High-VAE General

Lee Carlin

May 3, 2022

Project Page

1 Introduction

Goal   Develop a general variational autoencoder framework for high-cardinality and heterogenous tabular data.

2 Roadmap

Table 1: Roadmap

Step	Tasks	Due date

3 Competing Approaches

	Ref.	Architecture	Use-cases	Repo	Datasets
VAEs					
HI-VAE	Nazabal et al. (2020)	Hierarchical Decoder	Imputation	github	1-5
VAEM	Ma et al. (2020)	2-stage: ind. & dep. VAEs	Imputation	github	9-13
VSAE	Gong et al. (2021)	mask & data gen. models	Imputation	n/a	14
RVAE	Akrami et al. (2020, 2022)	Beta-divergence	Outlier robust	github	1,3,4-8
GANs					

4   Datasets:

	Name	Ref.	Summary
1	Adult	UCI-MLR	
2	Breast	UCI-MLR	
3	Credit Default	UCI-MLR	
4	Spam	UCI-MLR	
5	Wine	UCI-MLR	
6	KDDCup 99	KDD	
7	NSL-KDD	CiC	
8	UNSW-NB15	UNSW	
9	Bank	UCI-MLR	
10	Boston	UCI-MLR	
11	Avocado		
12	Energy	UCI-MLR	
13	MIMIC	MIT	de-identified health-related data (~40k)
14	Heart	UCI-MLR	

5   Something

dfd

References

Akrami, H., Aydore, S., Leahy, R. M., and Joshi, A. A. (2020). Robust variational autoencoder for tabular data with beta divergence. *arXiv preprint arXiv:2006.08204*.

Akrami, H., Joshi, A. A., Li, J., Aydöre, S., and Leahy, R. M. (2022). A robust variational autoencoder using beta divergence. *Knowledge-Based Systems*, 238:107886.

Gong, Y., Hajimirsadeghi, H., He, J., Durand, T., and Mori, G. (2021). Variational selective autoencoder: Learning from partially-observed heterogeneous data. In *International Conference on Artificial Intelligence and Statistics*, pages 2377–2385. PMLR.

Ma, C., Tschiatschek, S., Turner, R., Hernández-Lobato, J. M., and Zhang, C. (2020). Vaem: a deep generative model for heterogeneous mixed type data. *Advances in Neural Information Processing Systems*, 33:11237–11247.

Nazabal, A., Olmos, P. M., Ghahramani, Z., and Valera, I. (2020). Handling incomplete heterogeneous data using vaes. *Pattern Recognition*, 107:107501.