

# Maschinelles Lernen: Symbolische Ansätze



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Wintersemester 2008/2009

8. Projektaufgabe für den 12.2.2009

---

## Pre-Processing

---

Wählen Sie drei Datenset aus den fünf hier verfügbaren Datensets aus. Erstellen Sie für jedes Datenset eine diskretisierte Version (`weka.filters.supervised.attribute.Discretize`).

- a) Schätzen Sie die Genauigkeit von J48 mittels Cross-validation auf den ursprünglichen Daten und auf den diskretisierten Daten ab.
- b) Der Meta-Classifer `FilteredClassifier` erlaubt, eine Kombination einer Pre-processing Methode und eines Classifiers zu einem neuen Classifier zu machen. Erzeugen Sie die Kombination `Discretize` und `J48` und schätzen Sie deren Genauigkeit auf den ursprünglichen Daten ab.

Wie interpretieren Sie den Vergleich der Genauigkeiten und der Größe der gelernten Bäume dieser drei Experimente (die Ergebnisse können über die drei Datensets gemittelt werden)? Welche der drei Vorgehensweisen ist ihrer Meinung nach die richtige?