

Übung zur Vorlesung **Maschinelles Lernen**

Wintersemester 2008/2009
Blatt 5

Aufgabe

10 Punkte

Die Aufgabe auf dem letzten Blatt war die Implementierung eines einfachen Interfaces zum Splitten einer Menge von Beispielen anhand der Werte eines nominalen Attributes. Dabei hat die gegebene Implementierung des ID3-Modells

```
com.rapidminer.exercises.id3.ID3Modell
```

jeweils alle Attribute der Reihenfolge nach zum Aufteilen der Beispiele verwendet. In der Vorlesung wurden bessere Kriterien zur Wahl eines günstigen Merkmals zum Zerlegen der Beispiele vorgestellt. In dieser Aufgabe sollen Sie die Implementierung des ID3-Modells um ein Kriterium zur Attributauswahl erweitern.

1. Erstellen Sie eine Klasse (z.B. `MyFancyID3Modell`), die von `SimpleID3Modell` im Package `com.rapidminer.exercises.id3` erbt und die Methode

```
public void getSplitAttribute( ExampleSet )
```

überschreibt. In dieser Methode soll das Attribut für die nächste Zerlegung von Beispielen jeweils anhand des *Informationsgewinns* ausgewählt werden.

2. Implementieren Sie Ihren eigenen *Lern-Operator*, indem Sie von der Klasse

```
com.rapidminer.exercises.id3.AbstractID3Learner
```

erben und in der `learn(ExampleSet)`-Methode Ihre eigene Modell-Implementierung zurückliefern.

3. Fügen Sie einen Eintrag für Ihren Lern-Operator zu der Datei `operators.xml` hinzu und testen Sie Ihren Lerner wieder mit dem `NominalExampleSetGenerator`.
4. Erstellen Sie ein weiteres Experiment, bei dem Ihr ID3-Lerner auf dem *Golf*-Datensatz lernt. Diesen Datensatz finden Sie bei Ihrer RapidMiner-Installation im Verzeichnis `samples` und im Subversion-Repository.

Der Datensatz enthält zusätzlich numerische Attribute, die mit einem Attribut-Typ-Filter `FeatureValueTypeFilter` von RapidMiner entfernt werden können!

Hinweis: Das Ziel dieses Übungsblattes ist die eigenständige Implementierung eines eigenen Lern-Operators basierend auf den vorgegebenen Basisklassen. Ein Verständnis dieser Klassen ist dafür sehr hilfreich. Sie finden unter

<http://www-ai.cs.uni-dortmund.de/LEHRE/VORLESUNGEN/MLRN/WS0809/javadoc/>

die *Javadoc*-Dokumentation der zu verwendenden Basisklassen, bzw. können diese Dokumentation auch mit Hilfe von `ant` lokal aus Ihrem `mlv-uebung`-Projekt erzeugen, indem Sie den `ant`-Task `javadoc` ausführen - mit Hilfe von Eclipse oder über den Befehl

```
ant javadoc
```

im Projekt-Verzeichnis.

Die *JavaDoc*-Dokumentation aller übrigen RapidMiner-Klassen finden Sie unter:

<http://www-ai.cs.uni-dortmund.de/LEHRE/VORLESUNGEN/MLRN/WS0809/rm-api/>