4 is the significance of the empirical result. We agree that the current way the empirical results are presented
8 is confusing, and hope to clarify and address it here.

9 The contribution of this work is to establish the following result: Both the MSE error *and* the smoothness are important
10 factors governing the convergence of KNN over transformations; to the surprise of a common belief that the dimension
11 is important. As a result, to verify the significance of our contribution, we need to establish that:

12     1. Using MSE leads to significantly better correlation compared with using the dimension;
13     2. Examining the smoothness (represented by the norm), in addition to MSE, can lead to further improvement.

14 In terms of the comparison with LR Error and the fact that MSE seems to give only slightly better result than LR
15 Error – we agree that this is confusing and we apologize for it. LR Error is just a practical metric that is similar to
16 the MSE, and is included since LR Error is popular alternative of MSE, often used in practice (therefore, the similar
17 performance between the MSE error and the LR error is expected). However, **there is no theoretical understanding**
18 **linking LR Error to KNN convergence over feature transformations**. As a result, LR Error is not the baseline that
19 we are comparing with. The significance of our contribution is between (1) MSE + Norm, (2) MSE, and (3) Dimension.
20 We want to express again our appreciation to the reviewers for this very constructive feedback and we will revise our
21 draft accordingly to reflect this.

22 **Common Concern (Reviewer 2, Reviewer 4): Empirical Results for $k > 1$.** We agree that understanding the behav-
23 ior of KNN for different $k$ is important. We performed experiments for $k > 1$, over all datasets and all embeddings,
24 and the result were uploaded with the initial submission as supplementary material[1]. **We want to emphasize that all**
25 **the results for $k > 1$ confirm the theoretical findings, whilst for some $k$ they offer an improvement in the linear**
26 **correlation** (for example, SST2 one can get CCA score up to 0.97 when $k = 8$, whereas in the main body we report
27 0.917, for $k = 1$). When writing the submitted version of the paper we opted for $k = 1$ since it already serves the
28 purpose of showing that MSE is important on its own, whilst one could see that including the smoothness further
29 improves the correlation. However, we are now aware that we should summarize the findings for $k > 1$ in the main
30 body of the paper. We will do for the final version of the paper and we thank the reviewers for pointing this out.

31 We further address each reviewer individually.

32 **Reviewer 1.** Intuitively, *safety* explains how much of information is preserved after applying a transformation, with
33 Theorem 4.6 showing that it can be controlled by the $L^2$ loss. *Smoothness* is a common assumption in the work on
34 convergence rates of nearest neighbor estimators which allows the new point to learn from its neighbors. It is usually
35 given through the Lipschitz constant and we explore this in Definition 4.7 and Theorem 4.8. We paid particular attention
36 to defining all the necessary notions, in particular to novel definitions, and we will do a final check for the final version.

37 **Reviewer 2.** We agree with the reviewer that one could indeed examine the tightness of our theoretical bounds by
38 constructing a toy dataset with a known true posterior probability and a set of transformations. However, such a
39 transformation should either be an identity (if the toy dataset already has the desired property), or carefully constructed
40 for this purpose only. In this paper we opted for not constructing a single transformation ourselves, as our main goal is
41 to bridge the gap between the real-world applications (e.g. public pre-trained embeddings) with the theory of nearest
42 neighbors. For example, even Lemma 4.4, which establishes a sharp bound for the safety, works for any transformation.
43 Our probabilistic Lipschitz condition is as weak as possible and we believe that establishing tight bounds will form
44 an interesting future work. To this end, we believe that stronger assumptions would yield a better exponent in $MSE$,
45 which is why for simplicity we opted for presenting $MSE$ instead of $MSE^{1/4}$ in the paper. We thank the reviewer for
46 suggesting the study of CCA of $k^{-1/2}$, $\|w\|(k/n)^{1/d}$ and $MSE^{1/4}$ (or some other exponent), since it could be another
47 way of understanding the tightness of the bound for $k > 1$, providing an interesting future direction. **With all that in**
48 **mind, we will definitely include the above reasoning on the tightness and the challenges involved in evaluating**
49 **it, as a discussion in the final version of this manuscript.**

50 **Reviewer 4.** We agree that the figures should be more self-contained and we will address this in the final version.

---

[1] file code/results/`<DATASET>`/knn_accuracy/`<EMBEDDING>`.csv