Empirical Evaluation of the Performance of CEVAE under Misspecification of the Latent Dimensionality

Patrik Barták *(p.bartak@student.tudelft.nl)*   Supervisors: Stephan Bongers, Jesse Krijthe

1 Background

- Causal effect estimation (Fig. 1) aims to infer the impact of changing a **treatment** *t* on the **outcome** *y*.
- **Individualized treatment effect** (ITE) measures strength of this relationship.
- Confounders *x* create **statistical bias** by introducing spurious **associations** between *t* and *y*.
- Causal ML models aims to infer the **true effect** by training on datasets with **observed confounders**.
- CEVAE, the Causal Effect Variational Autoencoder [1] (Fig. 2) aims to infer **causal effects** even with **unobserved confounders** *z*, using **proxy variables** *x*.
- Model requires specification of the **dimension** of this confounder to model it (further called model dimensionality), yet authors claim it to **make less assumptions** than other methods.
- Focusing on **dimensionality**, how does CEVAE perform when this specification does not correspond to the specification used to generate **synthetic data**?

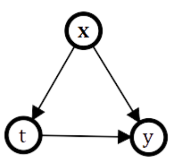

Fig. 1: Causal diagram for causal inference.

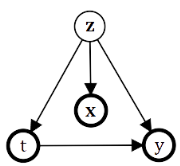

Fig. 2: Causal diagram for CEVAE.

2 Research Question

What is the impact of **misspecified dimensionality** on the **performance** of CEVAE?

3 Methodology

- Similar to factorial design with 3 factors.
- Two experiments, each with a different **data generating processes.**
- For each, vary the **dimension** of the **unobserved confounder.**
- For each confounder, vary the **model latent dimension.**
- Measure using **√PEHE** (mean squared error in ITE) and **eATE** (error in mean ITE).

4 Results

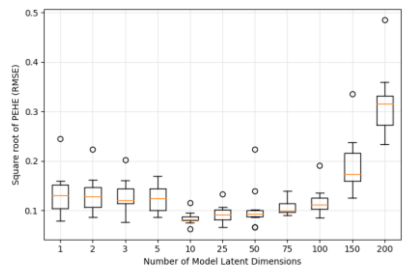

Fig. 3: √PEHE vs model dimensionality for 1 data dimension.

Model is surprisingly robust to having dimensionality **higher** than the data dimensionality but error **increases** eventually, likely due to overfitting (Fig. 3).

Applied to data with **higher** data dimensionality (Fig. 4), performance **decreases** for low model dimensionality (likely due to underfitting). Otherwise, the same trend as in (Fig. 3) can be seen.

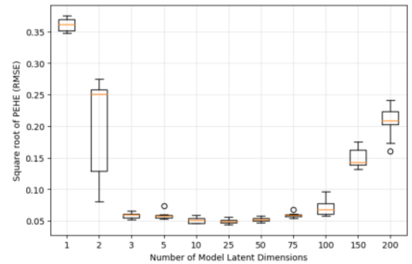

Fig. 4: √PEHE vs model dimensionality for 3 data dimensions.

5 Verification

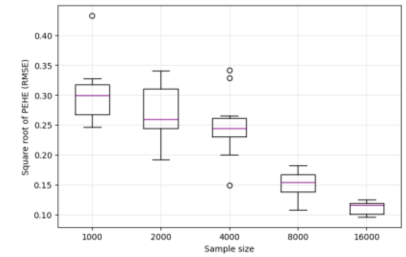

Fig. 5: Overfit model improves with increasing sample size.

When model was suspected to **overfit**, increasing the sample **decreased** error (Fig. 5). When it was suspected to **underfit**, error **did not decrease** (figure not shown), verifying these two cases.

6 Conclusions

- Both under and overspecification **hurt** performance.
- Underspecification is **worse** than overspecification.
- Overspecification **can be fixed** using more data, under **cannot**.
- Conclusions **apply to both** data types tested.

7 Future Work

- Varying the distribution instead of dimensionality.
- Repeat with different datasets.

[1] C. Louizos, U. Shalit, J. M. Mooij, D. Sontag, R. Zemel, and M. Welling, “Causal Effect Inference with Deep Latent-Variable Models,” *Neural Information Processing Systems*, 2017.