1. Causal Inference from Observational Data

**Goal:** Finding a model that estimates the Individual Treatment Effect (ITE) $ITE(x) = y^1(x) - y^0(x)$ from an observational dataset in the form of $(x_i, t_i, y_i)_{i=1:n}$ with:
- $x$: personal features
- $t$: received treatment chosen from a set of options
- $y$: the observed outcome after receiving the corresponding treatment

**Challenges:**
1. Partial information data: depending on the received treatment $t$, we observe $(outcome \ y^t)$ either $y^1$ or $y^0$, but not both. The other outcome (counterfactual outcome $y^{1-t}$) is unobservable.
2. Selection bias: both outcome $y$ and the treatment $t$ assignment are dependent on (some) context information $x$.
   - $e.g.$, younger (older) patients (part of $x$) are more likely to receive treatment $t$: surgery (medication) because they tend to have a faster (complicated) recovery (outcome $y$).

**Representation Learning:**
Reducing the selection bias by learning a common representation space $\phi(x)$ such that:
- $Pr(\phi(x) | t = 0)$ and $Pr(\phi(x) | t = 1)$ are as close as possible to each other
- $provided$ that $\phi(x)$ retains enough information to accurately predict factual outcomes
- By a learned hypothesis network for each treatment arm (i.e., $\phi_i(x)$) that estimates the corresponding outcomes

**Evaluation Criteria:**
- $PEHE$: $\mathbb{E}(|ITE(x)|) - \mathbb{E}(|ATE(x)|)$
- $PB$: Propensity Score Bias (lower is better)

**4. Experiments**

**Hyperparameter Selection:**
- As counterfactual outcomes are inherently unobservable, it is not possible to use standard internal cross-validation to select hyperparameters (e.g., $\alpha, \lambda$ etc.).
- An estimation of the true effect is needed as a surrogate for the $\alpha$-term
- Shalit et al. (2017) used the observed outcome $y^t$ of the nearest neighbor (1-NN) in the $x$ space (referred to as 1-NN) in the alternative treatment group $y^{1-t} = y^t - t = t - t$
- We also considered outcome prediction by the Bayesian Additive Regression Trees (BART)

**Benchmarks:***
- Infant Health and Development Program (IHDP)
- The observational study is sub-sampled from an RCT by removing a non-random subset of the treated population
- Includes 747 instances with 25 covariates
- Atlantic Causal Inference Conference 2018 (ACIC’18)
- The $x$ matrix is sub-sampled from the Linked Birth and Infant Death Data (LBIDD)
- The $y$s are synthesized by the challenge organizers
- Includes 100,000 instances with 177 features

**5. Results**

We compare performance of four methods:
- 1-NN: One nearest neighbor method for finding the counterfactual outcomes
- BART: Bayesian Additive Regression Trees method (Chipman et al., 2010)
- CFR: CounterFactual Regression method proposed in [Shalit et al., 2017]
- RCFT: Re-weighted CFR (Johansson et al., 2018)
- CFR-ISW: CounterFactual Regression with Importance Sampling Weights (our method)

**Comparisons:**
- PEHE: PEHE of our method is lower than the baseline models
- PB: PB of our method is lower than the baseline models

**Selected References:**
- Chipman, et al., 2010
- [Gretton et al., 2012]
- [Gretton, Scholkopf, et al., 2012]
- [Johansson et al., 2018]
- [Shalit et al., 2017]