Machine learning for estimation in IRT models

Mariana Curi
University of São Paulo, Brazil
mcuri@icmc.usp.br

work developed in collaboration with ACTNext, ACT Inc.:
Benjamin Deonovic
Pravin Chopade
Gunter Maris
thanks to ACTNext and FAPESP (2018/21063-4) for funding support

Content

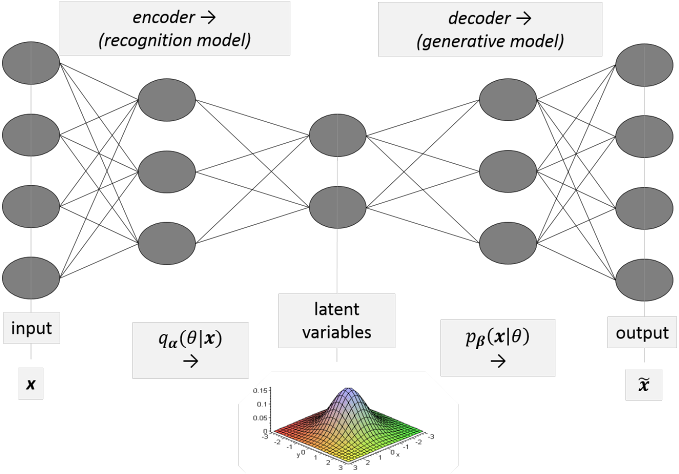
- Motivation
- Variational autoencoder (VAE)
- Proposed estimation method: *VAE-Q2PL*
- Simulation study
- Conclusions

Motivation

- traditional IRT estimation methods: MCMC and MML
- infeasible for high-dimensional $\boldsymbol{\theta}$ (latent traits)
- maximum of 8 latent traits found in literature
- time consuming
- guide the interpretation of $\boldsymbol{\theta}$ (experts knowledge)

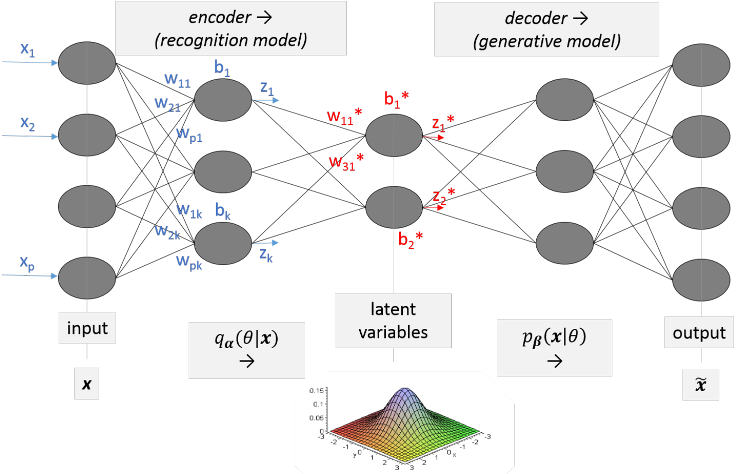
VAE: Kingma and Welling, 2014

UNSUPERVISED deep artificial neural network



VAE: Kingma and Welling, 2014

UNSUPERVISED deep artificial neural network



$$z_k = \frac{1}{1+e^{-\sum_{i=1}^{p} w_{ik}x_i + b_k}} \text{ and } z_1^* = \frac{1}{1+e^{-\sum_{h=1}^{k} w_{h1}^* z_h + b_1^*}}$$

VAE: loss function

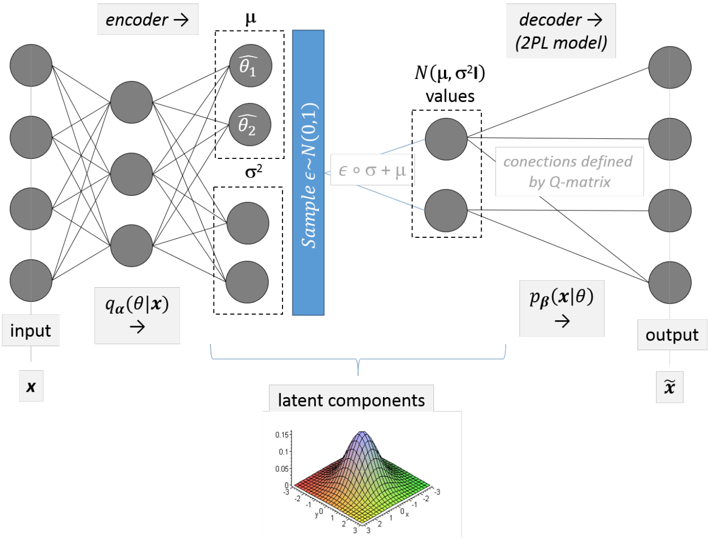
- approach $f(\boldsymbol{\theta}|\boldsymbol{x})$ by $q(\boldsymbol{\theta}|\boldsymbol{x})$ that can reconstruct the data
- loss function:

$$\underset{q(\boldsymbol{\theta}|\boldsymbol{x})\in G}{\operatorname{argmin}} KL[q(\boldsymbol{\theta}|\boldsymbol{x})||f(\boldsymbol{\theta}|\boldsymbol{x})],$$

- analogous to maximize ELBO:

$$E_{\boldsymbol{\theta}\sim q_{\boldsymbol{\alpha}}(\boldsymbol{\theta}|\boldsymbol{x})}[logP_{\beta}(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{\theta})]-KL[q_{\boldsymbol{\alpha}}(\boldsymbol{\theta}|\boldsymbol{x})||f(\boldsymbol{\theta})]$$

Proposed IRT estimation method: VAE-Q2PL



Proposed modification: no hidden layer in the decoder (2PL) and output layer not dense (non-negative weights).

VAE-Q2PL

Curi *et al*, 2019

- 28 item assessment with 3 latent skills
- 5K and 10K individuals: good results

Converse *et al*, 2019

- VAE better than (regular) autoencoder (Guo *et al*, 2017)

Simulations: Scenario

# Dimensions (d) and Estimation	# items (i) 28	56	90	180	270	360
3 MML	x					
3 VAE	x	x				
11 VAE		x	x			
21 VAE			x	x	x	x

- 2PL multidimensional IRT model
- (i x d) Q-matrix: some elements equal to 0
  3D: 28-item English language (Templin 2007, 2013)
  11D and 21D:

SIMPLE ITEMS	θ_1	COMPLEX ITEMS	
	$+ 1$	2 traits (double)	3 traits (triple)

- j = 1, ..., 10K individuals, *log*N for *a*'s and N(0,1) for *b*'s
- r = 1, ..., 10 replicates
- Pearson ρ, BIAS $= \sum_{r=1}^{10} \frac{\hat{\zeta}_r - \zeta}{10}$, RMSE $= \sqrt{\sum_{r=1}^{10} \frac{(\hat{\zeta}_r - \zeta)^2}{10}}$

Simulations: Scenario

# Dimensions (d) and Estimation	# items (seconds per simulation scenario) 28	56	90	180	270	360
3 MML	469					
3 VAE	27	38				
11 VAE		37	46			
21 VAE			49	54	51	55

- 2PL multidimensional IRT model
- (i x d) Q-matrix: some elements equal to 0

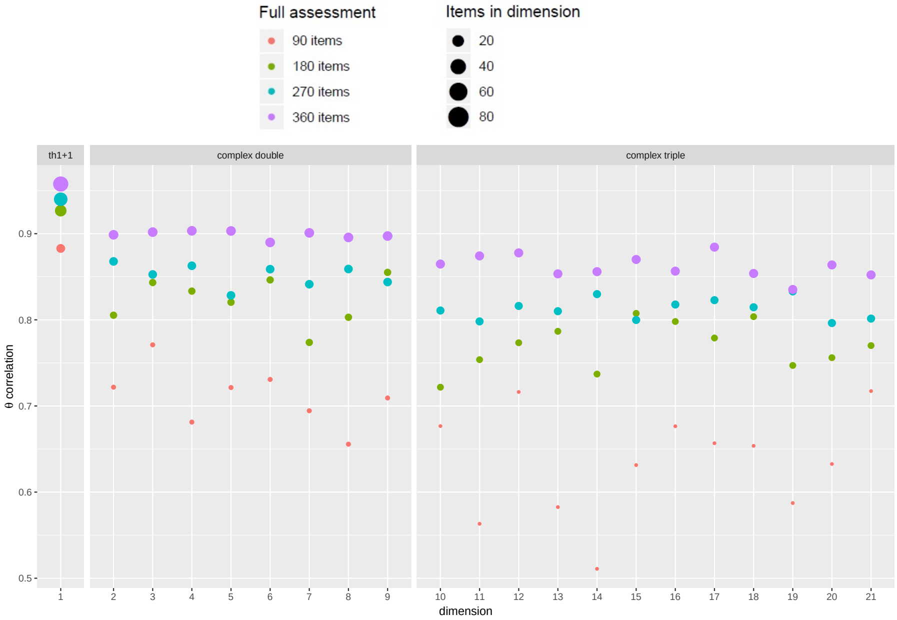
  3D: 28-item English language (Templin 2007, 2013)

  11D and 21D:

SIMPLE ITEMS	θ_1	COMPLEX ITEMS	
	$+1$	2 traits (double)	3 traits (triple)

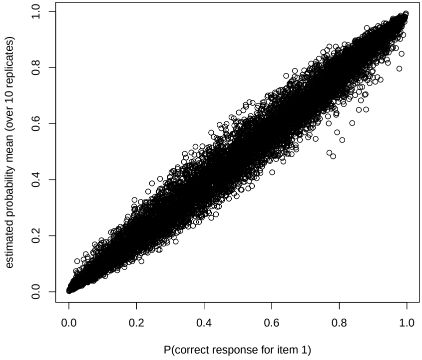
- j = 1, ..., 10K individuals, *log*N for *a*'s and N(0,1) for *b*'s
- r = 1, ..., 10 replicates
- Pearson ρ, BIAS $= \sum_{r=1}^{10} \frac{\hat{\zeta}_r - \zeta}{10}$, RMSE $= \sqrt{\sum_{r=1}^{10} \frac{(\hat{\zeta}_r - \zeta)^2}{10}}$

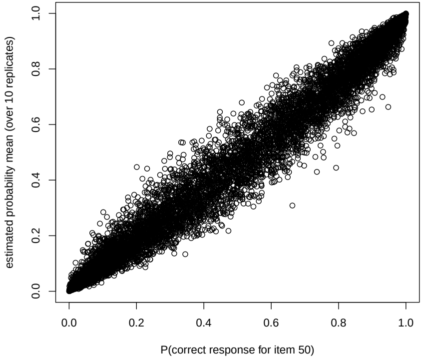
θ REAL AND ESTIMATES CORRELATIONS: 21D



P(correct response): real x estimates

21D and 360 item simulation





REAL X ESTIMATES

BIAS

b parameter

Items		28	56	90	180	270	360
3D	MML	0.004					
3D	VAE	0.13	0.07				
11D	VAE		0.01	-0.10			
21D	VAE			-0.02	-0.03	0.12	-0.01

a parameter

Items		28	56	90	180	270	360
3D	MML	0.01					
3D	VAE	-0.12	-0.19				
11D	VAE		-0.01	-0.15			
21D	VAE			-0.07	-0.32	-0.43	-0.47

θ parameter

Items		28	56	90	180	270	360
3D	MML	0.003					
3D	VAE	0.03	0.02				
11D	VAE		0.00	-0.01			
21D	VAE			-0.01	-0.04	0.14	-0.08

CORRELATION

b parameter

28	56	90	180	270	360	Items
0.999						MML
0.997	0.999					VAE
	0.98	0.98				VAE
		0.98	0.97	0.97	0.97	VAE

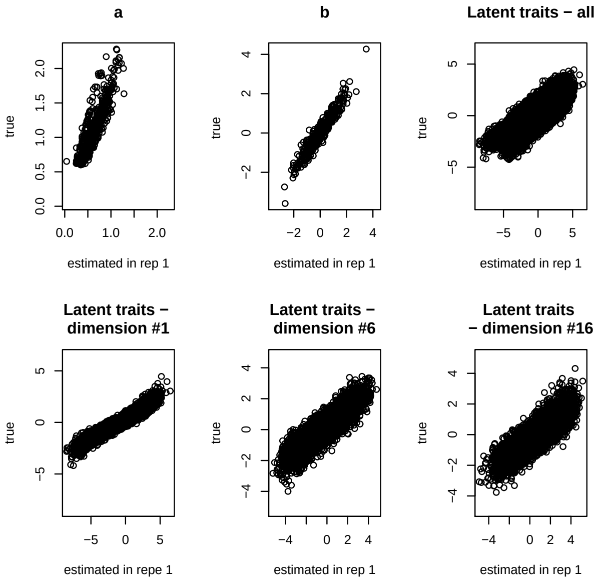
a parameter

28	56	90	180	270	360	
0.997						MML
0.96	0.97					VAE
	0.85	0.91				VAE
		0.71	0.92	0.96	0.96	VAE

θ parameter

28	56	90	180	270	360	
0.82						MML
0.80	0.88					VAE
	0.73	0.80				VAE
		0.67	0.80	0.83	0.88	VAE

Real x estimates: 21D and 360 item simulation



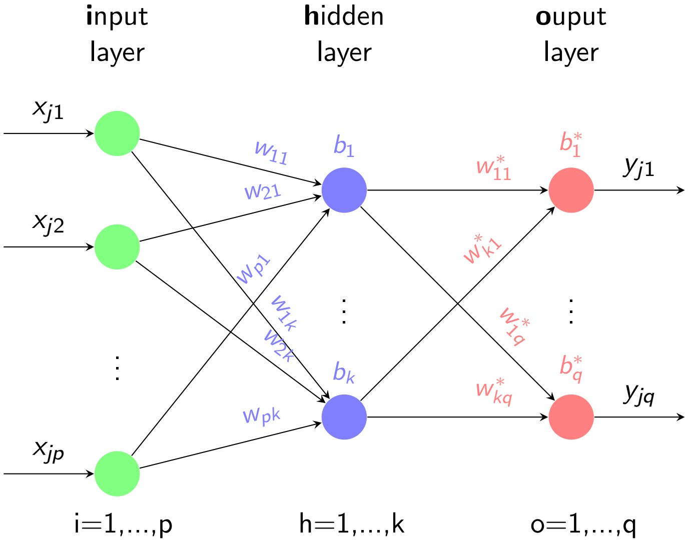
Conclusions

- VAE has a good recovery of the (M2PL model) parameters
- some bias, in particular for discrimination parameters
- good P(correct answer) prediction
- the more items, the better the results
- large sample required (5 to 10K)
- VAE much faster than traditional methods

References

- Guo, Q., Cutumisu, M. and Cui, Y. (2017). A Neural Network Approach to Estimate Student Skill Mastery in Cognitive Diagnostic Assessments. In: Xiangen Hu et al. (eds.) Proceedings of the 10th International Conference on Educational Data Mining. EDM 2017
- Kingma and Welling. Auto-Encoding Variational Bayes. International Conference on Learning Representations. ICLR, 2014
- Curi, Converse, G., M., Hajewsky, J. and Oliveira, S. Interpretable Variational Autoencoders for Cognitive Models. 2019 International Joint Conference on Neural Networks (IJCNN), 2019.
- Converse, G., Curi, M. and Oliveira, S. Autoencoders for Educational Assessment. 20th Annual Conference on Artifical Intelligence in Education (AIED), 2019.

Artificial neural networks (ANN)



j=1,...,n (subject); w_{ih}, w^*_{ho} (weight/parameter); b_h, b^*_o (bias/intercept)

ANN

$$net_{jh} = \sum_{i=1}^{p} w_{ih} x_{ji} + b_h$$

$$f(net_{jh}) = \frac{1}{1 + e^{-net_{jh}}} = z_{jh}$$

$$net^*_{jo} = \sum_{h=1}^{k} w^*_{ho} z_{jh} + b^*_o$$

$$\hat{Y}_{jo} = \hat{P}(Y_{jo} = 1) = f^*(net^*_{jo}) = \frac{1}{1 + e^{-net^*_{jo}}}$$

Activation function $(f(\cdot), f^*(\cdot))$: **sigmoid or logistic function**, hyperbolic tangent, relu, etc
Loss function $(L(\boldsymbol{y}, \hat{\boldsymbol{y}}))$: MSE, **cross-entropy**, etc