Treatment Effect Estimation of the DragonNet under Overlap Violations

Marco van Veen

Supervisor: Stephan Bongers, Responsible Professor: Jesse Krijthe

1. Background

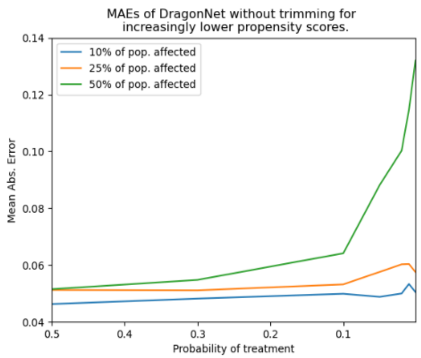
- Learn treatment effects from observational data using ML
- Propensity scores: probability of someone receiving treatment
- Overlap required: all propensity scores must be strictly between 0 and 1
- DragonNet [1] predicts outcomes and propensity scores and uses targeted regularization for desirable estimator properties
- Low overlap samples discarded in original performance tests
- How does DragonNet perform under (near) overlap violations?

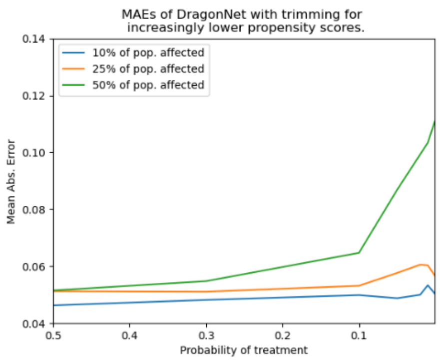
2. Methodology

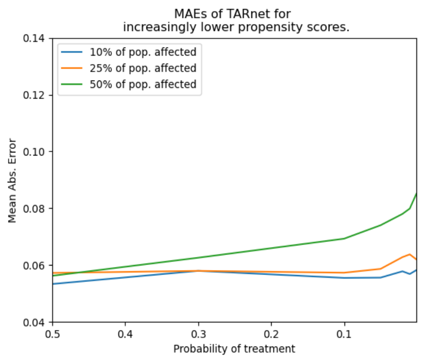
- Obtain errors of DragonNet on synthetic data with varying underlying propensity scores
- Compare results to DragonNet with trimming low propensity data and TARnet model [2]
- Obtain and compare errors when using more realistic semi-synthetic IHDP data by artificially lowering overlap

3. Results

Synthetic Data Results

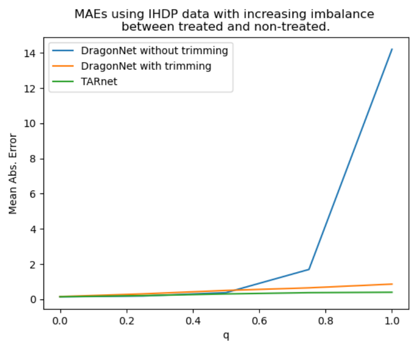




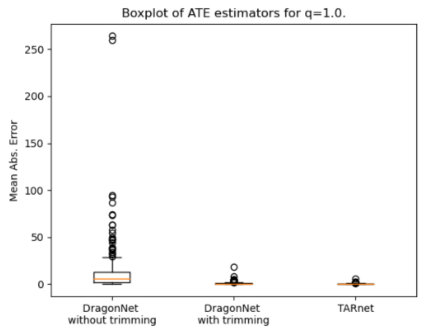


Mean absolute errors (MAE) of estimated average treatment effects (ATE) for increasingly lower propensity scores when 10%, 25%, or 50% of the population is affected by the decreasing propensity scores.

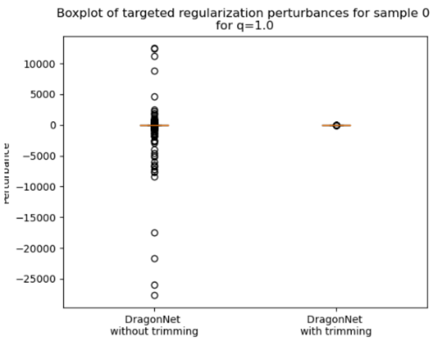
Semi-synthetic Data Results



MAEs of the three models for IHDP samples with decreasing levels of overlap



Variance in the MAEs obtained for each model



Variance in perturbance term in targeted regularization which uses estimated propensity scores

4. Conclusion

- DragonNet performs poorly when large portion of population suffers from low overlap
- Usage of estimated propensity scores in targeted regularization main cause of bad performance under low overlap
- Trimming helps performance, but leads to biased results
- Best to choose other model for effect estimation if substantial overlap violations suspected

5. References

[1] Shi, C., Blei, D., and Veitch, V. (2019). Adapting neural networks for the estimation of treatment effects. *Advances in neural information processing systems*, 32.

[2] Shalit, U., Johansson, F. D., and Sontag, D. (2017). Estimating individual treatment effect: generalization bounds and algorithms. In *International Conference on Machine Learning, pages* 3076–3085. PMLR.