
MIWAE: Deep Generative Modelling and Imputation of Incomplete Data Sets

supplementary material

Pierre-Alexandre Mattei¹ Jes Frellsen¹

1. General implementation details

We used TensorFlow probability (Dillon et al., 2017) and the Adam optimiser (Kingma & Ba, 2014) to train all DLVMs. All neural nets are initialised following the heuristics of Glorot & Bengio (2010) and He et al. (2015). Because we wanted to investigate the properties of the different bounds, all DLVMs are trained without any form of regularisation. Regarding gradient estimation, we used the pathway derivative estimate of Roeder et al. (2017) for MNIST, but found that it led to quite unstable training for the continuous UCI data sets, so we used the regular gradient estimates of Burda et al. (2016). These results are consistent with the recent findings of Tucker et al. (2019), who showed that the pathway derivative estimate is actually biased, but that this bias is almost invisible for MNIST.

2. A MAR (but not MCAR) version of MNIST

We consider a MAR version of MNIST where all bottom halves of the pixels are observed. For each digit, either the top half, top quarter, or second quarter, is missing (depending on the number of white pixels in the bottom half). Specifically, for each digit $\mathbf{x} \in \mathbb{R}^{784}$, we sample a binomial random variable h , with 2 trials and success probability

$$\pi(\mathbf{x}) = \frac{1}{\frac{784}{2}} \left(\sum_{j \in \text{bottom half of } \mathbf{x}} x_j \right) + 0.3.$$

Then, we

- remove the second quarter if $h = 0$,
- remove the top half if $h = 1$,
- remove the first quarter if $h = 2$.

Because $\pi(\mathbf{x})$ only depends on the (always observed) bottom half of \mathbf{x} , this scheme is MAR, but not MCAR.

¹Department of Computer Science, IT University of Copenhagen, Denmark. Correspondence to: Pierre-Alexandre Mattei <pima@itu.dk>, Jes Frellsen <jefr@itu.dk>.

3. A MAR (but not MCAR) UCI experiment

We consider the *Breast* data set, which contains 30 continuous features. After standardisation, we keep the first 15 features, and remove, for each observation $\mathbf{x} \in \mathbb{R}^{30}$, the second 15 features with probability

$$\pi(\mathbf{x}) = \text{sigmoid} \left(\frac{1}{15} \sum_{j=1}^{15} x_j \right).$$

Since $\pi(\mathbf{x})$ only depends on the first, always observed, 15 features, this scheme is MAR, but not MCAR.

We then perform a single imputation experiment similar to the one in the main paper. The results are presented in Table 1. Similarly to the MCAR experiments, MIWAE is more accurate than the other competitors.

	<i>Breast</i>
MIWAE	0.24 (0.04)
missForest	0.31 (0.04)
PCA	0.28 (0.05)
k NN	4.35 (1.93)
Mean	0.99 (0.08)

Table 1. Mean-squared error for single imputation for a MAR version of the *Breast* data set (mean and standard deviations over 5 randomly generated data sets).

References

- Burda, Y., Grosse, R., and Salakhutdinov, R. Importance weighted autoencoders. In *International Conference on Learning Representations*, 2016.
- Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., and Saurous, R. A. TensorFlow distributions. *arXiv preprint arXiv:1711.10604*, 2017.
- Glorot, X. and Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256, 2010.

He, K., Zhang, X., Ren, S., and Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 2014.

Roeder, G., Wu, Y., and Duvenaud, D. Sticking the landing: Simple, lower-variance gradient estimators for variational inference. In *Advances in Neural Information Processing Systems*, pp. 6928–6937. 2017.

Tucker, G., Lawson, D., Gu, S., and Maddison, C. J. Doubly reparameterized gradient estimators for monte carlo objectives. In *International Conference on Learning Representations*, 2019.