

Vorlesungsplan

- 17.10. Einleitung
- 24.10. Ein- und Ausgabe
- 31.10. Reformationstag
- **7.11. Einfache Regeln**, Naïve Bayes
- 14.11. Entscheidungsbäume, und Regeln
- 21.11. Assoziationsregeln
- 28.11. Lineare Modelle
- 5.12. Instanzbasiertes Lernen, Clustering
- 12.12. Evaluation I
- 19.12. Evaluation II
- 9.1. Entscheidungsbäume, Klassifikationsregeln
- 16.1. Lineare Modelle, Numerische Vorhersage
- 23.1. Clustering
- 30.1. Attribut-Selektion, Diskretisierung, Transformationen
- 6.2. Kombination von Modellen, Lernen von nicht-klassifizierten Beispielen

Zuerst die einfachen Methoden

- Einfache Algorithmen funktionieren oft sehr gut!
- Es gibt viele Beispiele für Einfachheit:
 - Ein Attribut erledigt die ganze Arbeit
 - Alle Attribute tragen gleichmäßig & unabhängig zum Ergebnis bei
 - Eine gewichtete Linearkombination reicht aus
 - Instanz-basiert: nutze wenige Prototypen
 - Nutze einfache logische Regeln
- Erfolg hängt vom Anwendungsgebiet ab

Ableiten rudimentärer Regeln

- 1R: lernt einen 1-level Entscheidungsbaum
 - Regeln, die nur ein bestimmtes Attribut testen
- Basic Version
 - Ein Zweig für jeden Wert
 - Jeder Zweig weist die häufigste Klasse zu
 - Fehlerrate: Anteil der nicht zur Mehrheitsklasse in den jeweiligen Zweigen gehört
 - Wähle Attribut mit niedrigster Fehlerrate

(nimmt *nominale Attribute* an)

Pseudo-code für 1R

```
For each attribute,  
  For each value of the attribute, make a rule as follows:  
    count how often each class appears  
    find the most frequent class  
    make the rule assign that class to this attribute-value  
  Calculate the error rate of the rules  
Choose the rules with the smallest error rate
```

- Bemerkung: “missing” wird als separater Attributwert behandelt

Auswerten der Wetterdatenattribute

Outlook	Temp	Humidity	Windy	Play	Attribute	Rules	Errors	Total errors
Sunny	Hot	High	False	No	Outlook	Sunny → No	2/5	4/14
Sunny	Hot	High	True	No		Overcast → Yes	0/4	
Overcast	Hot	High	False	Yes		Rainy → Yes	2/5	
Rainy	Mild	High	False	Yes	Temp	Hot → No*	2/4	5/14
Rainy	Cool	Normal	False	Yes		Mild → Yes	2/6	
Rainy	Cool	Normal	True	No		Cool → Yes	1/4	
Overcast	Cool	Normal	True	Yes	Humidity	High → No	3/7	4/14
Sunny	Mild	High	False	No		Normal → Yes	1/7	
Sunny	Cool	Normal	False	Yes	Windy	False → Yes	2/8	5/14
Rainy	Mild	Normal	False	Yes		True → No*	3/6	
Sunny	Mild	Normal	True	Yes				
Overcast	Mild	High	True	Yes				
Overcast	Hot	Normal	False	Yes				
Rainy	Mild	High	True	No				

* Gleichstand

Numerische Attribute

- Diskretisiere numerische Attribute
- Teile den Bereich jeden Attributes in Intervalle
 - Sortiere Instanzen nach den Werten des Attributes
 - Platziere Trenner wo die Klasse sich ändert (Mehrheitsklasse)
 - dies minimiert den Gesamtfehler

Numerische Attribute

- Beispiel: *temperature* aus den Wetterdaten

sunny,85,85,FALSE,no
 sunny,80,90,TRUE,no
 overcast,83,86,FALSE,yes
 rainy,70,96,FALSE,yes
 rainy,68,80,FALSE,yes
 rainy,65,70,TRUE,no
 overcast,64,65,TRUE,yes
 sunny,72,95,FALSE,no
 sunny,69,70,FALSE,yes
 rainy,75,80,FALSE,yes
 sunny,75,70,TRUE,yes
 overcast,72,90,TRUE,yes
 overcast,81,75,FALSE,yes
 rainy,71,91,TRUE,no

Problem: Overfitting

- Diese Prozedur ist sehr anfällig für Rauschen
- Eine Instanz mit falscher Klasse kann ein separates Interval erzeugen
- Einfache Lösung:
erzwinge eine minimale Anzahl von Instanzen der Mehrheitsklasse pro Interval
- Diskretisierung hängt von der Verarbeitungsrichtung ab

Mit Vermeidung von Overfitting

- Ergebnis Regelmenge:

Attribute	Rules	Errors	Total errors
Outlook	Sunny → No	2/5	4/14
	Overcast → Yes	0/4	
	Rainy → Yes	2/5	
Temperature	≤ 77.5 → Yes	3/10	5/14
	> 77.5 → No*	2/4	
Humidity	≤ 82.5 → Yes	1/7	3/14
	> 82.5 and ≤ 95.5 → No	2/6	
	> 95.5 → Yes	0/1	
Windy	False → Yes	2/8	5/14
	True → No*	3/6	

Diskussion von 1R

- 1R wurde in einem Artikel von Holte (1993) beschrieben
 - Enthält eine experimentale Auswertung mit 16 Datenmengen (mit *cross-validation* so dass die Ergebnisse repräsentativ sind)
 - Minimale Anzahl von Instanzen wurde auf 6 gesetzt, nach ersten Experimenten
 - 1R's einfache Regeln waren nicht viel schlechter als viel komplexere Entscheidungsbäume

- Einfache Methoden zuerst testen!!!

Very Simple Classification Rules Perform Well on Most Commonly Used Datasets

Robert C. Holte, Computer Science Department, University of Ottawa



Erweiterung von 1R

- Nutze für jede Klasse eine separate Regel
- jede Regel ist eine UND-Verknüpfung von Tests
 - Nominal testet auf Element in Teilmenge
 - Numerisch: testet auf enthalten in Intervall
- Für jede Klasse c
 - Nom. Attribut: Teilmenge der Attributwerte bei denen c in Trainingsmenge vorkommt
 - Num. Attribut: Intervall $=[\min(c), \max(c)]$ auf der Trainingsmenge
- Regeln überlappen, klassifiziere nach Regel mit den meisten zutreffenden Tests