Decision tree methods for psychological assessment

An introduction to classification and regression trees, model-based trees, and trees for longitudinal and multilevel data

ECPA15, Brussels, Belgium, 2019

Marjolein Fokkema, Leiden University

Today's workshop

Goal: Short introduction to decision-tree methods

How:
1. Plenary introduction + examples
2. Interactive / individual:
   - Replicate examples in R and/or
   - Apply methods to own data

All materials (slides, manual, datasets) available online, go to **https://tinyurl.com/ECPA15trees**

Disclaimer: Most datasets artificial, so that you can replicate all examples. Thus, results not valid in real world!

Contents

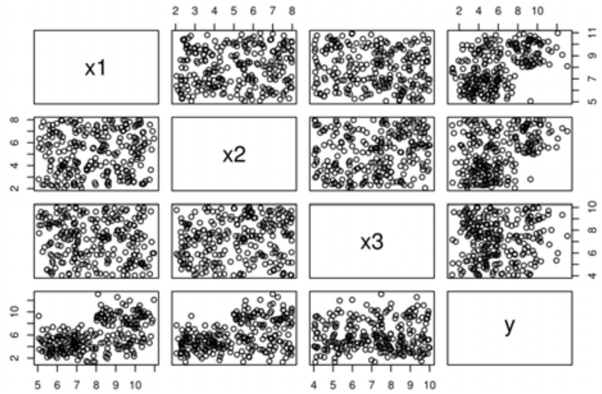
Short history of RPMs

Early methods:
- Automated interaction detection (Morgan & Sonquist, 1963)
- Classification and regression trees (Breiman et al., 1984)
- ID3 (Quinlan, 1986)
- C4.5 (Quinlan, 1993)

Today we focus on unbiased recursive partitioning:
- Conditional inference tree (ctree; Hothorn, Hornik & Zeileis, 2006)
- Model-based recursive partitioning (MOB; Zeileis, Hothorn & Hornik, 2008)

Toy dataset

N = 250

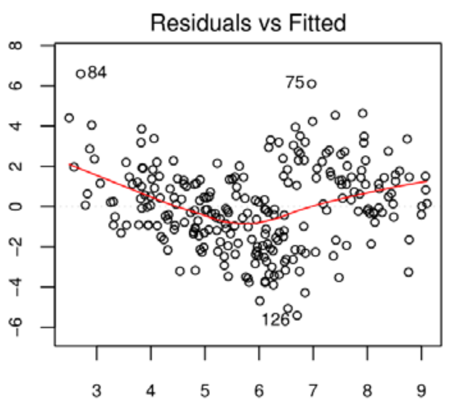


OLS regression model on toy data

```
lin_mod <- lm(y ~ x1 + x2 + x3, data = toy_data)
```

```
summary(lin_mod)
```

```
##
##
##                                       (>|t|)
##                                       .00231 **
##                                        2e-16 ***
##                                       31e-10 ***
##                                       .59521
##
##                                       ' 0.05 '.' 0.1 ' ' 1
##
##                                       ees of freedom
##                                       uared:   0.3574
##                                       ue: < 2.2e-16
```

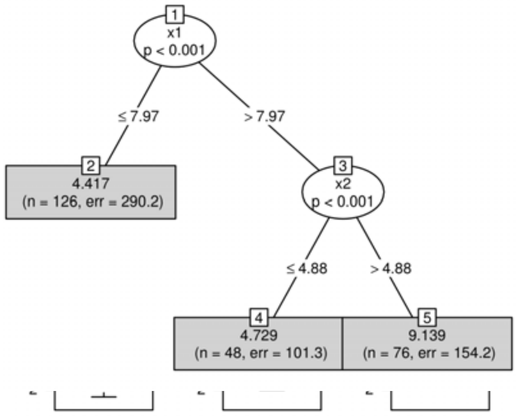


$\hat{y} = $ -… $+ 0.041\ x_3$

Regression tree on toy data

```
library("partykit")
```

```
tree <- ctree(y ~ x1 + x2 + x3, data = toy_data)
```



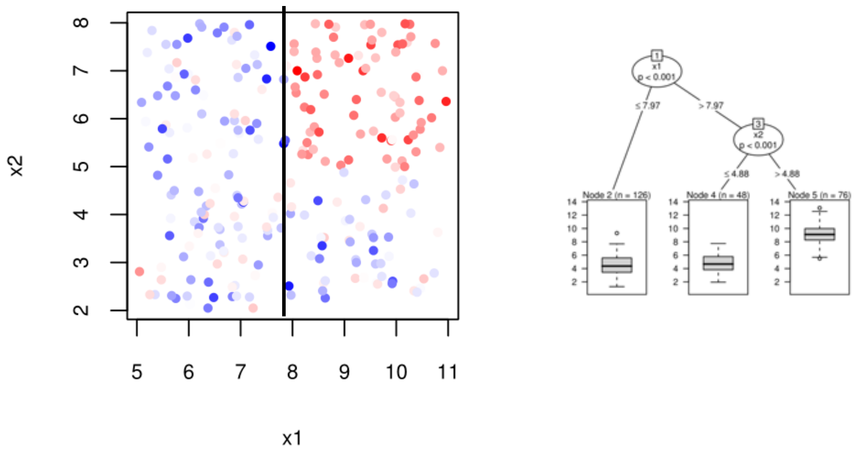
Split selection

Conditional inference tree algorithm (Hothorn, Hornik & Zeileis, 2006):

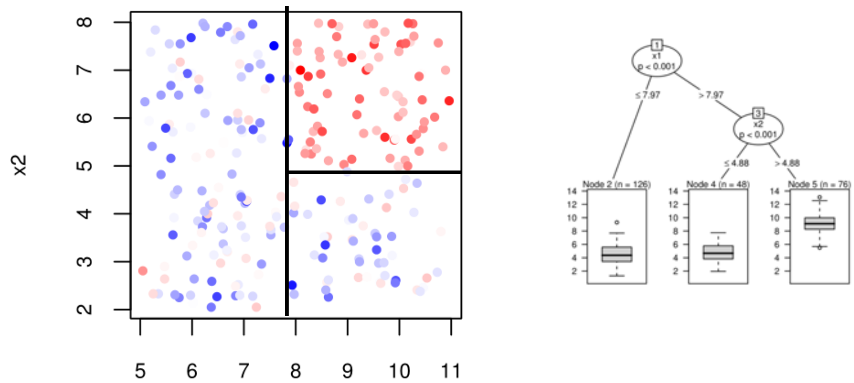
1. In current node, statistically test the association between partitioning variables and response. If at least one partitioning variable has *p* value ≤ α , variable with the lowest *p* value is selected for splitting.
2. Select the splitting value that minimizes loss function in the two resulting terminal nodes.
3. Repeat steps 1 and 2 in the two resulting nodes.

-> Separate selection of splitting variable (1) and value (2)
eliminates variable selection bias

-> Step 2 provides natural stopping criterion




```
##                      x1           x2         x3
## statistic 6.120936e+01 3.644008e+01 0.3462214
## p.value   1.539507e-14 4.722965e-09 0.9126252
```

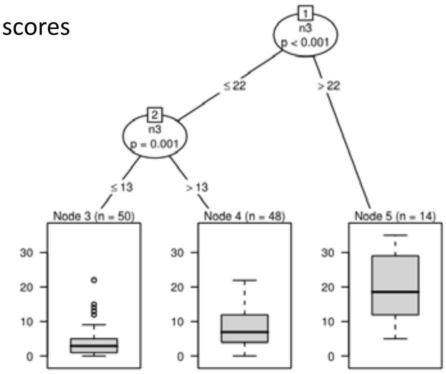



```
## $`5`
##                  x1         x2        x3
## statistic 0.4845901 0.07074638 1.6192195
## p.value   0.8644809 0.99077245 0.4941207
```

x3
45
28

Regression tree example

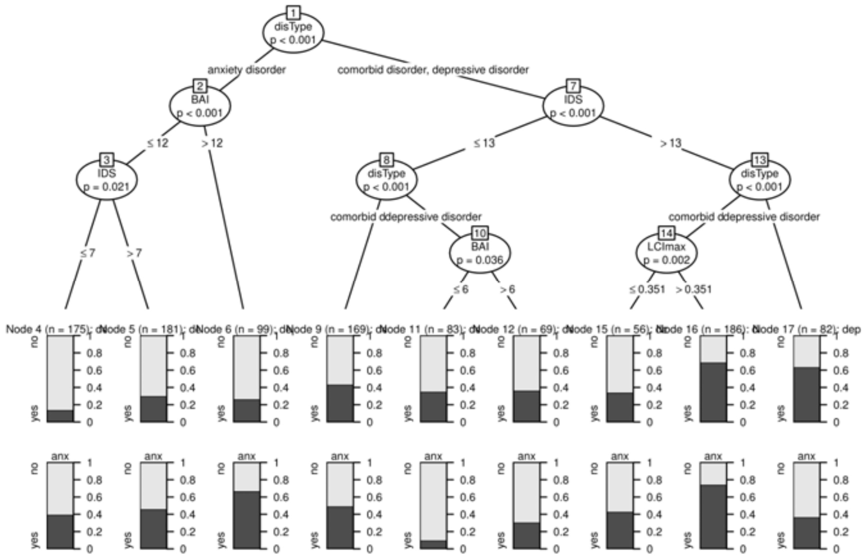
Study by Carrillo et al. (2001):
- Response: Beck Depression Inventory score
- 25 possible predictors:
  - NEO-PI-R facet and subscale scores
  - age, gender
- N = 112 respondents



Classification tree example

Dataset modeled after study by Penninx et al. (2011) on chronicity of anxiety and depression
- Sample: Respondents with current anxiety and/or depressive disorder (original N = 1,209; current N = 1,100)
- Response: Presence of anxiety and/or depressive disorder diagnoses, 2 years after baseline
- Possible predictor variables:
  - disType: baseline psychiatric status
  - gender, age, age of disorder onset, years of completed education.
  - IDS (depressive symptoms), BAI and FQ (anxiety symptoms)
  - LCImax: proportion of time in which symptoms of anxiety and depression were present in the four years prior to baseline
  - type of depressive disorder at baseline
  - type of anxiety disorder(s) at baseline
  - ….

Classification tree example



Model-based recursive partitioning (MOB; Zeileis, Hothorn & Hornik, 2008)

Rationale: Global parametric model (e.g., LM / linear regression, GLM / logistic regression) may not fit data well.

Perhaps can find subgroups, defined by partitioning variables, with better-fitting local models.

MOB algorithm:

1. Fit parametric model to observations in current node. Perform a *parameter stability test* for each of the partitioning variables. If at least one of the partitioning variable has *p* value ≤ α , select variable with lowest *p* value for splitting.
2. Select splitting value that yields optimal fit when the parametric model is fitted in each of the resulting two nodes.
3. Repeat steps 1 and 2 in the two resulting nodes.

MOB example: Treatment subgroups

Data modeled after SMART data from the Improving Access to Psychological Therapies (IAPT) project (Lucock et al., 2017).

IAPT: Patients receiving mental-health services in the Northern UK. Non-random assignment to:

- LI: low intensity treatment, e.g. guided self-help, computerized cognitive behavior therapy
- HI: high intensity treatment, e.g. face-to-face psychological therapies

Aim of SMART: Identify patients who will benefit most from HI vs. LI treatment

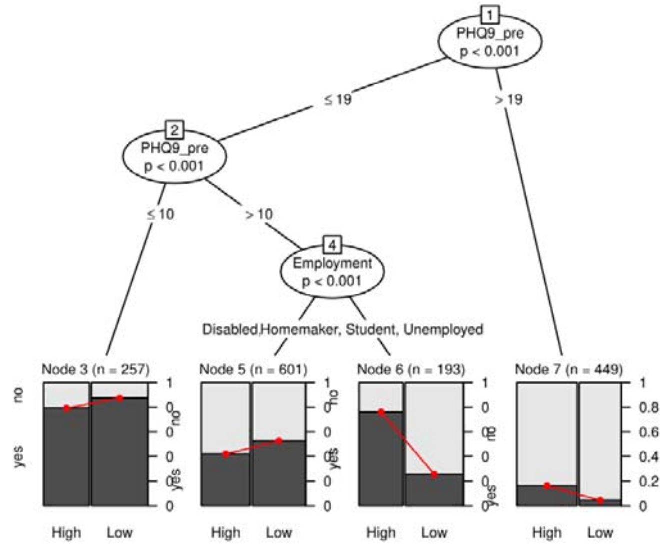
-> Fit tree to detect predictors and moderators of treatment outcome

Treatment subgroups

Dataset contains 1,500 observations and 13 variables

- Response: recovered (yes, no)
- Predictor: treatment type (HI vs LI)
- 10 possible partitioning variables:
  - PHQ9_pre (baseline depression measure)
  - GAD7_pre (baseline symptoms of generalized anxiety disorder)
  - WSAS_pre (baseline work and social functioning)
  - Age, Gender, Ethnicity
  - Diagnosis
  - Employment status, Disability
  - Medication use
- indicator for treatment center (will use later in the mixed-effects tree)

Treatment subgroups




```
## Model formula:
## recovered ~ Treatment | Age + PHQ9_pre + GAD7_pre + WSAS_pre +
##     Gender + Ethnicity + Diagnosis + Employment + Disability +
##     Medication
##
## Fitted party:
## [1] root
## |   [2] PHQ9_pre <= 19
## |   |   [3] PHQ9_pre <= 10: n = 257
## |   |        (Intercept) TreatmentLow
## |   |          1.3328057    0.5971041
## |   |   [4] PHQ9_pre > 10
## |   |   |   [5] Employment in Disabled, Employed, Retired: n = 601
## |   |   |        (Intercept) TreatmentLow
## |   |   |         -0.3305021    0.4325012
## |   |   |   [6] Employment in Homemaker, Student, Unemployed: n = 193
## |   |   |        (Intercept) TreatmentLow
## |   |   |           1.147402    -2.221917
## |   [7] PHQ9_pre > 19: n = 449
## |        (Intercept) TreatmentLow
## |          -1.641477    -1.449565
##
## Number of inner nodes:    3
## Number of terminal nodes: 4
## Number of parameters per node: 2
```

Local logistic regression models (GLM)

Dependent observations

Observations are known to be dependent in *nested* data, e.g.,

- Multilevel data
- Longitudinal data

Account for dependence by:

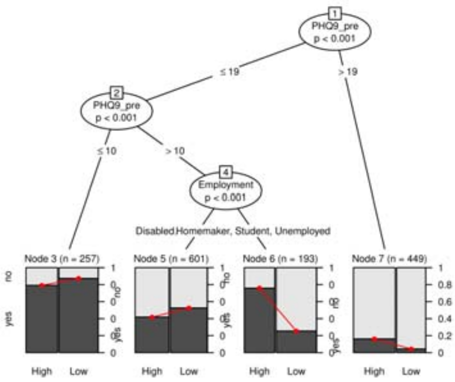
- Estimating random effects and/or
- Adjusting variable selection tests

Mixed-effects trees

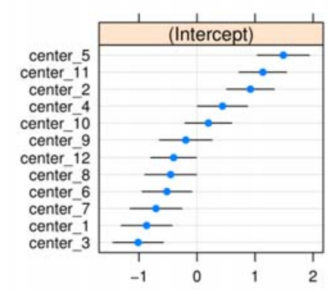
Generalized linear mixed effect regression tree (glmertree; Fokkema et al. 2018):

- Estimate local 'fixed-effects' models
  - That is: find subgroups and estimate local parameters
- Globally estimate random effects
  - That is: using all observations

Mixed-effects tree example: Treatment subgroups with multilevel structure



+ random intercept w.r.t. treatment center:



Mixed-effects tree example: Subgroup detection in growth trajectories

Data: Subsample of N = 400 children from Early Childhood Longitudinal Study-Kindergarten (ECLS-K)

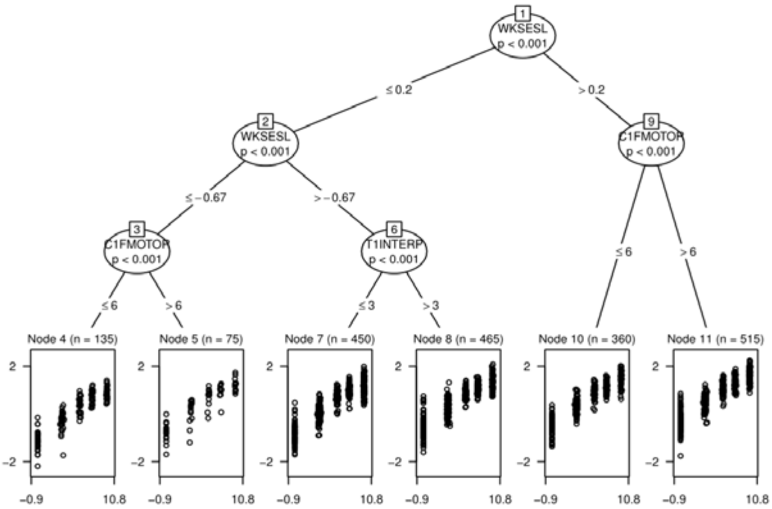
Response variable: theta (IRT ability) scores of reading skills

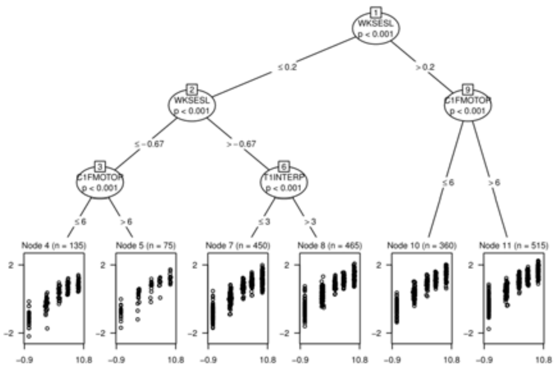
Predictor variable: time (square root of months since baseline)

Potential partitioning variables (baseline):
- GENDER , RACE and AGE
- WKSESL (measure of socio-economic status)
- C1GMOTOR and C1FMOTOR (measures of gross and fine motor skills)
- T1INTERN , T1EXTERN , T1INTERP and T1CONTRO (measures of internalizing, externalizing, interpersonal problems and self control)

Partitioning variables measured at level II, not individual observation level (level I) -> need to account in tests for variable selection

Mixed-effects tree example: Subgroup detection in growth trajectories





```
coef(ECLSK_t)
```

```
##    (Intercept)      time
## 4   -1.0273482 0.2125477
## 5   -0.6956136 0.2113902
## 7   -0.7277500 0.2168037
## 8   -0.4681510 0.2000110
## 10  -0.4711683 0.2122275
## 11  -0.2576150 0.2011876
```

```
VarCorr(ECLSK_t)
```

```
##  Groups   Name        Std.Dev.
##  CHILDID  (Intercept) 0.23342
##  Residual             0.23515
```

```
as.numeric(VarCorr(ECLSK_t)) / var(ECLSK$score)
```

```
## [1] 0.08497081
```

Concluding remarks

- Go and create yourself some trees!
- Manual for replicating examples is online ("Fitting_trees_in_R.pdf")
- I'm happy to answer questions now, during conference, or via m.fokkema@fsw.leidenuniv.nl

References

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees.* New York: Wadsworth.

Carrillo, J. M., Rojo, N., Sanchez-Bernardos, M. L., Avia, M. D. (2001). Openness to experience and depression. *European Journal of Psychological Assessment, 17*(2), 130-136.

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.

Hothorn, T., Hornik, K., Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics, 15*(3), 651-674.

Kass, G.V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics, 29*(2), 119–127.

Lucock, M., Barkham, M., Donohoe, G., Kellett, S., McMillan, D., Mullaney, S., ... & Delgadillo, J. (2017). The role of Practice Research Networks (PRN) in the development and implementation of evidence: The Northern improving access to psychological therapies PRN case study. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(6), 919-931.

Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association, 58*, 415–434.

Penninx, B. W. J. H., Nolen, W. A., Lamers, F., Zitman, F. G., Smit, J. H., Spinhoven, P., ..., Beekman, A. T. F. (2011). Two-year course of depressive and anxiety disorders: Results from the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Affective Disorders, 133*(1), 76-85.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning.* San Francisco: Kaufmann Publishers.

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