VAEs in the Presence of Missing Data

Mark Collier, Alfredo Nazabal, Christopher K.I. Williams

July 16, 2020

Introduction

- The standard VAE formulation assumes no missing data.
- In practice many datasets have missing data.
- Current approaches to applying VAEs to missing data have many significant disadvantages.
- By assuming missing data arises from a corruption process and modelling the observed data as a conditional VAE we solve the disadvantages of the current approaches and get better empirical results.

Standard VAE Graphical Model

Standard VAE assumes fully observed data.

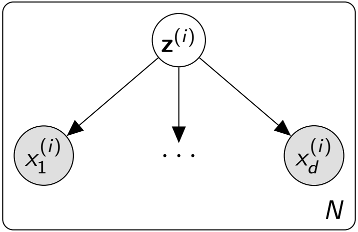

Figure: VAE graphical model for fully observed data.

Previous Approach: Treat Missing Variables as Latents [1]

Under MCAR assumption missing elements can be integrated out of the VAE ELBO.

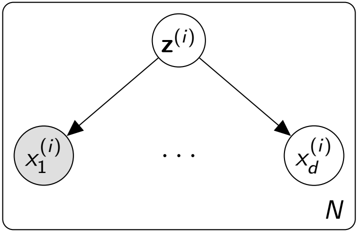

Figure: Missingness as latents

Our Latent Variable Model for Missing Data I

Key idea is to think of missing data as a corruption process.

Our Latent Variable Model for Missing Data II

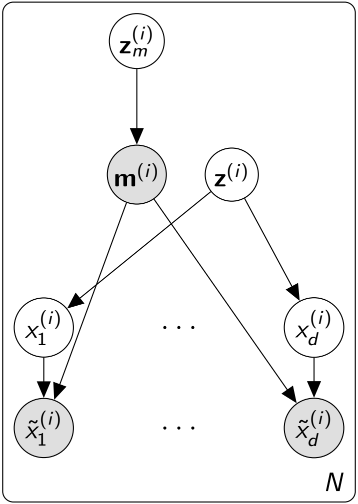

Figure: MCAR corruption process

Our Latent Variable Model for Missing Data III

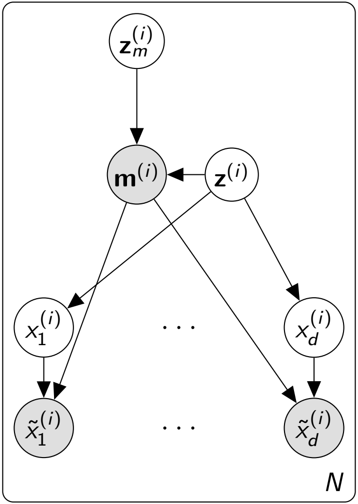

Figure: MNAR corruption process

Our ELBO

We model the observed data as a conditional VAE, conditioned on the missingness mask $\mathbf{m}$ and derive a corresponding VAE ELBO.

$$\log p(\mathbf{x}_o \mid \mathbf{m}) \geq \mathbb{E}_{q(\mathbf{z}|\tilde{\mathbf{x}},\mathbf{m})}\left[\log p(\mathbf{x}_o \mid \mathbf{z}, \mathbf{m})\right] - D_{KL}(q(\mathbf{z} \mid \tilde{\mathbf{x}}, \mathbf{m}) || p(\mathbf{z})). \tag{1}$$

Advantages of Our Approach I

1. The encoder network which in our approach parameterizes $q(\mathbf{z} \mid \tilde{\mathbf{x}}, \mathbf{m})$ is conditioned on $\mathbf{m}$. Thus the encoder network can distinguish between an observed data element and a missing value by looking at the corresponding element in $\mathbf{m}$.
2. As the encoder network has access to $\mathbf{m}$ its output can change based on the missingness pattern. The encoder can approximate a non-linear parameterizarion of the exact Factor Analysis solution of learning a separate encoder network for each missingness pattern [2], while sharing the same encoder weights across all missingness patterns.
3. Our model applies to both **MCAR** and **MNAR** data with no change in the training objective.

Advantages of Our Approach II

4. Our model is computationally efficient with no restrictions on the neural networks architecture used, scaling to high dimensional $\mathbf{x}$ as per standard VAEs.

Example MNIST MCAR Reconstructions

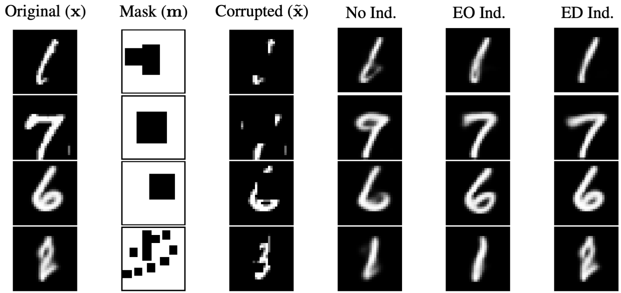

Figure: MNIST MCAR example images. Shown are the original image, missingness mask, the corrupted image and the mean reconstructions provided by the No Ind. EO Ind. and ED Ind. methods.

Example SVHN MCAR Reconstructions

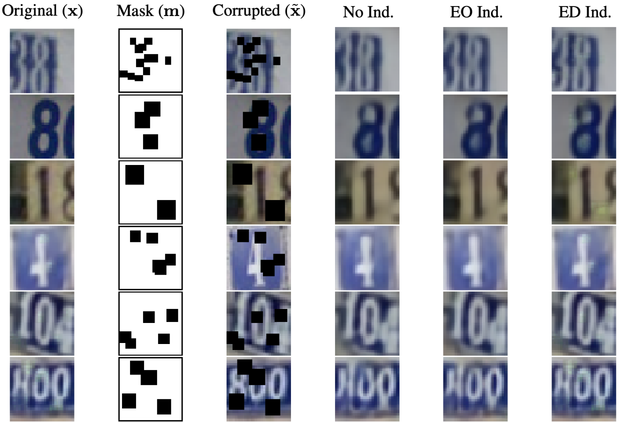

Figure: SVHN MCAR example images. Shown are the original image, missingness mask, the corrupted image and the mean reconstructions provided by the No Ind. EO Ind. and ED Ind. methods.

MNIST Results

Table: MCAR/MNAR results. Each cell contains two values the metric for MCAR missingness and the metric for MNAR missingness. For each metric we conduct a paired sample t-test between the method and ED Ind. * indicates p $<$ 0.001.

Method	**MNIST**	
	$\log p(\mathbf{x}_o \mid \mathbf{m})$	$\mathbb{E}_q\left[\log p(\mathbf{x}_m \mid \mathbf{z}, \mathbf{m})\right]$
No Ind.	$-63.95^*/-64.21^*$	$-64.03^*/-65.66^*$
EO Ind.	$-63.04/-63.25$	$-58.28^*/-59.09^*$
ED Ind.	$\mathbf{-62.99/-63.23}$	$\mathbf{-56.11/-57.01}$

SVHN Results

Table: MCAR/MNAR results. Each cell contains two values the metric for MCAR missingness and the metric for MNAR missingness. For each metric we conduct a paired sample t-test between the method and ED Ind. * indicates p $<$ 0.001.

Method	**SVHN**	
	$\log p(\mathbf{x}_o \mid \mathbf{m})$	$\mathbb{E}_q\left[\log p(\mathbf{x}_m \mid \mathbf{z}, \mathbf{m})\right]$
No Ind.	$-5934.42^*/-5967.43^*$	$-3065.70^*/-3074.08^*$
EO Ind.	$-5853.79/-5863.37$	$-3175.00^*/-3198.89^*$
ED Ind.	$-\mathbf{5828.02}/-\mathbf{5841.13}$	$-\mathbf{2844.40}/-\mathbf{2944.98}$

References

Alfredo Nazabal, Pablo M Olmos, Zoubin Ghahramani, and Isabel Valera.
Handling incomplete heterogeneous data using vaes.
arXiv preprint arXiv:1807.03653, 2018.

Christopher KI Williams, Charlie Nash, and Alfredo Nazábal.
Autoencoders and probabilistic inference with missing data: An exact solution for the factor analysis case.
arXiv preprint arXiv:1801.03851, 2018.