Trees and rules

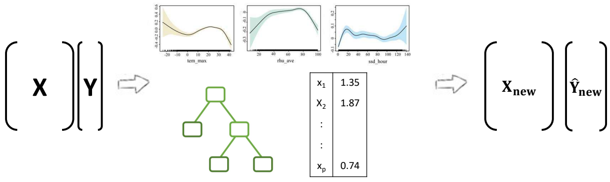
Bridging interpretable and explainable machine learning

Marjolein Fokkema

m.fokkema@fsw.leidenuniv.nl

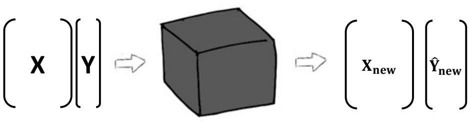
1

Interpretable ML



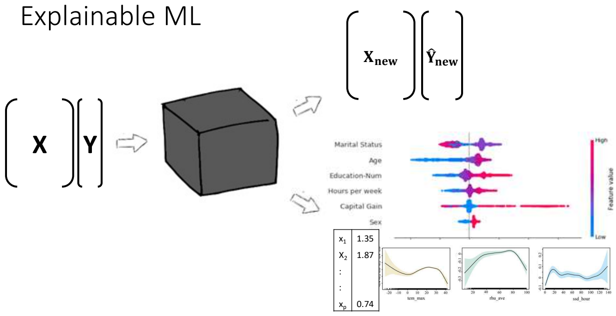
2

Explainable ML



3

Explainable ML



4

Predicting university major

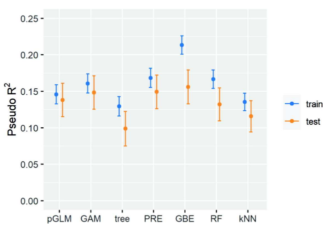
Fokkema et al. (2022), Eur J Psych Assessment

- N = 55,593 (75% training; 25% test)
- Took psychology as a major at university: Yes (19.4%) vs. No
- Predictors: 48 items on a vocational preference scales

5

Predicting university major

Fokkema et al. (2022), Eur J Psych Assessment



(p)GLM = (penalized) logistic regression
GAM = generalized additive model with smoothing splines
PRE = prediction rule ensemble
GBE = gradient boosted tree ensemble
RF = random forest
*k*NN = *k* nearest neighbors

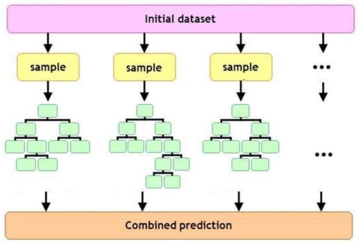
6

Prediction rule ensembles

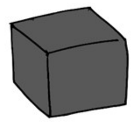
-> Can have simple, interpretable ML model with near-optimal accuracy?

7

Decision tree ensembles



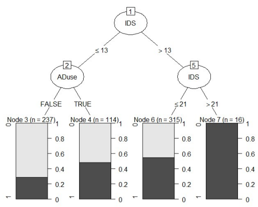
- Bagging
- Random forests
- Gradient boosting
- .....



8

From trees to rules

$r_2(\mathbf{x}) = I(IDS \leq 13)$
$r_3(\mathbf{x}) = I(IDS \leq 13) \cdot I(ADuse = FALSE)$
$r_4(\mathbf{x}) = I(IDS \leq 13) \cdot I(ADuse = TRUE)$
$\cancel{r_5(\mathbf{x}) = I(IDS > 13)}$
$r_6(\mathbf{x}) = I(IDS > 13) \cdot I(IDS \leq 21)$
$r_7(\mathbf{x}) = I(IDS > 13) \cdot I(IDS > 21)$



9

From trees to rules

$r_2(\mathbf{x}) = I(IDS \leq 13)$
$r_3(\mathbf{x}) = I(IDS \leq 13) \cdot I(ADuse = FALSE)$
$r_4(\mathbf{x}) = I(IDS \leq 13) \cdot I(ADuse = TRUE)$
$r_6(\mathbf{x}) = I(IDS > 13) \cdot I(IDS \leq 21)$
$r_7(\mathbf{x}) = I(IDS > 13) \cdot I(IDS > 21)$
...
$l_1(\mathbf{x}) = IDS$
$l_2(\mathbf{x}) = ADuse$
...

$$\boxed{F(\mathbf{x}) = \hat{\alpha}_0 + \sum_{m=1}^{M} \alpha_m f_m(\mathbf{x})}$$

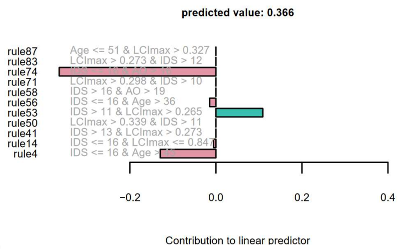
IDS	ADuse	...	r_2	r_3	r_4	r_6	r_7	...
5	FALSE	...	1	1	0	0	0	...
15	FALSE	...	0	0	0	1	0	...
18	TRUE	...	0	0	0	1	0	...
25	TRUE	...	0	0	0	0	1	...
...	...	...	...	...	...	...	...	...

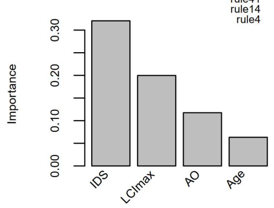
10

PRE for predicting chronic depression
Fokkema & Strobl (2020) Psychological methods

Description	Coeffient
1	-0.221
IDS > 10 & LCImax > 0.2632	0.224
IDS > 13 & LCImax > 0.3621	0.213
IDS <= 16 & AO > 17	-0.175
IDS > 10 & LCImax > 0.3276	0.140
LCImax > 0.26 & IDS > 9	0.122
IDS <= 16 & GAD %in% c("Negative")	-0.080
IDS > 10 & Age > 22	0.020
IDS <= 17 & AO > 13	-0.015
IDS > 14 & pedigree %in% c("Yes")	0.002

11





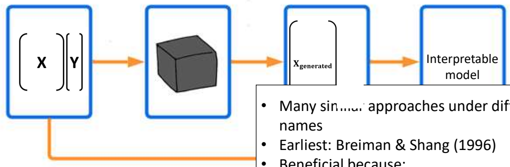
12

-> PRE is interpretable <u>and</u> explains black box

Problems:
- Complexity (too many rules retained by lasso)
- Lower accuracy than black box
- Instability (trees, lasso)

13

Model-based data generation



- Many sin... approaches under different names
- Earliest: Breiman & Shang (1996)
- Beneficial because:
  - Less noise in $\hat{y}$?
  - Very large N_{gen}?

14

Experiments with PRE

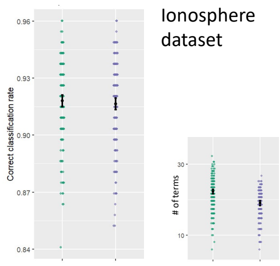
Black box: Gradient boosting (not tuned)

Data generation:
- $N_{gen} = 10xN_{orig}$
- Features 100% permuted

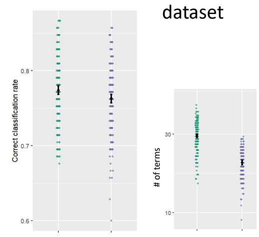
Interpretable model: Prediction Rule Ensemble

15

PRE



Ionosphere dataset



Sonar dataset

16

Experiment with trees

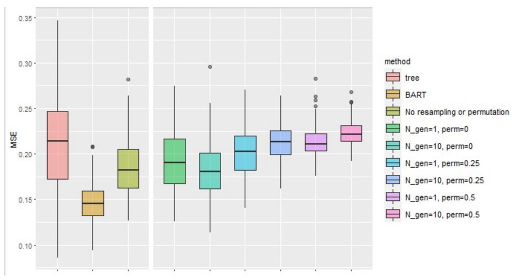
Black box: Bayesian additive regression trees

Data generation:
- N_{gen} = 1, 5 and 10 x N_{orig}
- Features 0, 25, 50 and 100% permuted

Interpretable model: Single tree

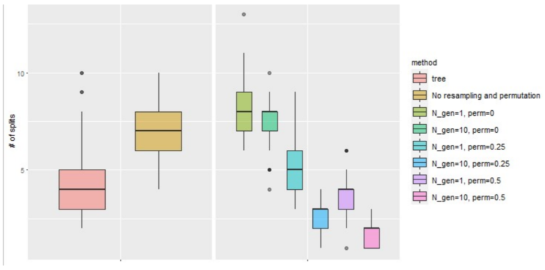
17

Trees (Sonar dataset)



18

Trees (Sonar dataset)



19

Discussion

- Can use black box to improve accuracy, sparsity/stability of interpretable method
- Interpretable ML meets global explainability

\+ Accuracy of interpretable model and black box quantified

- Data generation approach is critical for succes

20

References

Breiman, L., & Shang, N. (1996). Born again trees. *University of California, Berkeley, Berkeley, CA, Technical Report*, *1*(2), 4.

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, *4*(1), 266-298.

Fokkema, M., Iliescu, D., Greiff, S., & Ziegler, M. (2022). Machine learning and prediction in psychological assessment: Some promises and pittfalls. *European Journal of Psychological Assessment 38*(3), 165-175. https://doi.org/10.1027/1015-5759/a000714

Fokkema, M. (2020). Fitting prediction rule ensembles with R package pre. *Journal of Statistical Software, 92*(12), *1-30.* http://doi.org/10.18637/jss.v092.i12

Fokkema, M. & Strobl, C. (2020). Fitting prediction rule ensembles to psychological research data: An introduction and tutorial. *Psychological Methods 25*(5), 636–652. https://arxiv.org/abs/1907.05302

Fokkema, M. & Christoffersen, B. **pre**: Prediction Rule Ensembles. https://github.com/marjoleinF/pre https://cran.r-project.org/package=pre

Markovitch, B., & Fokkema, M. (2021). Improved prediction rule ensembling through model-based data generation. *arXiv preprint arXiv:2109.13672*. https://arxiv.org/abs/2109.13672

21