

Seminar Datenanalyse

Dr. Constantin Weiser; Winter 2019/2020

Zeitplan

Datum	Raum	
24.10.2019 14.15-15.45 Uhr	HS VI	Einführung & Themenvergabe
13.11.2019 12:00Uhr		Abgabe der Exposés (per Mail)
14.11.2019 14.15-17.45 Uhr	HS VI	Präsentation der Exposés
18.11.2019		Start Bearbeitungszeit der Bachelorarbeiten
08.01.2020		Abgabe der Bachelorarbeiten beim Studienbüro

Klassifikation mit CART

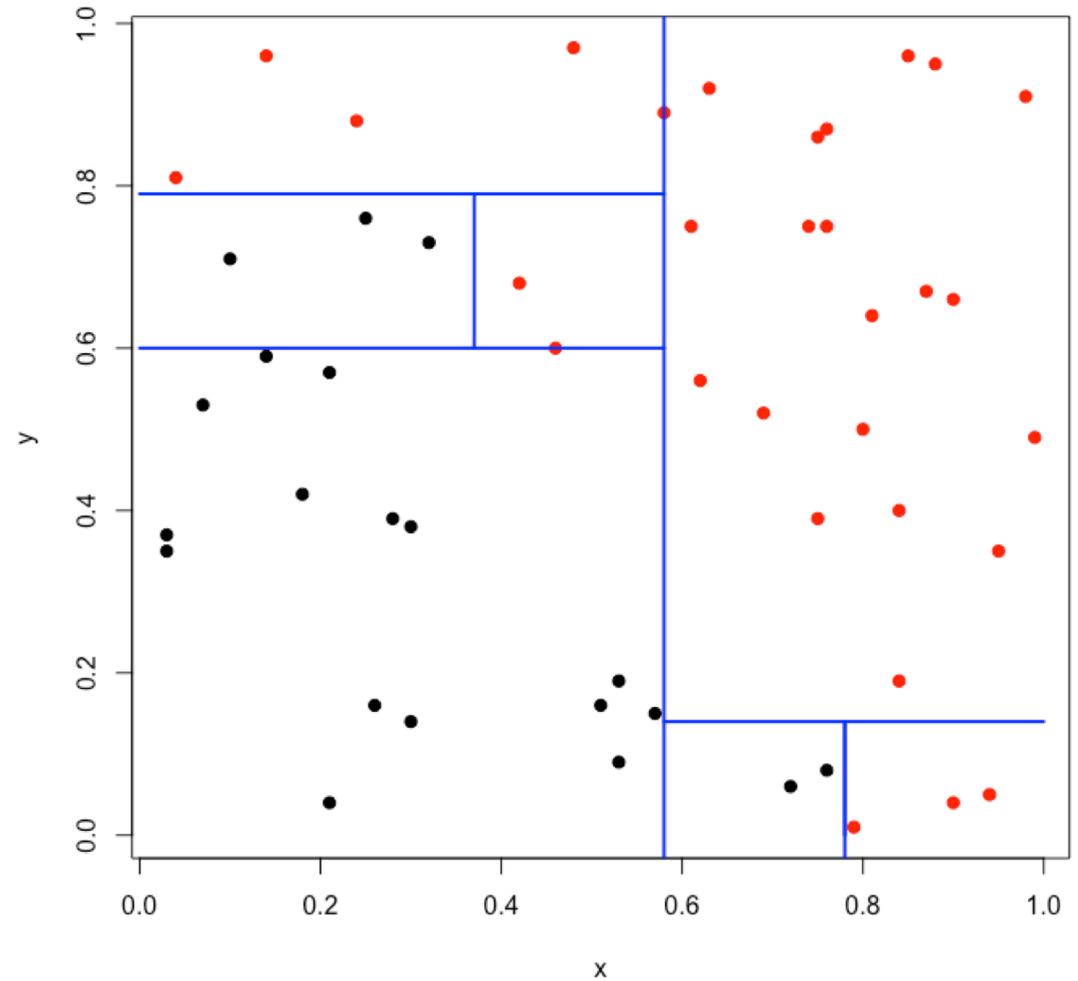
Konzeptionelle Einführung

Beispiel: Vorhersage von Klausurergebnissen

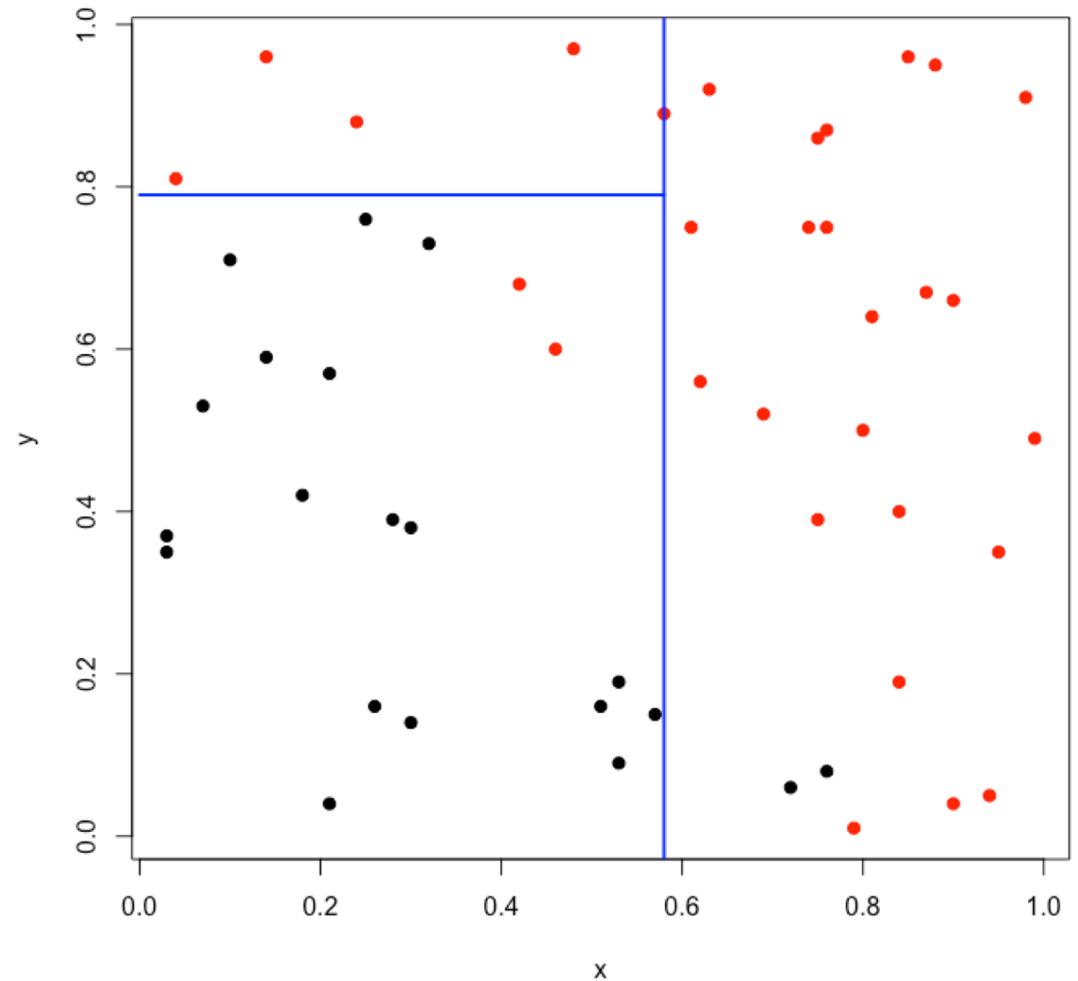
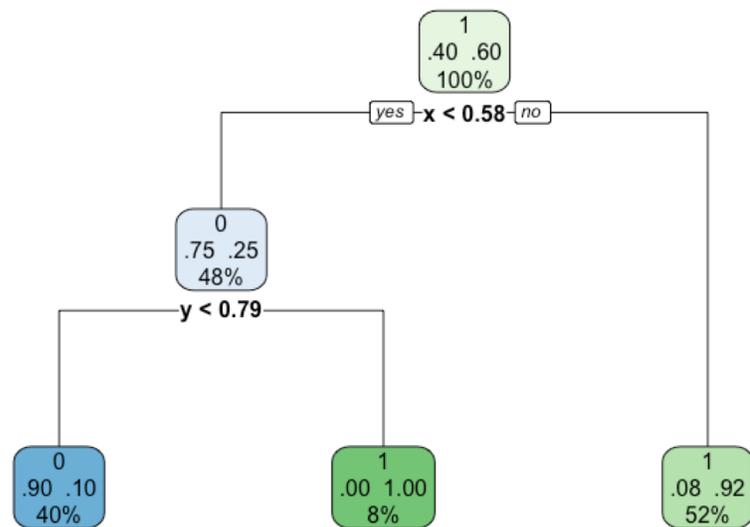
abinote	bundesl	mathelk	manote	alter	geschl	guteNote
2.5	9	1	1	1	1	FALSE
2.7	6	0	1	0	0	TRUE
2.6	10	0	1	1	0	FALSE
3	6	0	4	0	0	FALSE
1.4	6	0	0	1	1	FALSE
2.8	6	0	2	1	0	FALSE
3.3	6	0	3	2	0	FALSE
1.2	10	0	NA	4	0	FALSE
1.3	6	0	0	2	0	TRUE
1.6	0	1	0	1	0	TRUE
1.9	6	0	2	1	0	FALSE
2.4	11	0	1	1	1	FALSE
1.4	6	1	1	2	0	TRUE
2.4	6	0	1	1	0	FALSE
2.4	10	0	2	1	0	TRUE
2.2	6	0	1	0	0	FALSE
3.1	NA	1	2	NA	0	FALSE
3.1	NA	1	2	NA	1	FALSE

Illustratives Beispiel

Illustratives Beispiel

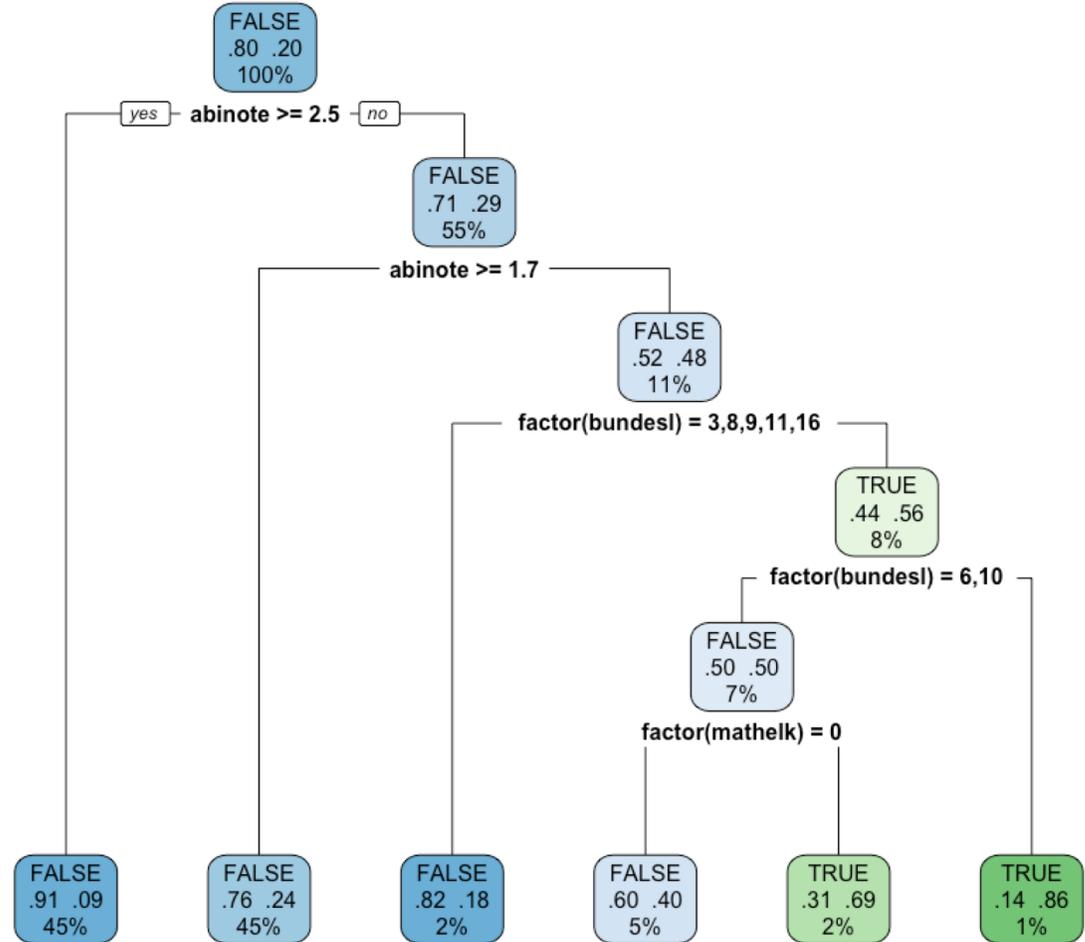


Illustratives Beispiel



Beispiel: Vorhersage von Klausurergebnissen

abinote	bundesl	mathelk	manote	alter	geschl	guteNote
2.5	9	1	1	1	1	FALSE
2.7	6	0	1	0	0	TRUE
2.6	10	0	1	1	0	FALSE
3	6	0	4	0	0	FALSE
1.4	6	0	0	1	1	FALSE
2.8	6	0	2	1	0	FALSE
3.3	6	0	3	2	0	FALSE
1.2	10	0	NA	4	0	FALSE
1.3	6	0	0	2	0	TRUE
1.6	0	1	0	1	0	TRUE
1.9	6	0	2	1	0	FALSE
2.4	11	0	1	1	1	FALSE
1.4	6	1	1	2	0	TRUE
2.4	6	0	1	1	0	FALSE
2.4	10	0	2	1	0	TRUE
2.2	6	0	1	0	0	FALSE
3.1	NA	1	2	NA	0	FALSE
3.1	NA	1	2	NA	1	FALSE



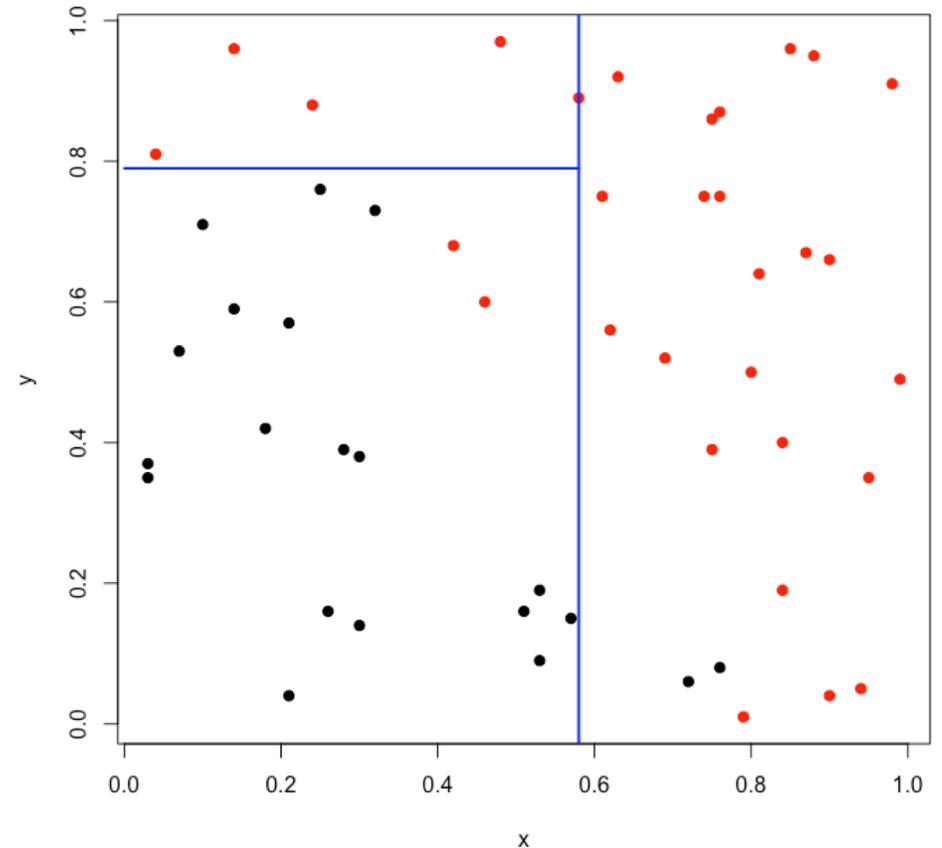
Anwendungsbeispiele (in den WiWi)

- ▶ Verhaltensvorhersage (Kreditausfall, Kaufentscheidungen, ...)
- ▶ Auswahl von Prediktoren und Theoriebildung
- ▶ ...

- ▶ Allgemein
 - ▶ Strukturerkennende Methode
 - ▶ Theoriebildung

CART - Algorithmus

- ▶ L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone (1984): *CART: Classification and Regression Trees*.
- ▶ Rekursiver Algorithmus
- ▶ Viele Implementierungen und Varianten
 - ▶ STATA: Wim van Putten, 2006. "[CART: Stata module to perform Classification And Regression Tree analysis](#)"
 - ▶ R: <https://CRAN.R-project.org/package=rpart>



Stärken und Schwächen von CART

- ▶ Black-Box Methode zur Strukturerkennung
- ▶ Automatische Variablenselection

- ▶ Keine Möglichkeit Hypothesen zu testen (im statistischen Sinne)
- ▶ Nicht geeignet als “Erklärungsmodell”
- ▶ Lösungsbaum nicht stabil (→ existierende Lösungsvorschläge)
- ▶ Overfitting

Arbeitsschritte der Cluster-Analyse

1. Datensatzaufarbeitung
2. CART-Analyse
3. Evaluation
 1. Cross-Validation
 2. (Random Forests)

Themenvorschläge

The background features abstract, overlapping geometric shapes in various shades of green, ranging from light lime to dark forest green. The shapes are primarily triangles and polygons, creating a dynamic, layered effect. The overall composition is clean and modern, with the text 'Themenvorschläge' centered on the left side of the page.

Literaturüberblick: Einsatz von CART in den WiWi

- ▶ Für welche Analysen wurden CART im wirtschafts-wissenschaftlichen Kontext genutzt?
 - ▶ Literaturübersicht
 - ▶ Usecases vorstellen
 - ▶ Vor-/Nachteile in den entsprechenden Anwendungsfeldern

Detaillierte Beschreibung des Algorithmus und dessen Implementierung

- ▶ (einfache) Umsetzung in R oder Python
- ▶ Diskussion der Komplexität des Algorithmus
- ▶ Ausblick/Vergleich mit Varianten des Algorithmus

Anwendungsbeispiel

- ▶ Vorstellung der Methode (inkl. übliche Arbeitsschritte bei der Anwendung) + Illustration an einem einfachen Beispieldatensatz

Weiteres Vorgehen

- ▶ Themenauswahl (allein oder in 2er-Gruppen)
 - ▶ Ziele für die Präsentation:
 - ▶ Methodischen/Theoretischen Überblick
 - ▶ Darstellung/Erklärung der Ideen und Funktionsprinzipien
-
- ▶ Konkretisierte Fragestellung für die Abschlussarbeit
 - ▶ Beabsichtigte Gliederung der späteren Abschlussarbeit
 - ▶ Ausstehende Arbeitsschritte und beabsichtigtes vorgehen

Sprechstunde

auf Anfrage (weiserc@uni-mainz.de)