Dr. Feilim Mac Gabhann and Dr. Jason Papin
Editors-in-Chief
PLOS Computational Biology

January 30, 2024

Dear Editors-in-Chief,

We wish to submit an original research article titled *Calibrating dimension reduction hyperparameters in the presence of noise* for consideration by *PLOS Computational Biology*.

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere. All authors have approved the manuscript and agree with its submission to *PLOS Computational Biology*.

In this paper, we discuss the overfitting problem in the context of dimension reduction and demonstrate how its disregard leads to miscalibration of t-SNE/UMAP hyperparameters. More specifically, we show low values of perplexity/n_neighbors tend to overfit the data. This is especially troublesome because nonlinear dimension reduction techniques such as t-SNE and UMAP are known to produce unfaithful results when mishandled.

We originally submitted this paper to *PLOS Computational Biology* at the end of last year. Dr. Sushmita Roy graciously provided feedback, stating more examples would increase the scope and relevance of our work. In this revision, we have included experiment results for one more simulated dataset and two more real datasets, as well as more thorough experimentation of previous datasets. Our experiments now include a CyTOF, scRNA-seq, and microbiome dataset, each with promising results. We believe the addition of more examples should convince readers our work is relevant to all types of high-dimensional data.

We have no conflicts of interest to disclose. Thank you for your consideration of this manuscript.

Sincerely,

Justin Lin
PhD Candidate, Department of Mathematics
Indiana University
linjus@iu.edu

Julia Fukuyama,
Assistant Professor, Department of Statistics
Indiana University
jfukuyam@iu.edu