

Übung zur Vorlesung **Maschinelles Lernen**

Wintersemester 2008/2009
Blatt 3

Wiederholung

1. Erläutern Sie die konzeptionellen Unterschiede zwischen der *linearen Regression* und der Methode der *nächsten Nachbarn*!
2. Was bedeuten die Begriffe *Varianz* und *Bias* im Kontext des maschinellen Lernens?
3. Was ist der *Fluch der hohen Dimensionen*?
4. Was ist mit *Modellselektion* gemeint? Welche Kriterien zur Selektion kennen Sie?

Aufgabe

10 Punkte

Mit dem *k*NN-Algorithmus haben Sie ein instanzbasiertes Lernverfahren kennengelernt. Implementieren Sie ihren eigenen *k*NN-Operator in RapidMiner. Nutzen Sie dazu die abstrakte Klasse

```
com.rapidminer.exercises.IncompleteKNNModel
```

als Ausgangspunkt ihrer Implementierung.

Hinweis: Sie sollen anhand der Klassen des *mlv-uebung*-Projektes auch gleich die Vorgehensweise zur Implementierung eines Lernverfahrens in RapidMiner nachvollziehen: Es existiert ein *Lern-Operator* – im Falle der Übung ist das der Operator, der in der Klasse

```
com.rapidminer.solutions.KNearestNeighborOperator
```

implementiert wird. Dieser erzeugt basierend auf einer Trainingsmenge ein *Modell*, d.h. der Operator bekommt als Input ein *IOObject* vom Typ *ExampleSet* und liefert als Ausgabe unter anderem ein *IOObject* vom Typ *Model*.

Dieses Modell (im aktuellen Fall eine Instanz von *SimplePredictionModel*) stellt je nach Art des Modells unterschiedliche Funktionen bereit, für Vorhersagemodelle beispielsweise die Methode

```
public double predict( Example example ) throws OperatorException
```

Zur Lösung der Aufgabe muss sowohl eine Implementierung des *IncompleteKNNModel* geschrieben werden, als auch der Operator *KNearestNeighborOperator* angepasst werden, so dass dieser ein Objekt Eurer Modell-Implementierung zurückliefert.