

Praxisübung zur 10. Übung „Knowledge Discovery“

Wintersemester 2008/2009

Klassifikation mit k -Nearest-Neighbors

In der ersten Praxisübung haben Sie RapidMiner innerhalb der Eclipse-Entwicklungsumgebung installiert und das prinzipielle Vorgehen beim Entwickeln eines Plugins gelernt. Ziel dieser Übung ist das eigenständige Implementieren des Klassifikations-Algorithmus k -Nearest-Neighbors innerhalb der RapidMiner-Plattform.

Implementieren Sie folgende Variante des k -Nearest-Neighbors-Algorithmus:

- Alle Attribute sind vor der Benutzung auf den Wertebereich $[0, 1]$ zu normieren.
- Als *Distanzfunktion* nutzen Sie bitte die Euklidische-Distanz.
- Beim Abstimmen der k nächsten Nachbarn soll die Stimme jedes Nachbarn mit seiner Entfernung gewichtet werden.

Nutzen Sie die Datei `KDDKNearestNeighbors.java` als Grundgerüst für Ihre Implementierung. Behalten Sie Klassen- und Paketnamen bei und implementieren Sie Ihren Algorithmus in den angegebenen Methoden. Selbstverständlich können Sie ggf. weitere Methoden in der Klasse anlegen.

Beachten Sie bitte zusätzlich folgende Formalitäten:

- Abgabe der Praxisübung: bis **10.2.2009, 23:59 Uhr MEZ** per E-Mail an jaeschke@cs.uni-kassel.de. Spätere Einreichungen werden nicht berücksichtigt!
- Alle benötigten Methoden, Attribute, etc. müssen sich in der Java-Datei `KDDNearestNeighbors.java` befinden, deren Grundgerüst Sie hier finden: <http://www.kde.cs.uni-kassel.de/lehre/ws2008-09/kdd/daten/KDDKNearestNeighbors.java>.
- Am 12.2.2009 werden Sie uns Ihren Algorithmus kurz vorführen, erklären und ggf. Fragen beantworten. Termine (pro Teilnehmer ca. 5 Minuten) werden wir vorher vereinbaren.
- Für eine positive Bewertung muss ihr Code problemlos compilieren, einen Durchlauf durch den Iris-Datensatz aus der vorherigen Praxisübung ohne Fehler bestehen, eine nachvollziehbare Klassifikation liefern (Accuracy > 90% bei $k = 1$ und $split_ratio = 0.7$) und es muss im Gespräch erkennbar sein, dass Sie *Ihre Implementierung* des Algorithmus verstanden haben.

Sollten Sie Fragen haben, wenden Sie sich bitte an Robert Jäschke (jaeschke@cs.uni-kassel.de).
Zusätzlich bieten wir Ihnen an, dass sie immer donnerstags vor und nach der Übung kurz Fragen stellen können. Weitere Hinweise und Dateien finden Sie auch auf der Webseite <http://www.kde.cs.uni-kassel.de/lehre/ws2008-09/kdd/uebung/rapidminer.html>.