An Interpretable Neural Network-based Nonproportional Odds Model for Ordinal Regression

Akifumi Okuno[1,2], Kazuharu Harada[3]

[1]Institute of Statistical Mathematics, [2]RIKEN AIP, [3]Tokyo Medical University

Abstract

This study proposes an interpretable neural network-based nonproportional odds model (N3POM) for ordinal regression. N3POM is different from conventional approaches to ordinal regression with nonproportional models in several ways: (a) N3POM is defined for both continuous and discrete responses, whereas standard methods typically treat the continuous variables as if they were discrete, (b) instead of estimating response-dependent finite-dimensional coefficients of linear models from discrete responses as is done in conventional approaches, we train a nonlinear neural network to serve as a coefficient function. Thanks to the neural network, N3POM offers flexibility while preserving the interpretability of conventional ordinal regression. We establish a sufficient condition under which the predicted conditional cumulative probability locally satisfies the monotonicity constraint over a user-specified region in the covariate space. Additionally, we provide a monotonicity-preserving stochastic (MPS) algorithm for effectively training the neural network. This talk is based on our paper (Okuno and Harada, 2024).

References

Okuno, A. and Harada, K. (2024). An interpretable neural network-based nonproportional odds model for ordinal regression. *Journal of Computational and Graphical Statistics*, 33(4):1454–1463.