

# An Interactive Tool for Analyzing High-Dimensional Clusterings

Justin Lin

Department of Mathematics, Indiana University  
and

Julia Fukuyama

Department of Statistics, Indiana University

April 7, 2025

## **Abstract**

The text of your abstract. 200 or fewer words.

*Keywords:* Dimension Reduction, R package, R Shiny

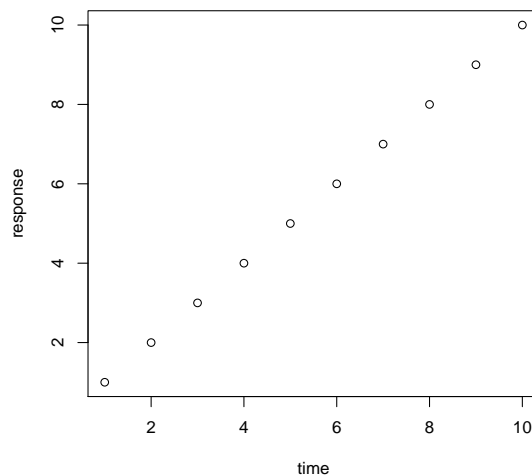


Figure 1: Consistency comparison in fitting surrogate model in the tidal power example.

Table 1: D-optimality values for design  $X$  under five different scenarios.

one	two	three	four	five
1.23	3.45	5.00	1.21	3.41
1.23	3.45	5.00	1.21	3.42
1.23	3.45	5.00	1.21	3.43

## 1 Introduction

Body of paper. Margins in this document are roughly 0.75 inches all around, letter size paper.

- Note that figures and tables (such as Figure 1 and Table 1) should appear in the paper, not at the end or in separate files.
- In the latex source, near the top of the file the command `\newcommand{\blind}{1}` can be used to hide the authors and acknowledgements, producing the required blinded version.

- Remember that in the blind version, you should not identify authors indirectly in the text. That is, don't say "In Smith et. al. (2009) we showed that ...". Instead, say "Smith et. al. (2009) showed that ...".
- These points are only intended to remind you of some requirements. Please refer to the instructions for authors at <http://amstat.tandfonline.com/action/authorSubmission?journalCode=ubes20&page=instructions#>
- For more about ASA style, please see <http://journals.taylorandfrancis.com/amstat/asa-style-guide/>
- If you have supplementary material (e.g., software, data, technical proofs), identify them in the section below. In early stages of the submission process, you may be unsure what to include as supplementary material. Don't worry—this is something that can be worked out at later stages.

## 2 Methods

Don't take any of these section titles seriously. They're just for illustration.

## 3 Verifications

This section will be just long enough to illustrate what a full page of text looks like, for margins and spacing.

Chen and Friedman (2017)

The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. **With this spacing we have 26 lines per page.** The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick

brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog.

[illegible][illegible]

The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the  
lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped  
over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox  
jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick  
brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The  
quick brown fox jumped over the lazy dog.

The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the  
lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped  
over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox  
jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick  
brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The  
quick brown fox jumped over the lazy dog.

[illegible][illegible][illegible]

The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the  
lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped  
over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox  
jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick  
brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The  
quick brown fox jumped over the lazy dog.

The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the  
lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped

over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog.

The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog. The quick brown fox jumped over the lazy dog.

## 4 Conclusion

### SUPPLEMENTARY MATERIAL

**Title:** Brief description. (file type)

**HIV data set:** Data set used in the illustration of MYNEW method in Section 3.2. (.txt file)

## References

- Aggarwal, C., A. Hinneburg, and D. Keim (2001). On the surprising behavior of distance metrics in high dimensional space. In *ICDT '01: Proceedings of the 8th international conference on database theory*, pp. 420–434.
- and Fernanda Viégas, M. W. and I. Johnson. How to use t-SNE effectively. <https://distill.pub/2016/misread-tsne/>.
- Becht, E., L. McInnes, J. Healy, C.-A. Dutertre, I. Kwok, L. Ng, F. Ginhoux, and E. Newell (2019). Dimensionality reduction for visualizing single-cell data using umap. *Nature Biotechnology* 37, 38–44.

- Bhattacharya, B. (2019). A general asymptotic framework for distribution-free graph-based two-sample tests. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 81(3), 575–602.
- Chen, H., X. Chen, and Y. Su (2018). A weighted edge-count two-sample test for multivariate and object data. *Journal of the American Statistical Association* 113(523), 1146–1155.
- Chen, H. and J. Friedman (2017). A new graph-based two-sample test for multivariate and object data. *Journal of the American Statistical Association* 112(517), 397–409.
- Coenen, A. and A. Pearce. Understanding umap. <https://pair-code.github.io/understanding-umap/>.
- Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine* 29(6).
- Di Gesù, V. and V. Starovoitov (1999). Distance-based functions for image comparison. *Pattern Recognition Letter* 20(2), 207–214.
- Diaconis, P. and D. Freedman (1984). Asymptotics of graphical projection pursuit. *The Annals of Statistics* 12(3), 783–815.
- Friedman, J. and L. Rafsky (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Annals of Statistics* 7(4), 697–717.
- González, I., S. Déjean, P. Martin, and A. Baccini (2008). Cca: An R package to extend canonical correlation analysis. 23(12).
- Gower, J. and G. Ross (1969). Minimum spanning trees and single linkage cluster analysis. *Journal of the Royal Statistical Society Series C: Applied Statistics* 18(1), 54–64.

- King, B. and B. Tidor (2009). MIST: Maximum information spanning trees for dimension reduction of biological data sets. *Bioinformatics* 25(9), 1156–1172.
- Li, S. (2011). Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics* 4(1), 66–70.
- McInnes, L., J. Healy, N. Saul, and L. Großberger (2018). UMAP: Uniform manifold approximation and projection. *The Journal of Open Source Software* 3(3).
- Probst, D. and J.-L. Reymond (2020). Visualizing very large high-dimensional data sets as minimum spanning trees. *Journal of Cheminformatics* 12(12).
- Qurashi, A., V. Holmes, and A. Johnson (2020). Document processing: methods for semantic text similarity analysis. *IEEE Xplore*.
- Robinson, D. and L. Foulds (1981). Comparison of phylogenetic trees. *Mathematical Biosciences* 53, 131–147.
- Rozál, G. and J. Hartigan (1994). The MAP test for multimodality. *Journal of Classification* 11, 5–36.
- Stuetzle, W. (2003). Estimating the cluster tree of a density by analyzing the minimal spanning tree of a sample. *Journal of Classification* 20, 25–47.
- Tenenbaum, J., V. de Silva, and J. Langfor (2000). A global geometric framework for nonlinear dimensional reduction. *Science* 290(2319).
- Tuzhilina, E., L. Tozzi, and T. Hastie (2023). Canonical correlation analysis in high dimensions with structured regularization. *Statistical Modelling* 23(3), 203–227.
- Wong, M., D. Ong, F. Lim, K. Teng, N. McGovern, S. Narayanan, W. Ho, D. Cerny, H. Tan, R. Anicete, B. Tan, T. Lim, C. Chan, P. Cheow, S. Lee, A. Takano, E.-H. Tan, J. Tam, E. Tan, J. Chan, and E. Newell (2016). A high-dimensional atlas of human T cell



diversity reveals tissue-specific trafficking and cytokine signatures. *ScienceDirect* 45(2), 442–456.