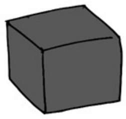
Born-again trees for predicting treatment outcomes

First fit a forest, then fit a tree

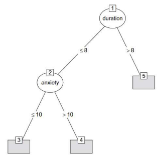
TSIL 2021
Marjolein Fokkema
Leiden University

Slides and R scripts:
https://github.com/marjoleinF/TSIL2021

1



Neural nets, tree ensembles, support vector machines:
- High accuracy
- Black box



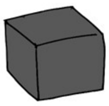
Single trees:
- Interpretable
- Lower accuracy

2

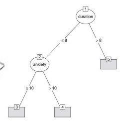
Breiman & Shang (1996): Born again trees

Fit tree on predictions of a black box, instead of original outcome *y*

	x1	x2	y
1	5	2	8
2	9	0	3
3	2	1	6



	x1	x2	
1	5	2	
2	9	0	
3	2	1	



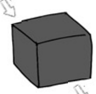
More stable and accurate tree

3

Predicting treatment outcomes

1. Fit black box model on original training data (x, t, y)
2. For each training observation, generate:
   $\hat{y}_{t=A}$ (pred. outcome given treatment A)
   $\hat{y}_{t=B}$ (" given treatment B)
3. Fit a linear (mixed) model tree on augmented dataset

	x1	x2	t	y
1	5	2	A	8
2	9	0	A	3
3	2	1	B	6



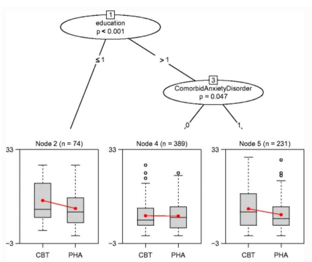
	x1	x2		
1	5	2		
2	9	0		
3	2	1		
1	5	2		
2	9	0		
3	2	1		

4

Linear model tree (Zeileis et al., 2008)

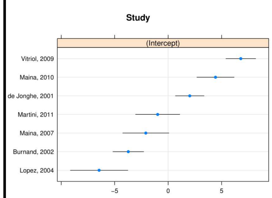
An effect of interest (e.g., treatment effect) may differ between observations in a dataset

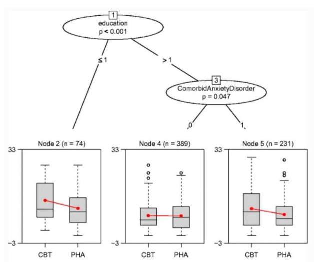
Using additional covariates, we may find subgroups that show differences in parameter estimates (e.g., treatment effect)



5

Linear mixed model tree (Fokkema et al., 2018)





6

Experiment

Can the born-again approach improve treatment outcome predictions of LM(M) trees?

Black box: Bayesian Additive Regression Trees (BART)

Two IPDMA datasets:

1. STPP+PHA versus PHA (Driessen et al., 2018).
   N = 367 (completers only); 5 potential moderators.
2. CBT versus PHA (Cuijpers et al., 2014):
   N = 694 (completers only); 4 potential moderators.

7

Results (1)

All results based on 10 repeats of 10-fold CV.

Prediction error (MSE):

IPDMA Dataset	Var(y)	LM tree	BART	BART-based LM tree
STPP vs PHA	89.78	60.75 (SD=14.21)	55.95 (SD=12.77)	56.74 (SD=12.53)
CBT vs PHA	39.25	38.13 (SD=6.17)	36.98 (SD=6.17)	37.90 (SD=5.93)

Number of splits in tree:

IPDMA dataset	LM tree	BART-based LM tree
STPP vs PHA	2.91 (SD=0.60)	9.94 (SD=1.26)
CBT vs PHA	2.69 (SD=0.86)	1.83 (SD=0.38)

8

Results (2)

Mixed-effects models to account for patients nested in studies:

Prediction error (MSE):

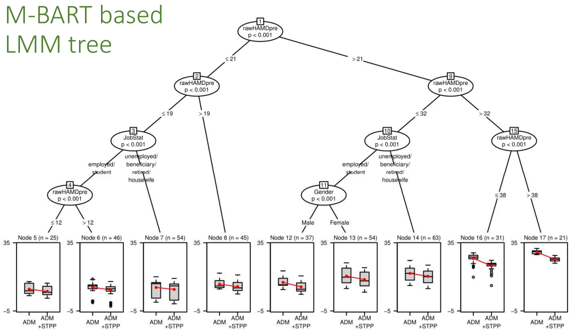
IPDMA dataset	Var(y)	LMM tree	M-BART	M-BART-based LMM tree
STPP vs PHA	89.78	46.12 (SD=12.16)	44.94 (SD=11.67)	43.54 (SD=11.35)
CBT vs PHA	39.25	36.38 (SD=5.86)	35.85 (SD=5.67)	35.91 (SD=5.75)

Number of splits in tree:

IPDMA dataset	LMM tree	M-BART-based LMM tree
STPP vs PHA	2.58 (SD=0.57)	9.53 (SD=1.23)
CBT vs PHA	2.58 (SD=0.73)	1.94 (SD=0.28)

9

M-BART based LMM tree



10

Discussion

- First fit a forest, then fit a tree: Born-again approach improves treatment outcome predictions, performance close to that of BART
- Can prune trees based on criterion of clinically relevant change
  - Born-again approach similar to:
    - PAI (e.g., DeRubeis et al. 2014)
    - Virtual Twins (e.g., Foster et al., 2011)
    - Causal models (e.g., Kunzel et al., 2019; Hahn et al., 2020)

11

Breiman, L. (1996). *Bias, variance, and arcing classifiers*. Tech. Rep. 460, Statistics Department, University of California, Berkeley, CA, USA.

Breiman, L., & Shang, N. (1996). Born again trees. *University of California, Berkeley, CA, Technical Report*.

Cuijpers, P., Weitz, E., Twisk, J., Kuehner, C., Cristea, I., David, D., & Hollon, S. D. (2014). Gender as predictor and moderator of outcome in cognitive behavior therapy and pharmacotherapy for adult depression: An individual-patients data meta-analysis. *Depression and Anxiety*, *31*(11), 941–951.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PloS one*, *9*(1), e83875.

Driessen, E., Abbass, A. A., Barber, J. P., Gibbons, M. B. C., Dekker, J. J., Fokkema, M., ... & Cuijpers, P. (2018). Which patients benefit specifically from short-term psychodynamic psychotherapy (STPP) for depression? Study protocol of a systematic review and meta-analysis of individual participant data. *BMJ open*, *8*(2), e018900.

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior research methods*, *50*(5), 2016-2034.

Foster, J. C., Taylor, J. M., & Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in medicine*, *30*(24), 2867-2880.

Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Analysis*, *15*(3), 965-1056.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *PNAS*, *116*(10), 4156-4165.

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492-514.

Slides and R scripts:
https://github.com/marjoleinF/TSIL2021

12