

Maschinelles Lernen Symbolische Ansätze:

Projekt Aufgaben 1-3



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Aufgabe 1 - Regellernen: Anwendung und Interpretation

Verwendete Datensätze und 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



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 und dann in 2 gleich große Teile aufteilen
→ Danach mit JRip auf Trainingsdatensatz lernen und evaluieren

Aufgabe 2 - Evaluation von Regellernern

Vergleich Verschiedener Validierungsmethoden



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Resultierende 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

→ Definitiv ist Testen auf der Trainingsmenge nicht empfehlenswert, da die Accuracy des gelernten Modells überschätzt wird (Overfitting)

→ Die verschiedenen Cross-Validation Ergebnisse liefern bis jetzt kein aussagekräftiges Muster bzgl. ihrer Qualität (weitere Untersuchung nötig)

Aufgabe 2 - Evaluation von Regellernern

Unterschiedliche Random Seeds für 10x10 Cross-Validation



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Vergleich 1x10 vs. 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 nicht allzu stark vom Ursprungsergebnis ab.

Aber während der Berechnung ist aufgefallen, dass je nach Random Seed Schwankungen von $\pm 4\%$ zwischen den einzelnen 1x10 CVs auftreten.

→ Für robustere Aussagen sind also mehrere Läufe mit jeweils neuer Random Initialisierung sinnvoll

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 meistens eine Accuracy der selben Größenordnung. In einem Fall ist die Abschätzung aber auffallend niedriger.
→ Für eine realistischere Validierung ist es sinnvoll, mit ungesehenen Testdaten zu evaluieren (bessere Überprüfung der Generalisierbarkeit)

Aufgabe 3 - ROC-Kurven

Klassifikationsdatensatz



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- ▶ Datensatz “Vote” jeweils mit J48 und NaiveBayes klassifiziert
- ▶ Anschließend jeweils beide ROC-Kurven für die Klassen “republicans” und “democrats” generiert

Aufgabe 3 - ROC-Kurven

Vergleich der Erzeugten Kurven

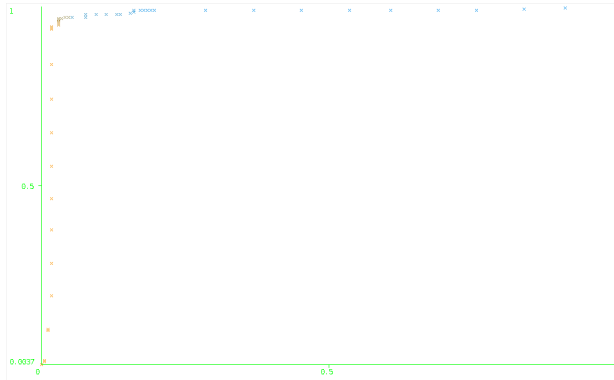


Abbildung: J48 - ROC-Kurve für die Klasse "democrats"

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



TECHNISCHE
UNIVERSITÄT
DARMSTADT

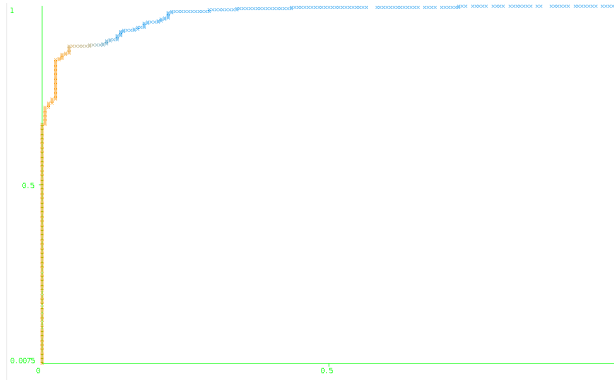


Abbildung: NaiveBayes - ROC-Kurve für die Klasse “democrats”

Aufgabe 3 - ROC-Kurven

Vergleich der Erzeugten Kurven

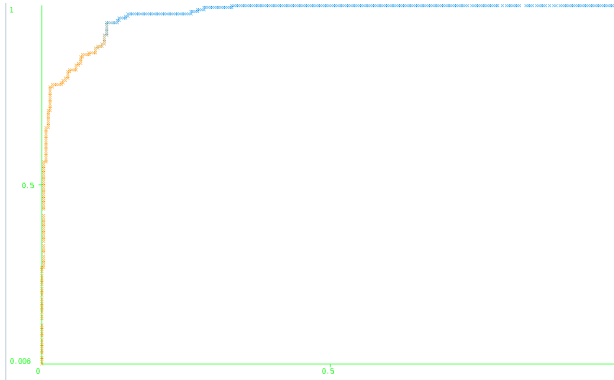


Abbildung: NaiveBayes - ROC-Kurve für die Klasse "republicans"

Aufgabe 3 - ROC-Kurven

Interpretation der Resultate



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Zunächst einmal fällt auf dass die Kurven sehr konvex sind

→ Gute Trennung der Klassen

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 diesen Datensatz scheint J48 also nahezu immer ein besserer Klassifizierer als NaiveBayes zu sein

→ Nur für sehr steile Kostenverhältnisse (fpr minimal, sehr hohe Precision) bzw. für sehr flache Kostenverhältnisse (tpr maximal, sehr hoher Recall) kann NaiveBayes sinnvoll sein



FRAGEN?

Gruppenmitglieder



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Matthias Krebs, 1620340

Thomas Pignede, 1626386

Svenja Stark, 1658147