

Maschinelles Lernen: Symbolische Ansätze



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Wintersemester 2008/2009

1. Projektaufgabe für den 25.11.2008

Aufgabe 1: Regellernen: Anwendung und Interpretation

In dieser einführenden Aufgabe sollen Sie die Verwendung von WEKA erlernen und dabei die Ergebnisse dreier Regellerner auf drei unterschiedlichen Datensätzen vergleichen. Wenden Sie hierzu die Regellerner `ConjunctiveRule`, `JRip` und `Prism` auf die Datensätze `contact-lenses.arff`, `titanic.arff` und `zoo.arff` an. Diese Datensätze liegen entweder in der Installation von WEKA vor oder können unter den angegebenen Links heruntergeladen werden.

Anmerkung: Bei dem Datensatz `zoo.arff` ist das Attribut `legs` numerisch. Da `Prism` numerische Attribute nicht behandeln kann, muß in der Datei der Wert `INTEGER [0,9]` auf `{0,2,4,5,6,8}` geändert werden.

- Vergleichen Sie die Anzahl der Regeln, der Bedingungen und der vorhergesagten Klassen der resultierenden Regelmengen jeweils in Bezug auf
 - die einzelnen Datensätze
 - die jeweiligen Regellerner
- Existiert bei allen Algorithmen eine Default-Rule? Wenn ja:
 - Welche Klasse wird üblicherweise als Default-Rule ausgewählt?
 - Wie interpretieren Sie die Güte dieser Default-Rule?
- Läßt sich anhand der vorherigen Teilaufgaben eine Aussage treffen, welche der drei Datenmengen am leichtesten bzw. am besten zu lernen ist?
- Vergleichen Sie die Regelmengen der Algorithmen `JRip` und `Prism` für den Datensatz `zoo.arff`. Wie schätzen Sie die Allgemeinheit der von `JRip` bzw. `Prism` gefundenen Regeln ein? Beachten Sie hierbei, daß `JRip` als Heuristik `Information Gain` und `Prism` `Precision` verwendet.