Übungen Musterklassifikation

Abgabe am 20.4.2004 12.15 Uhr

Sommersemester 2004

Prof. Dr. Stefan Posch

Dipl.Inform. Denis Williams

Institut für Informatik

Universität Halle

Blatt 1

Aufgabe 1.1 (bewertet)

Welche Schwierigkeiten können Sie sich vorstellen, eine repräsentative Stichprobe zu erstellen? Nennen Sie konkrete Beispiele.

Aufgabe 1.2 (nicht bewertet)

Schreiben Sie ein Programmgerüst (am besten objektorientiert), daß zum allgemeinen Klassifizieren einfacher Mustern dienen soll und in den noch folgenden Übungsserien erweitert wird.

Beginnen Sie, indem Sie eine interne Repräsentation für Stichproben entwerfen. Eine Stichprobe kann klassifiziert sein und trägt einen Namen zur leichten Identifizierung. Schreiben Sie Methoden zur Ein- und Ausgabe einer Stichprobe aus und in eine Datei. Wir verwenden Dateien mit folgender Struktur:

```
# alle Kommentare beginnen mit # und gehen bis zum Ende der Zeile
Stichprobe1              # Name der Stichprobe
5                        # 1.. Anzahl n der folgenen Mustervektoren
3                        # 1.. Dimension d der Mustervektoren
1                        # 0=unklassifiziert, 1=klassifiziert
# alle folgenden Spalten getrennt durch spaces oder tabs (ein oder mehrere)
#   d Spalten Mustervektorkomponenten
#   falls klassifiziert: 1 Spalte Klassenzugeh. (0.., wobei 0=Rueckweisung)
1.231233312 2.12223 3       1
1.2e-3      -5.4234E5   -21.2 2
-234123.23e-04 -33 123.321    0
1 2 3 3
0 0.0 -0.0 1
```

Aufgabe 1.3 (nicht bewertet)

Erweitern Sie Ihr Programmgerüst zu einem benutzbaren Programm, indem Sie ein (einfaches - z.B. zeilenorientiertes) Menü hinzufügen, daß es erlaubt

- eine Stichprobe aus einer Datei einzulesen und Namenskonflikte mit bereits eingelesenen Stichproben aufzulösen (indem ein neuer Name erfragt wird),
- alle Namen der im Speicher befindlichen Stichproben aufzulisten,
- eine Stichprobe per Name auszuwählen (aktuelle Stichprobe),
- die aktuelle Stichprobe in einer Datei abzuspeichern.

Aufgabe 1.4 (bewertet)

Machen Sie einen Vorschlag zum Lernen von Hyperquadern für die Klassifikation beim Zwei-Klassen-Problem. (Der Problemkreis zerfällt hier in zwei Klassen.)

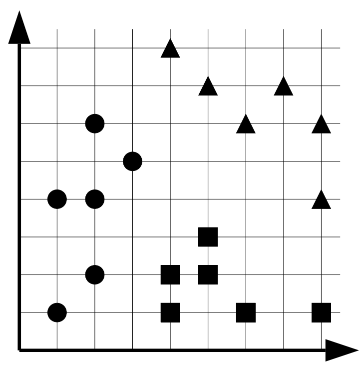
Aufgabe 1.5 (nicht bewertet)

Repräsentieren sie Klassifikator-Parameter zusammen mit ihrem Typ (z.B. Hyperquader) und der Anzahl der Klassen. Jeder Parameter-Satz soll wieder einen Namen zur Identifikation tragen. Bauen Sie in Ihr Programm einen Menüpunkt ein, der die aktuelle Stichprobe nutzt, um mit Hilfe des Algorithmus aus Aufgabe1.4 Parameter für einen zwei Klassen Hyperquader-Klassifikator zu lernen. Ein weiterer Menüpunkt soll es gestatten einen Parametersatz auszusuchen und auf die aktuelle Stichprobe anzuwenden. Das Ergebnis ist eine neue Stichprobe, die die klassifizierten Mustervektoren enthält.

Überlegen Sie, ob, wie und mit welchen Werkzeugen Sie in der Lage wären, eine Stichprobe (mit maximal 2 dimensionalen Vektoren) automatisch zu visualisieren (gnuplot, jai, qt, gl, ...?) (nur Lösungsidee).

Aufgabe 1.6 (bewertet)

Gegeben seien drei Objektklassen, die durch zwei Merkmale charakterisiert werden. Es existiere die in der nebenstehenden Abbildung gezeigte Menge von Referenzobjekten mit bekannten Merkmalswerten.
Skizzieren Sie die sich ergebenden Klassengebiete, wenn mit Hilfe der Nächster-Nachbar-Regel klassifiziert wird. Betrachten Sie dabei die alternative Verwendung folgender Metriken: Euklidischer Abstand, Maximum Norm bzw. City-Block-Abstand. Markieren Sie auch Bereiche, in denen keine Entscheidung für eine bestimmte Klasse möglich ist



Aufgabe 1.7 (bewertet)

Wie sehen die Klassengebiete bei der Klassifikation durch Repräsentanten aus, wenn je Klasse genau ein Repräsentant und der euklidische Abstand verwendet wird?

Aufgabe 1.8 (bewertet)

Machen Sie einen Vorschlag zur Modifikation des (m)-Nächster-Nachbar-Klassifikators zur Reduktion von Speicher- und Rechenaufwand bei im wesentlichen gleicher Klassifikationsleistung

Aufgabe 1.9 (bewertet)

Machen Sie einen Vorschlag zur Modifikation des (m)-Nächster-Nachbar-Klassifikators zur Berücksichtigung unbekannter Bereiche im Merkmalsraum. Dabei soll ein Muster, das keinem Muster der Stichprobe "ähnlich" ist, zurückgewiesen werden.