On the utility of scoring multidimensional assessments using variational autoencoders

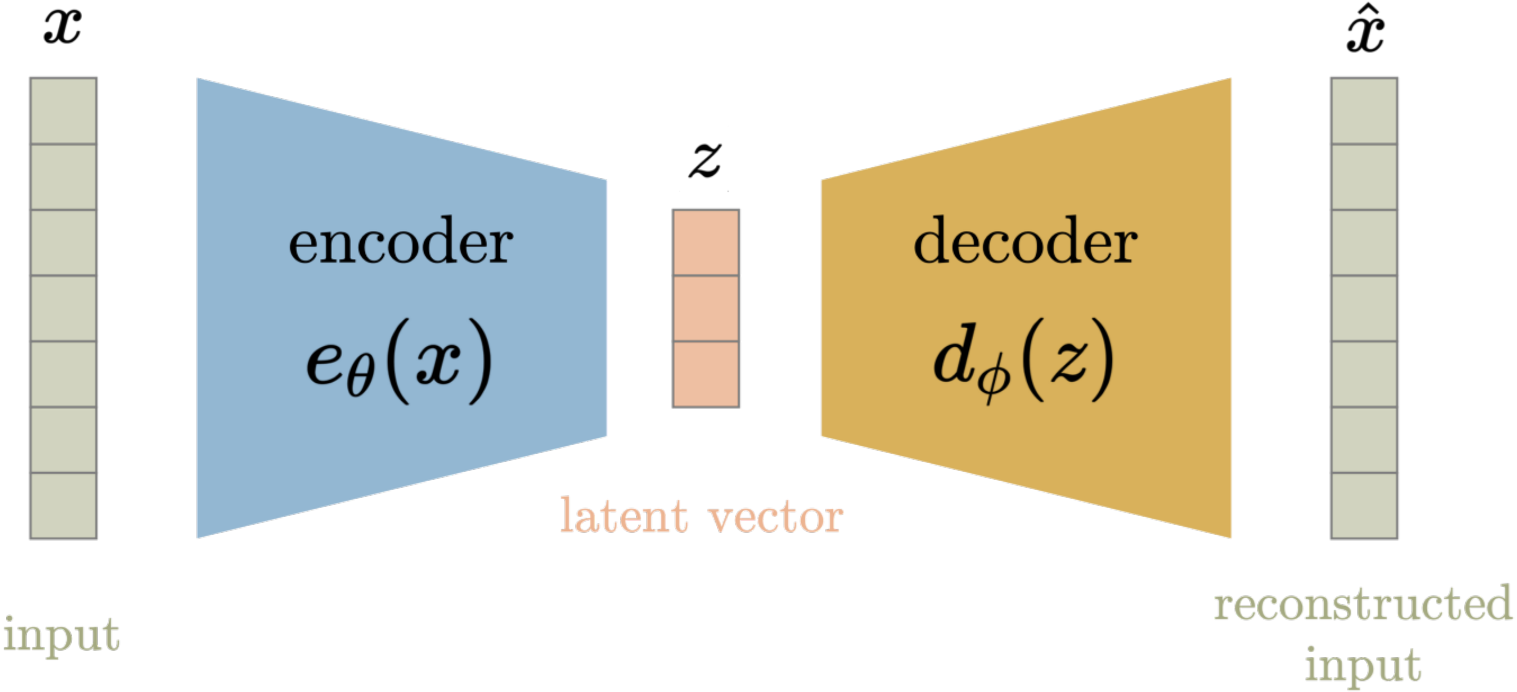
Sebastian Marin

UNIVERSITY OF MINNESOTA
Driven to Discover℠

What is an autoencoder?

- Autoencoders consist of two neural networks—an encoder network and a decoder network
  - Data are input into an encoder network, and the network compresses the data into a lower-dimensional, latent space
  - A decoder is trained to best reconstruct the original inputs to the encoder network from the latent space

What is an autoencoder?

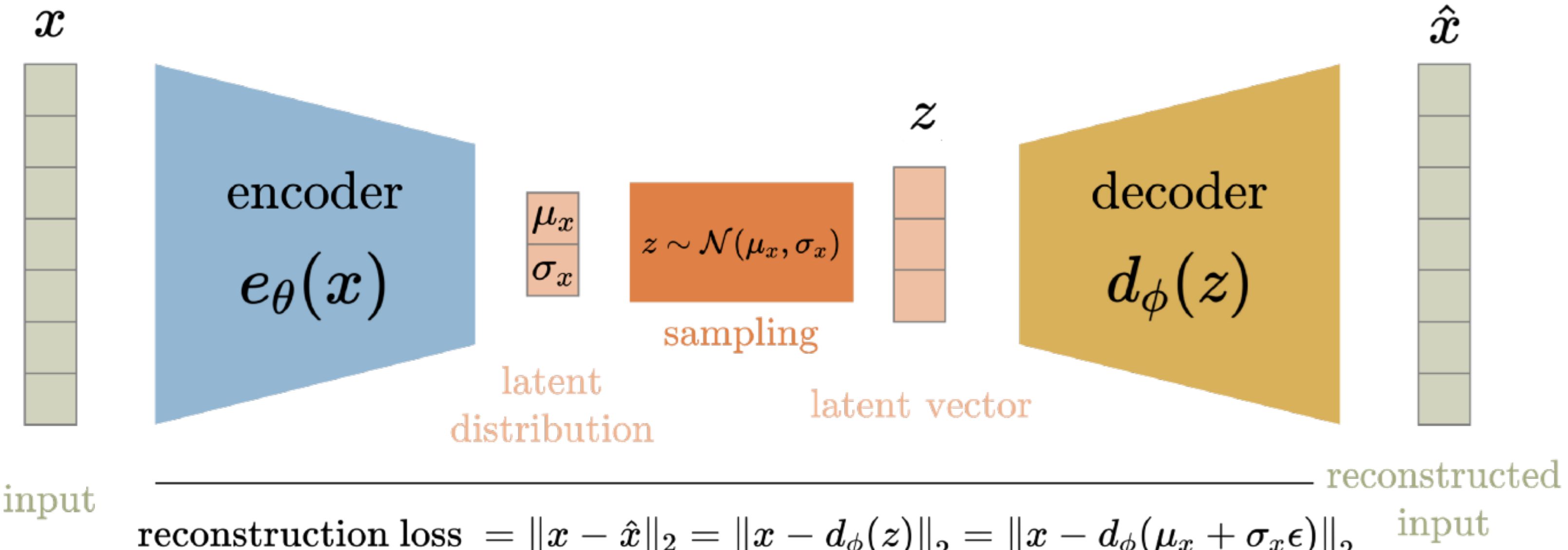


$$loss = \|x - \hat{x}\|_2 = \|x - d_\phi(z)\|_2 = \|x - d_\phi(e_\theta(x))\|_2$$

What is a variational autoencoder?

- VAEs differ from regular autoencoders due to generative modeling
  - Encoder fits data into to a normal distribution at lower-dimensional space
  - The output of the encoder corresponds to the probability distribution parameters (mean, variance) of the latent variables
  - Distributions parameters from the latent space are sampled
  - Sampled distributions are fed forward through the decoder
  - New data are produced optimizing towards
    - 1) reconstruction of the original data input
    - 2) similarity of the original data distribution

What is a variational autoencoder?



$$\text{reconstruction loss} = \|x - \hat{x}\|_2 = \|x - d_\phi(z)\|_2 = \|x - d_\phi(\mu_x + \sigma_x \epsilon)\|_2$$

$$\mu_x, \sigma_x = e_\theta(x), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$similarity\ loss = KL\ Divergence = D_{KL}(\mathcal{N}(\mu_x, \sigma_x) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I}))$$

$$loss = reconstruction\ loss + similarity\ loss$$

What are the advantages of using VAEs?

- VAEs produce exact approximations of parameters recovered from multidimensional 2PL models
- Unity is achieved between theta estimates of each factor for 2PL models and VAEs
- VAEs require way less compute
- VAEs recover parameters at faster speeds (by orders of magnitude)
- VAE architecture can be built to meet IRT assumptions

Applied example

- VAEs have recently been explored in educational assessment for estimating parameters of cognitive diagnostic models (Curi et al., 2019)
- Cognitive diagnostic models (CDMs) are a class of models that have been used in ed. assessment to classify skill mastery (e.g., skill mastered or not)
- A Q-matrix is commonly used in CDMs
  - The Q-matrix is a **binary** matrix that defines the relationship between the items of an assessment and the latent traits (or skills) required to obtain a correct answer
  - Dimensions of a Q matrix are $I \times K$, where
    - K is the number of latent traits (or skills evaluated in the test)
    - I is the number of items in the test
    - $Q_{i,k} = 1$ when item i requires skill k
- The Q-matrix is specified by SMEs

Q-matrix example

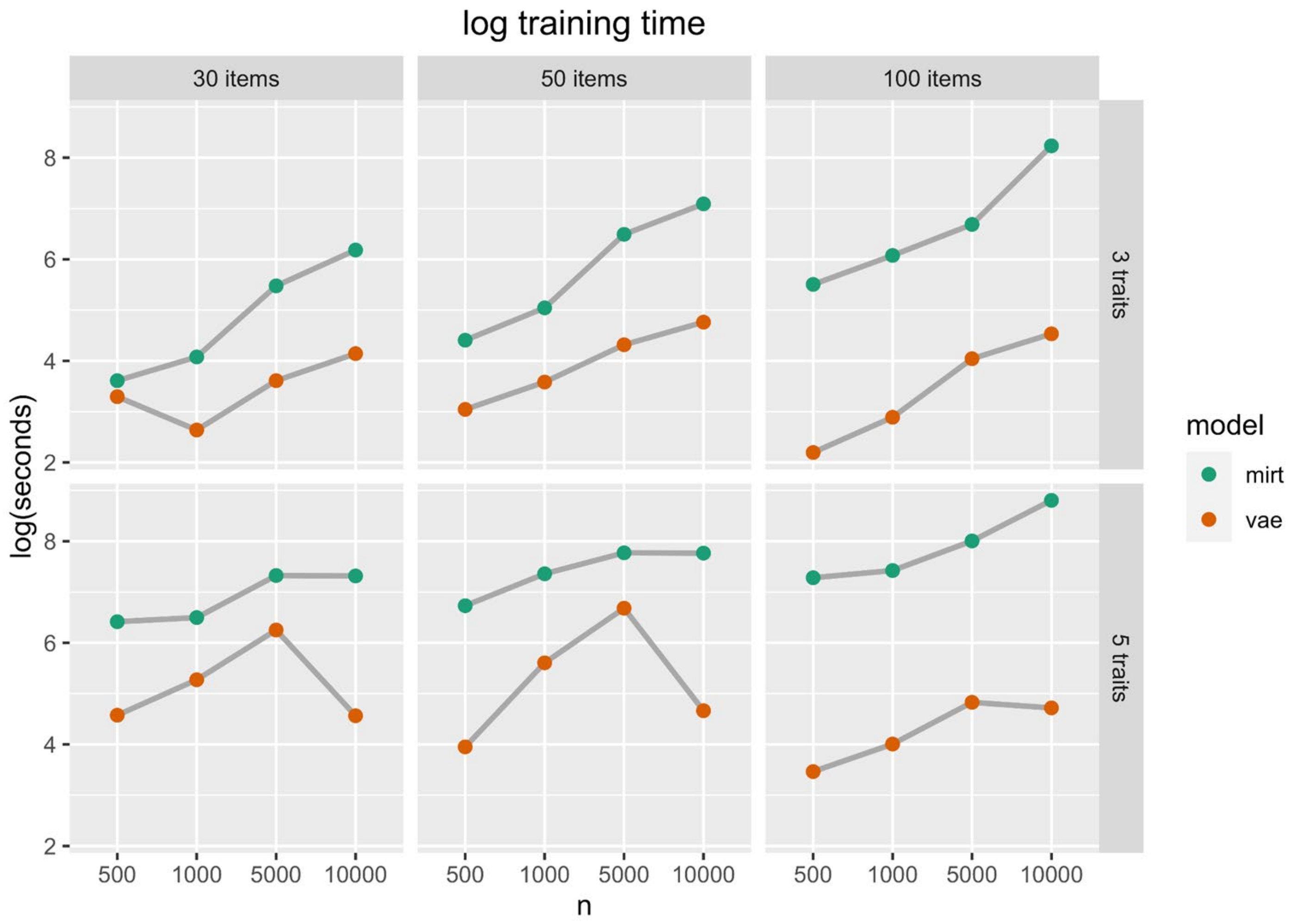
	Skill 1	**Skill 2**	**Skill 3**	**Skill 4**	**Skill 5**
Item 1	0	0	1	0	0
Item 2	1	0	0	1	0
Item 3	0	0	1	1	1
…	…	…	…	…	…
Item *N*	0	1	0	0	1

Applied example cont'd

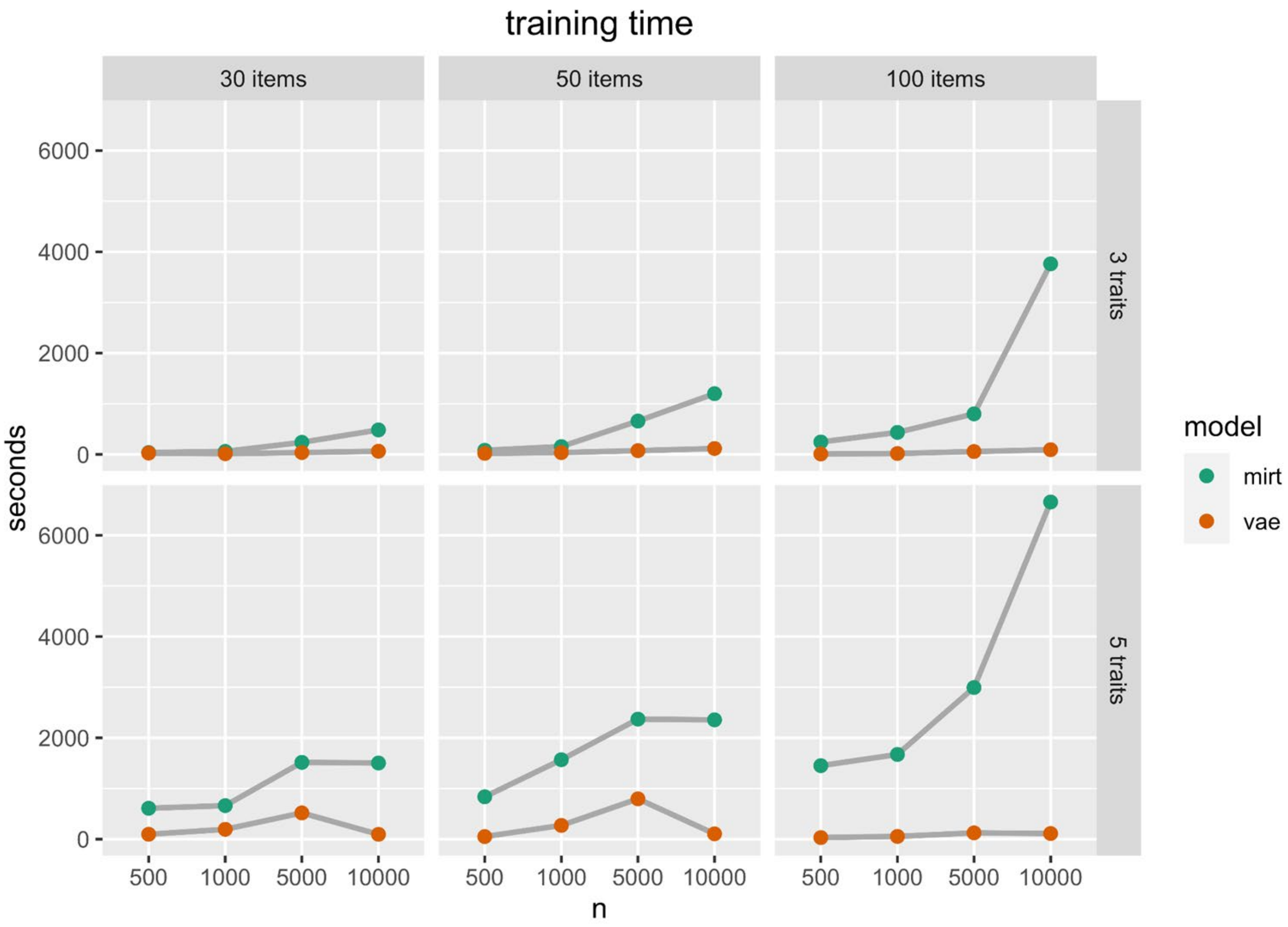
- To estimate IRT parameters, several modifications are made to the VAE architecture (Curi et al., 2019; Converse et al., 2019; Converse et al., 2021):
  - No hidden layer in the decoder (i.e., latent vectors → output)
  - Sigmoidal activation function on the output layer nodes (with non-negative weights)
  - A Q-matrix to determine the connections between the latent traits and the output items
    - Latent vector → output are specified and **not** fully connected

VAE experimenting

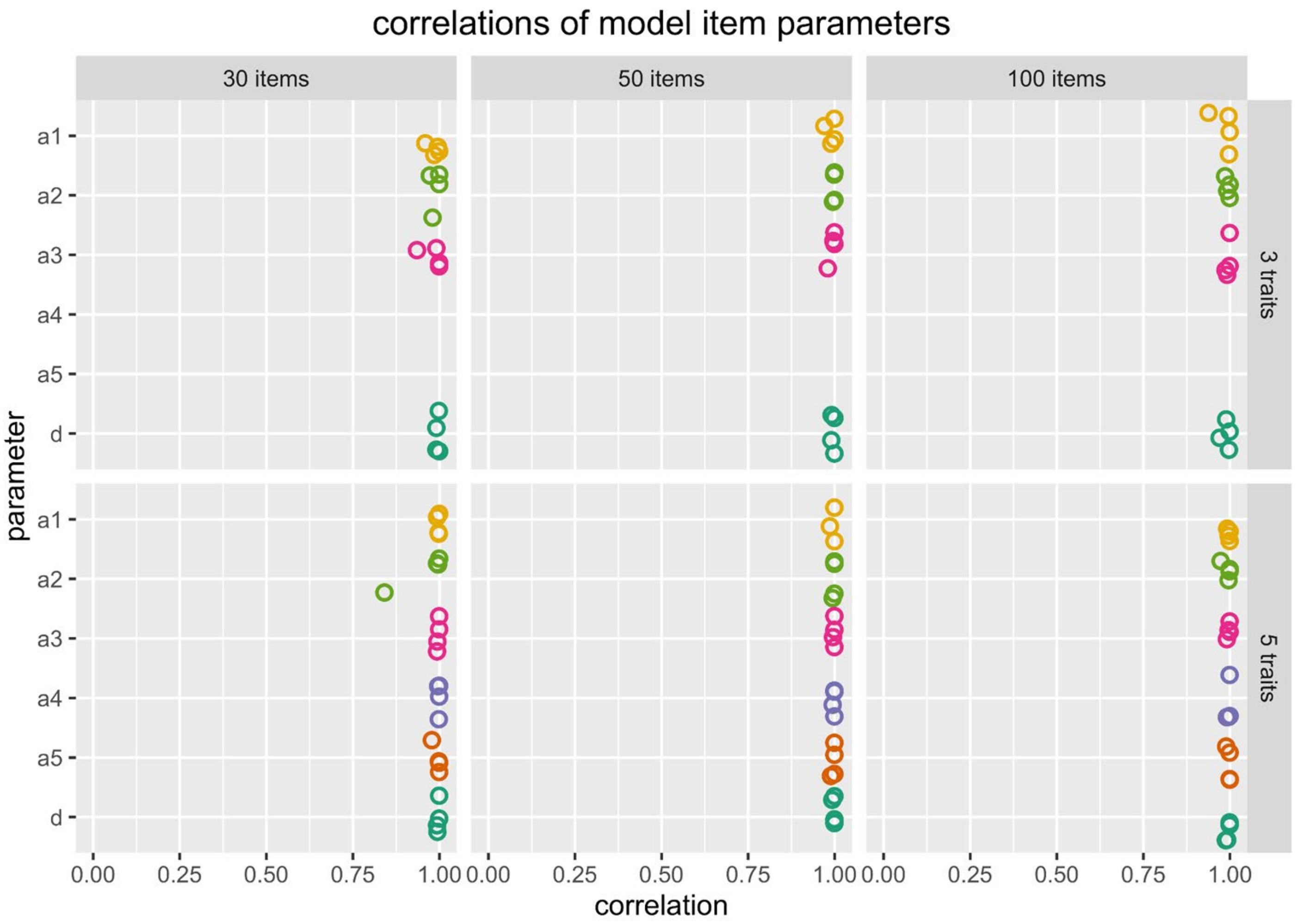
- Data
  - Simulated response data and Q matrices using **`simcdm`** package in R
  - 30, 50, 100 items
  - 3, 5 factors (correlated)
- IRT model
  - 2PL models built using **`mirt`** package in R
  - MIRT model specified to Q-matrix structure
  - Monte Carlo EM estimation (1000 cycles)
- VAE
  - VAE models built using **`ML2Pvae`** package in R
  - 2 encoder hidden layers
    - 3 factors: 16 nodes (hidden 1), 8 nodes (hidden 2)
    - 5 factors: 24 nodes (hidden 1), 16 nodes (hidden 2)
  - Epoch number was tuned
  - Batch size was tuned

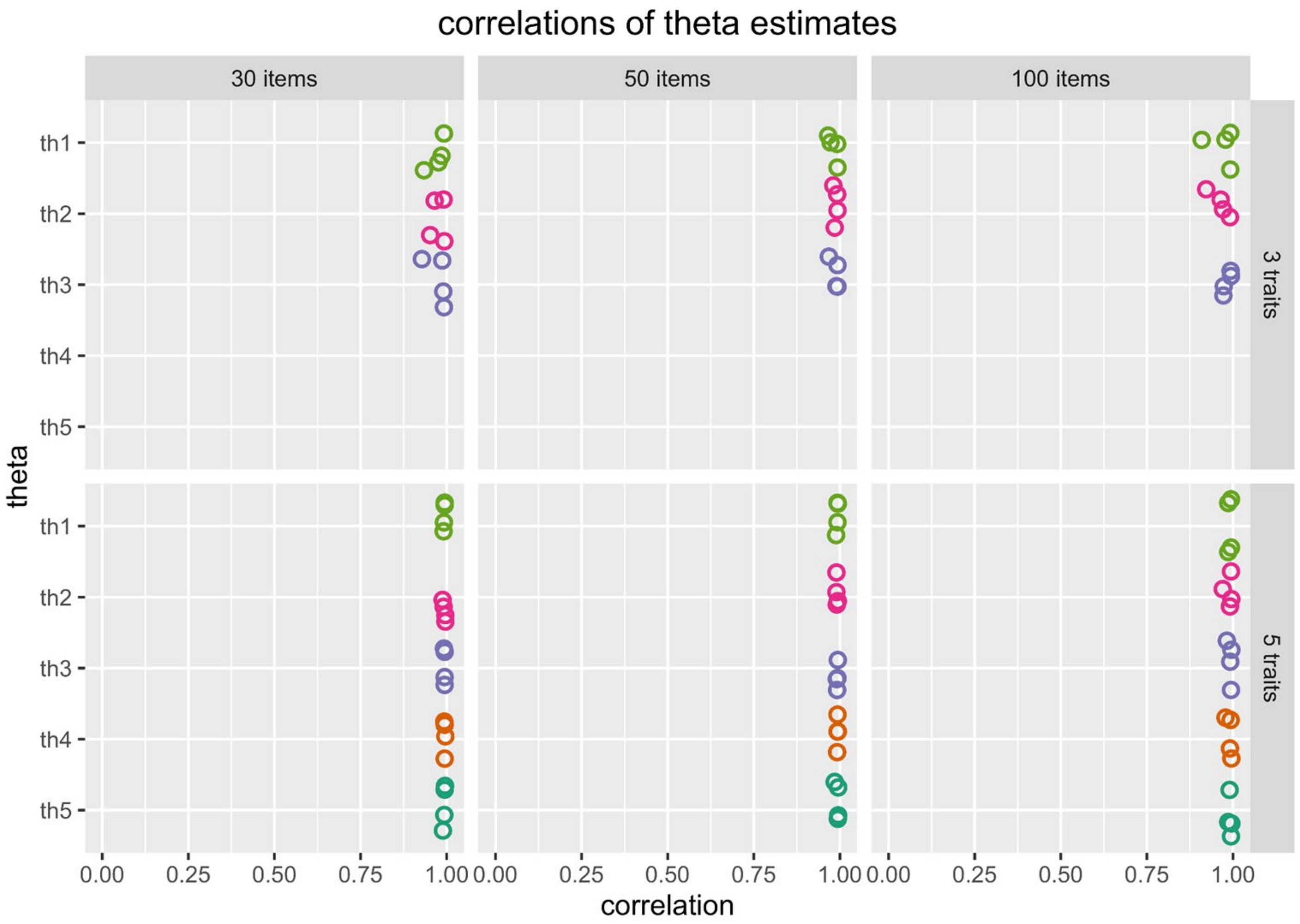


log training time
30 items
50 items
100 items
3 traits
5 traits
log(seconds)
8
6
4
2
500
1000
5000
10000
n
model
mirt
vae



training time
30 items
50 items
100 items
3 traits
5 traits
seconds
6000
4000
2000
0
n
500
1000
5000
10000
model
mirt
vae


correlations of model item parameters
30 items
50 items
100 items
3 traits
5 traits
parameter
a1
a2
a3
a4
a5
d
correlation
0.00
0.25
0.50
0.75
1.00


correlations of theta estimates
30 items
50 items
100 items
3 traits
5 traits
theta
th1
th2
th3
th4
th5
correlation
0.00
0.25
0.50
0.75
1.00

Why is this useful for selection?

- VAEs can be built to approximate IRT parameters at faster speeds and with less compute
  - Useful for high dimensional data, e.g., large item banks and large response data sets
- CDMs can be a framework for scoring assessments used in employee selection (e.g., SJTs; Sorrel et al., 2016), employee training, and professional credentialing/certification
- Highlights the notion of multidimensional test information
  - It is underused in selection, and VAEs can challenge us to think about ways to leverage multidimensional information

References

- Balamuta, J. J., Culpepper, S. A., and Hudson A. (2019) simcdm: Simulate cognitive diagnostic model (CDM) data. https://cran.r-project.org/package=simcdm.
- Converse, G. (2020). ML2Pvae: Variational autoencoder models for IRT parameter estimation. R package version 1.0.0. https://CRAN.R-project.org/package=ML2Pvae
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. Journal of Statistical Software, 48(6), 1-29. doi:10.18637/jss.v048.i06
- Converse, G., Curi, M., & Oliveira, S. (2019, June). Autoencoders for educational assessment. In International Conference on Artificial Intelligence in Education (pp. 41-45). Springer, Cham.
- Converse, G., Curi, M., Oliveira, S., & Templin, J. (2021). Estimation of multidimensional item response theory models with correlated latent variables using variational autoencoders. Machine Learning, 110(6), 1463-1480.
- Curi, M., Converse, G. A., Hajewski, J., & Oliveira, S. (2019, July). Interpretable variational autoencoders for cognitive models. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
- Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models. Organizational Research Methods, 19(3), 506-532.

Resources

Code

https://bit.ly/siop22-ml2pvae