ML and Causal Inference: An Overview of Recent Work

(NOT A COMPREHENSIVE LITERATURE REVIEW)

Themes from ML/Causal Inference Literature

Off-the-shelf ML methods…

- Cannot be used for statistical inference
- Are biased due to MSE optimization
- Are biased due to confounding
- Are not optimized for the objective

But a variety of modifications and tricks can help…

- Sample splitting (e.g., with subsampling for forests)
- Changing optimization criterion
- Estimate nuisance parameters using std. ML and use orthogonalization

Average Treatment Effects (ATE) w/ Unconfoundedness

- Targeted ML (van der Laan et al, series)
- Double-LASSO (Belloni, Chernozhukov, Hansen)
- Residual Balancing (Athey, Imbens, Wager (2016), Hirshberg and Wager (2017))
- Double ML (Chernozhukov et al 2017)
- Methods based on averaging CATE:
  - Generalized Random Forests (Athey, Tibshirani, Wager 2016)
  - BART (Chipman and George, 2010), e.g. Hill

Conditional ATE (CATE) w/ Unconf.:

Low-Dimensional Treatment Effect Heterogeneity

“Moving the Goalposts”

- Targeted ML (van der Laan et al)
- LASSO-based methods (Imai and Ratkovic)
- Causal Trees (Athey and Imbens, PNAS 2016)
- X-Learners (Kunzel, Sekhon, Bickel, Yu, 2017)
- Chernozukov and Duflo (2018)

Conditional ATE w/ Unconf.:

Non-parametric Case

- Causal Forests (Wager and Athey, 2015)
- Generalized Random Forests (Athey, Tibshirani, Wager 2016)
- Nie and Wager (2017)

Optimal (Personalized) Policy Estimation

Offline

- From ML Literature: Strehl et al. (2010); Dudik et al. (2011); Li et al. (2012); Dudik et al. (2014); Li et al. (2014); Swaminathan and Joachims (2015); Jiang and Li (2016); Thomas and Brunskill (2016); Kallus (2017).
- ML + Semiparametric efficiency: Efficient Policy Estimation (Athey and Wager, 2017)

Policy Estimation Online:

Contextual Bandits

- Very large ML literature; see e.g. Li et al. (2010), Goldenshluger and Zeevi (2013), Li et al. (2017), Bastani and Bayati (2015), and Feraud et al. (2016).
- Estimation issues (Dimakopoulou, Athey, Imbens 2017)

Supplementary Analyses	• Robustness of parameter estimates (Athey, Imbens 2015) • Confoundedness (Athey, Imbens, Pham, Wager, 2017) • See also Athey-Imbens 2017 survey
Instrumental Variables (IV) Estimates (Average or Low-D CATE)	• Targeted ML • LASSO-Based methods (Belloni, Chernozhukov, Hansen et al, series)
IV: Heterogeneous Effects (CLATE)	• ML/GMM Trees (Zeiles et al 2008; Athey, Tibshirani, Wager 2016; Asher et al 2016 • Generalized Random Forests (Athey, Tibshirani, Wager 2016) • Deep IV [neural nets] (Lewis, Leyton-Brown and Taddy, 2016)
Heterogeneous Parameter Estimation in GMM/ML Models (Non-parametric heterogeneity)	• Generalized Random Forests (Athey, Tibshirani, Wager 2016)

Regression Discontinuity

- Local Linear Forests (in progress; Athey, Friedberg, Wager 2017)

Panel Data

(Diffs-in-Diffs, Synthetic Controls, etc.)

- Synthetic Controls with Regularized Regression for weights (Doudchenko and Imbens 2016)
- Matrix Completion w/ Nuclear Norm (Athey et al 2017)

Combining Observational and Experimental Data

- Surrogates (Athey, Chetty, Imbens, Kang, 2017)
- Peysakhovich and Lada (2016)
- Randall Lewis (in progress)

Large-Scale Structural Models (Consumer Demand)

- Bayesian matrix factorization for independent categories (Athey, Blei, Donnelly, Ruiz, 2017; Athey, Blei, Donnelly, Ruiz, Schmidt, 2018)
- Matrix factorization for multi-step shopping decisions (Wan et al, 2017)
- Estimating complements/substitutes with many items in Bayesian model (Ruiz, Athey, Blei 2017)