

# Maschinelles Lernen Symbolische Ansätze:

## Projekt Aufgaben 1-3



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

### Aufgabe 1 - Regellernen: Anwendung und Interpretation

- Verwendete Datensätze & Regellerner

- Vergleich der Ergebnisse: Datensätze

- Vergleich der Ergebnisse: Regellerner

- Zusammenfassung

### Aufgabe 2 - Evaluation von Regellernern

- Verwendete Datensätze & Vorgehen

- Vergleich der Validierungsmethoden

- Unterschiedliche Random Seeds für 10x10 Cross-Validation

- Genauigkeitsmessung auf Testdatensatz

### Aufgabe 3 - ROC-Kurven

- Verwendeter Klassifikationsdatensatz

- Vergleich der erzeugten Kurven

- Interpretation der Ergebnisse

# Aufgabe 1 - Regellernen: Anwendung und Interpretation

## Verwendete Datensätze & Regellerner



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- ▶ Car Evaluation Database
- ▶ Database for Fitting Contact Lenses
- ▶ Zoo Database
  - ▶ Attribut 14 ist numerisch (Anzahl der Beine)
  - ▶ Preprocessing mit Unsupervised Discretiser nötig
  - ▶ Liefert fünf Bins (Intervalle) für 0, 2, 4, 6 oder 8 Beine

→ Klassifizieren mit ConjunctiveRule, JRip und Prism

→ Test mit training set als Testoption //TODO!!!

# Aufgabe 1 - Regellernen: Anwendung und Interpretation

## Vergleich der Ergebnisse: Datensätze



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Car Datensatz (4 Klassen):

- ▶ **ConjunctiveRule**: produziert (`true`  $\Rightarrow$  `class=unacc`) als einzige, bedingungslose Regel
  - ▶ Es wird immer nur eine Klasse vorhergesagt.
- ▶ **JRip**: 49 Regeln mit durchschnittlich 4 Bedingungen
  - ▶ Im Mittel: Vorhersage aller Klassen mit 88% Precision und 87% Recall
- ▶ **Prism**: deutlich mehr Regeln und eine etwas bessere Accuracy als JRip
  - ▶ Aber evtl. Overfitting da sehr viele Bedingungen pro Regel

# Aufgabe 1 - Regellernen: Anwendung und Interpretation

## Vergleich der Ergebnisse: Datensätze



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Contact Lenses Datensatz (3 Klassen):

- ▶ **ConjunctiveRule**: produziert mit (`true`  $\Rightarrow$  `contact-lenses=none`) eine bedingungslose Regel
  - ▶ Es wird nur die häufigste Klasse vorhergesagt (höchster Prior).
- ▶ **JRip**: 3 Regeln mit 0, 1 oder 2 Bedingungen um alle 3 Klassen vorherzusagen
- ▶ **Prism**: Mehr und spezifischere Regeln
  - ▶ Teilweise alle 4 Attribute als Bedingung
  - ▶ Schlechtere Accuracy als JRip

# Aufgabe 1 - Regellernen: Anwendung und Interpretation

## Vergleich der Ergebnisse: Datensätze



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Zoo Datensatz (7 Klassen):

- ▶ **ConjunctiveRule**: produziert (`milk=true`  $\Rightarrow$  `type=mammal`) als einzige Regel mit einer Bedingung
  - ▶ Deshalb werden nicht alle Klassen vorhergesagt.
- ▶ **JRip**: 7 Regeln mit maximal 3 Bedingungen
  - ▶ Accuracy von 89% für alle Klassen zusammen
- ▶ **Prism**: Deutlich speziellere Regeln
  - ▶ Für die meisten Tiernamen direkt den Typ gelernt  $\rightarrow$  fast keine Generalisierung!
  - ▶ Im Vergleich zu JRip: weniger Beispiele korrekt klassifiziert

# Aufgabe 1 - Regellernen: Anwendung und Interpretation

## Vergleich der Ergebnisse: Regellerner



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

ConjunctiveRule:

- ▶ Es wird immer genau eine Regel gelernt.
  - ▶ Die Regel, die häufigste Klasse vorhersagt!
- ▶ In 2 Fällen entspricht die Regel dem höchsten Prior, da sie keine Bedingung hat.
- ▶ Beim Zoo Datensatz wird für Beispiele, bei denen die Bedingung `milk=true`  
 $\Rightarrow$  `type=mammal` nicht erfüllt ist, als Default-Value die nächst häufigste Klasse `bird` zugewiesen

# Aufgabe 1 - Regellernen: Anwendung und Interpretation

## Vergleich der Ergebnisse: Regellerner



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

JRip:

- ▶ Vergleichsweise werden wenige generelle Regeln gelernt.
- ▶ Passt zur Information Gain Heuristik, da diese allgemeinere Regeln bevorzugt
- ▶ Die Default-Rule wählt die häufigste Klasse aus, wenn keine andere Regel davor zutrifft
  - ▶ Prinzip: "Wenn kein weiteres Wissen vorhanden ist, dann wähle die Klasse, die am meisten vorkommt."

# Aufgabe 1 - Regellernen: Anwendung und Interpretation

## Vergleich der Ergebnisse: Regellerner



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Prism:

- ▶ Vergleichsweise werden viele spezielle Regeln gelernt.
- ▶ Passt zur Precision Heuristik, die zu Overfitting neigt.
- ▶ Es gibt keine Default-Rule.



# Aufgabe 1 - Regellernen: Anwendung und Interpretation

## Zusammenfassung



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- ▶ Insgesamt lernt JRip die Datenmengen am Besten.
- ▶ Das Ergebnis passt zur benutzten Heuristik, die allgemeinere Regeln bevorzugt und Overfitting vermeidet.
- ▶ Außerdem lässt sich der Car-Datensatz am genauesten Lernen
  - ▶ (Vermutlich, weil er am Größten ist und dadurch viele Beispiele zum Lernen existieren)
- ▶ Für fundiertere Aussagen sind weitere Untersuchungen mit größeren Datensätzen notwendig, da insgesamt eine große Abhängigkeit von den jeweiligen Daten zu beobachten ist.

## Aufgabe 2 - Evaluation von Regellernern

### Verwendete Datensätze & Vorgehen



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- ▶ Balance Scale Weight & Distance Database
- ▶ Car Evaluation Database
- ▶ Thyroid Disease Records (“Sick” Datensatz)
- ▶ Sonar: Mines vs. Rocks
- ▶ 1984 United States Congressional Voting Records Database

- Datensätze jeweils zufällig mischen.
- Datensätze in 2 gleich große, stratifizierte Teile aufteilen.
- Datensätze mit JRip auf Trainingsdatensatz lernen und evaluieren.

## Aufgabe 2 - Evaluation von Regellernern

### Vergleich der Validierungsmethoden



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Genauigkeitsabschätzungen:

Datensatz	1x5 CV	1x10 CV	1x20 CV	Leave-One-Out	Trainingsmenge
Balance	78.2051	80.1282	75.641	78.2051	83.0128
Car	77.4306	79.8611	80.2083	79.2824	87.5
Sick	98.0382	97.9852	98.1972	98.3563	99.0456
Sonar	75.9615	75	75	73.0769	94.2308
Vote	94.0367	95.4128	93.578	95.4128	95.4128

- ▶ Testen auf der Trainingsmenge ist nicht empfehlenswert → Problem: Overfitting
  - ▶ Accuracy des gelernten Modells wird überschätzt.
- ▶ Die unterschiedlichen Cross-Validation Ergebnisse liefern kein aussagekräftiges Muster bzgl. ihrer Qualität
  - ▶ Es werden weitere Untersuchungen benötigt.

## Aufgabe 2 - Evaluation von Regellernern

### Unterschiedliche Random Seeds für 10x10 Cross-Validation



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Vergleich 1x10 und 10x10 Cross-Validation

Datensatz	1x10 CV	10x10 CV
Balance	80.1282	77.82052
Car	79.8611	79.71065
Sick	97.9852	98.27148
Sonar	75	73.36538
Vote	95.4128	94.9541

- ▶ Gemittelte Genauigkeiten der 10 verschiedenen Durchläufe weichen sehr wenig vom Ursprungsergebnis ab.
- ▶ Auffällig: je nach Random Seed treten während der Berechnung Schwankungen von  $\pm 4\%$  zwischen den einzelnen 1x10 CVs auf.
  - ▶ Verlässliche Aussagen erhält man nach mehreren Durchläufen mit jeweils neuer Random Initialisierung.

## Aufgabe 2 - Evaluation von Regellernern

### Genauigkeitsmessung auf Testdatensatz



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Vergleich 10x10 Cross-Validation vs. Validierungsmenge

Datensatz	10x10 CV	Testmenge
Balance	77.82052	78.2748
Car	79.71065	81.9444
Sick	98.27148	98.0382
Sonar	73.36538	67.3077
Vote	94.9541	95.8525

- ▶ Evaluation des gelernten Modells auf Testdaten liefert in den meisten Fällen eine Accuracy der selben Größenordnung.
  - ▶ Nur im Fall Sonar ist die Abschätzung deutlich niedriger.
- ▶ Für realistischere Validierungen evaluiert man sinnvollerweise mit ungesehenen Testdaten
  - ▶ Damit kann die Generalisierbarkeit besser geprüft werden.

## Aufgabe 3 - ROC-Kurven

### Verwendeter Klassifikationsdatensatz



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- ▶ Verwendung des Datensatz Vote
  - ▶ Klassifizierung jeweils mit J48 und NaiveBayes
- ▶ Danach Generierung beider ROC-Kurven für die Klassen republicans und democrats

# Aufgabe 3 - ROC-Kurven

## Vergleich der erzeugten Kurven

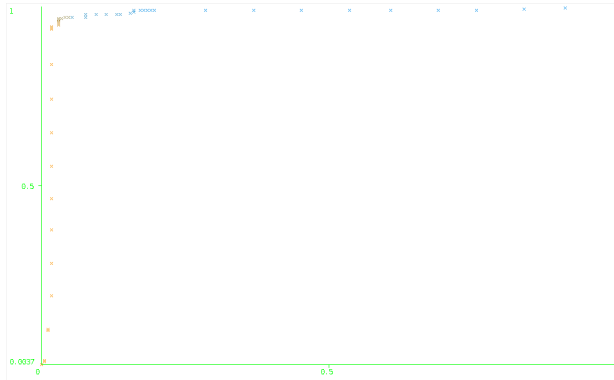


Abbildung : J48 - ROC-Kurve für die Klasse democrats

# Aufgabe 3 - ROC-Kurven

## Vergleich der erzeugten Kurven



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



Abbildung : J48 - ROC-Kurve für die Klasse `republicans`



# Aufgabe 3 - ROC-Kurven

## Vergleich der erzeugten Kurven

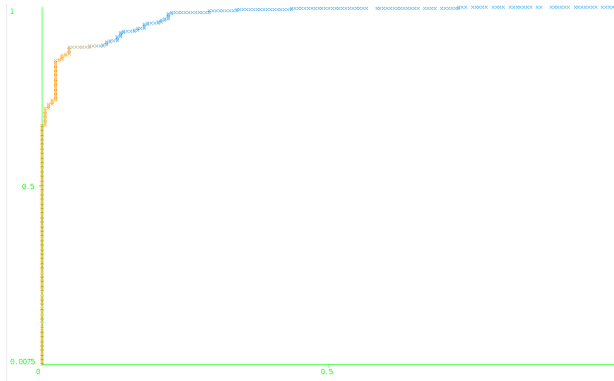


Abbildung : NaiveBayes - ROC-Kurve für die Klasse democrats

# Aufgabe 3 - ROC-Kurven

## Vergleich der erzeugten Kurven



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

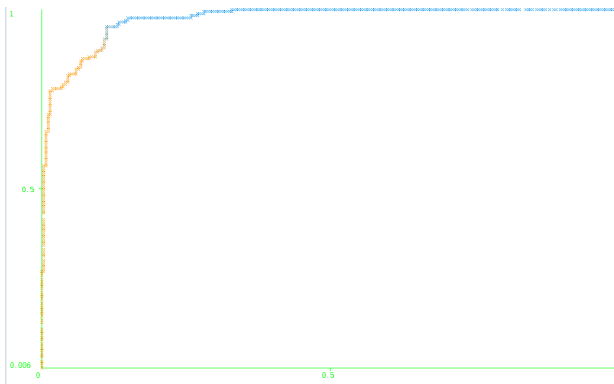


Abbildung : NaiveBayes - ROC-Kurve für die Klasse `republicans`

## Aufgabe 3 - ROC-Kurven

### Interpretation der Ergebnisse



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- ▶ Auffällig ist, dass die Kurven konvex verlaufen.
  - ▶ Es ist eine gute Trennung der Klassen vorhanden.
- ▶ Die ROC-Kurven für J48 nähern sich am meisten dem Punkt der optimalen Theorie an (true-positive-rate 100% und false-positive-rate 0%)
- ▶ Die “Area Under ROC” von J48 hüllt diejenige von NaiveBayes fast in ihrer Gesamtheit ein.
- ▶ Für den verwendeten Datensatz ist J48 fast immer ein besserer Klassifizierer als NaiveBayes.
- ▶ NaiveBayes kann allerdings für sehr steile bzw. flache Kostenverhältnisse sinnvoll sein
  - ▶ Zum Beispiel für minimales fpr und sehr hohe Precision bzw. maximales tpr und sehr hoher Recall



## Aufgabe 1 - Regellernen: Anwendung und Interpretation

- Verwendete Datensätze & Regellerner

- Vergleich der Ergebnisse: Datensätze

- Vergleich der Ergebnisse: Regellerner

- Zusammenfassung

## Aufgabe 2 - Evaluation von Regellernern

- Verwendete Datensätze & Vorgehen

- Vergleich der Validierungsmethoden

- Unterschiedliche Random Seeds für 10x10 Cross-Validation

- Genauigkeitsmessung auf Testdatensatz

## Aufgabe 3 - ROC-Kurven

- Verwendeter Klassifikationsdatensatz

- Vergleich der erzeugten Kurven

- Interpretation der Ergebnisse