Exploring Factors Affecting Life Expectancy Using Linear Techniques

Spencer Slaton, Leon Weingartner

*Abstract*—We recreate the study of statistical significance testing within the context of neural networks. This research was produced by Dennis Ulmer, Christian Hardmeier, and Jes Frellsen with a paper titled “Deep Significance – Easy and Meaningful Statistical Significance Testing in the Age of Neural Networks” ICLR (2022) *[1].* The authors comment on the lack of statistically rigorous testing within the rapidly expanding field of deep learning. To rectify this, they provide a Python package for implementing a novel and powerful significance test called *Almost Stochastic Order* (ASO).Here, we will revisit the ASO statistic and compare it to a variety of well-established significance tests across numerous scenarios to illustrate the usefulness of ASO and statistical testing in general under machine learning architectures including a simulation study using convolutional neural networks (CNNs) and multi-layer perceptrons (MLPs).

# INTRODUCTION

W

hen choosing between two different machine learning architectures, the industry standard uses a single

# Project Goals

## Data

## Analysis Topics

## Model Diagnostics

## Final Regression Model

## Additional Work

# Conclusion

Statistical significance testing in machine learning is an underused model comparison tool that can lead to more reliable model selection and strengthen ML research. We’ve

References

1. D. Ulmer, C. Hardmeier, J. Frellsen “Easy and Meaningful Statistical Significance Testing in the Age of Neural Networks” URL https://arxiv.org/abs/2204.06815
2. Nils Reimers and Iryna Gurevych. Why comparing single performance scores does not allow to draw conclusions about machine learning approaches. arXiv preprint arXiv:1803.09578, 2018.
3. Rotem Dror, Gili Baumer, Segev Shlomov, and Roi Reichart. The hitchhiker’s guide to testing statistical significance in natural language processing. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1383–1392, 2018
4. Alessio Benavoli, Giorgio Corani, Janez Demšar, and Marco Zaffalon. Time for a change: a tutorial for comparing multiple classifiers through bayesian analysis. The Journal of Machine Learning Research, 18(1):2653–2688, 2017.
5. Rotem Dror, Segev Shlomov, and Roi Reichart. Deep dominance - how to properly compare deep neural models. In Anna Korhonen, David R. Traum, and Lluís Màrquez (eds.), Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers, pp. 2773–2785. Association for Computational Linguistics, 2019. doi: 10.18653/v1/p19-1266. URL <https://doi.org/10.18653/v1/p19-1266>.
6. Taylor Berg-Kirkpatrick, David Burkett, Dan Klein, An Emperical Investigation of Statistical Significance in NLP. URL <https://aclanthology.org/D12-1091/>
7. **Project Overview Video URL:** [**https://www.youtube.com/watch?v=5Sb3s7QCiV4**](https://www.youtube.com/watch?v=5Sb3s7QCiV4)
8. Github code files URL: https://github.com/sbslaton/230-254-Final-Projects/tree/main/254