DELAWARE Data Science SYMPOSIUM NOV 17, 2021

UNIVERSITY OF DELAWARE DATA SCIENCE INSTITUTE

Dimensionality reduction of autism data using a basic autoencoder

Ginnie Sawyer-Morris, Human Development & Family Sciences, University of Delaware

Introduction and motivation

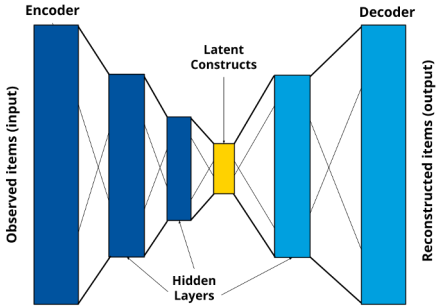
Human behavior is complex. Information captured through linear modeling, on its own, is insufficient to explain the complex interactions that likely exist across psychological constructs.

In autism spectrum disorder (ASD) research, behavioral and symptoms questionnaires are numerous, complex, and multifaceted. Furthermore, fields within ASD each have their own set of theoretical constructs (Ewen, 2020), and the relevant degree and redundancy of information across items within theorized constructs remain unknown.

Large dynamic databases coupled with analytic tools such as unsupervised AI offer the ability to explore complex, potentially nonlinear interactions, through methods such as dimension reduction.

Methodology

An autoencoder was used to represent autism questionnaire data via lower dimensional summaries (latent dimensions = 4, 8, & 16). The Tensorflow library was utilized to fit the autoencoder using adaptive moment estimation optimization. The algorithm was set with a step size of 100, 500 epochs and batch sizes of 256. The number of hidden nodes and network architecture were explored as described later.



Findings and Conclusions

The final analytic sample included 22,662 individuals with autism (75.2% male; 85% < age 18), ranging in age from 0 – 92 years (mean = 11.5; median = 9.3).

The best fitting model was one with 16 dimensions and one decoded layer (i.e., Model 16.1).