

Übung zur Vorlesung **Maschinelles Lernen**

Wintersemester 2008/2009
Blatt 4

Wiederholung

1. Was genau versteht man unter dem Begriff Modellselektion? Beschreiben Sie das generelle Vorgehen der meisten maschinellen Lernverfahren unter dem Gesichtspunkt der Modellselektion!
2. Betrachten Sie die Modellselektion beim k NN-Verfahren. Welches Ergebnis hätte eine Optimierung des Parameters k bei k NN?

Aufgabe

10 Punkte

Sie haben in der Vorlesung das Prinzip der Baumlerner kennengelernt. Am Beispiel des ID3-Algorithmus' sollen Sie sich das Konzept noch einmal vor Augen führen und in diesem und den nächsten Übungsblättern implementieren. Die erste Fassung Ihres ID3-Algorithmus soll zunächst nur auf polynomiellen und binomiellen Attributwerten arbeiten.

Ein Grundgerüst für Ihre Implementierung finden Sie bereit im `mlv-uebung` Projekt. Das zu implementierende Interface für diese Aufgabe ist

```
com.rapidminer.exercise.id3.ID3Builder1
```

Die zugrundeliegende Implementierung des ersten ID3-Modells wählt dabei die Attribute zum Splitten des `ExampleSets` in der Reihenfolge ihres Vorkommens aus (Nummer der Spalte).

Hinweis: Sie finden im `exercises`-Verzeichnis noch die Klasse

```
com.rapidminer.exercise.ExampleSetTools
```

die einige statische Hilfsmethoden enthält, die Sie für Ihre Implementierung nutzen können.

Zum Testen Ihrer Implementierung können Sie sich mit Hilfe des *NominalExampleSet-Generator* ein eigenes `ExampleSet` mit nominalen Werten generieren lassen.