

## Übung zur Vorlesung **Maschinelles Lernen**

Wintersemester 2013/2014

Blatt 9

Der ID3 Baumlerner soll nun erweitert werden, so dass er auch Splits an numerischen Attributen vornehmen kann.

### **Aufgabe 1**

10 Punkte

Hierzu muss die Klasse

- `com.rapidminer.exercises.id3.ID3Split.java`

implementiert werden, so dass diese nominale und numerische Attribute berücksichtigt. Zur Berechnung der Splits bei numerischen Attributen muss dabei ein `SplitValue` im `SplitPoint` angegeben werden. Dieser wird dadurch ermittelt, dass für alle numerischen Werte an diesem Attribut, die in dem `ExampleSet` vorkommen, genau der Wert als `SplitValue` gewählt wird, der bei Vorhersage der Mehrheitsklasse den kleinsten Fehler liefert.

Wenn ein Beispiel den Wert 7.2 hat, dann werden die übrigen Beispiele in zwei Partitionen unterteilt: Die Werte, die kleiner oder gleich 7.2 und die Werte, die größer als 7.2 sind. In beiden Partitionen wird das häufigste Label vorhergesagt. Die Beispiele, die in dieser Partition nicht der Vorhersage entsprechen, werden zum Fehler hinzugefügt. Der Wert mit dem kleinsten Fehler wird als `SplitValue` gewählt.

Der Operator ist dann erfolgreich implementiert, wenn der Test-Prozess

`//MLV/processes/id3/id3-test`

erfolgreich durchläuft und das Ergebnis mit den Ergebnis

`//MLV/processes/id3/id3-test-expected-port-1`

übereinstimmt. Auf der Übungsseite unter Software steht beschrieben, wie man überprüfen kann, ob der Test-Prozess erfolgreich ist und wie man die Implementierung einreicht.