Ludwig-Maximilians-Universität München
Institut für Informatik
Dr. Peer Kröger
Johannes Aßfalg, Karsten Borgwardt

Knowledge Discovery in Databases

WS 2005/06

Übungsblatt 6

Abgabe aller mit Hausaufgabe *markierten Aufgaben bis Donnerstag, 15.12.2005, 8:30 Uhr, vor der Vorlesung beim Dozenten oder im Übungsbriefkasten*

Aufgabe 6-1 Naive Bayes
Hausaufgabe

Die Ski-Saison ist eröffnet. Um zuverlässig zu entscheiden, wann Sie Skifahren gehen können/sollen, und wann nicht, können Sie einen Klassifikator (z.B. Naive Bayes) benutzen. Der Klassifikator wird mit Ihren Erfahrungswerten aus dem letzten Jahr trainiert. Berücksichtigt werden dabei folgende Attribute:
Das Wetter: Das Attribut `Wetter` kann die folgenden drei Werte annehmen: `Sonne`, `Regen` und `Schneefall`.
Die Schneehöhe: Das Attribut `Schneehöhe` kann die folgenden zwei Werte annehmen: ≥ 50 (Es liegen mehr als 50 cm Schnee) und < 50 (Es liegen weniger als 50 cm Schnee).

Angenommen, Sie wollten letztes Jahr 7-mal zum Skifahren gehen. Die folgende Tabelle gibt Ihre jeweiligen Entscheidungen wieder:

Wetter	**Schneehöhe**	**Skifahren ?**
Sonne	< 50	nein
Sonne	≥ 50	ja
Regen	≥ 50	nein
Schnee	≥ 50	ja
Schnee	< 50	nein
Sonne	≥ 50	ja
Schnee	≥ 50	ja

(a) Berechnen Sie die *a priori* Wahrscheinlichkeiten für die beiden Klassen `Skifahren = ja` und `Skifahren = nein` (auf den Trainingsdaten)!

(b) Berechnen Sie für alle möglichen Attributswerte die Wahrscheinlichkeit, dass aufgrund des Wertes und der Trainingsdaten eine bestimmte Klasse gewählt wird!

(c) Entscheiden Sie, ob Sie bei den folgenden Wetter- und Schneebedingungen Skifahren gehen oder nicht! Verwenden Sie dazu den naiven Bayes-Klassifikator.

	Wetter	**Schneehöhe**
Tag A	Sonne	≥ 50
Tag B	Regen	< 50
Tag C	Schneefall	< 50

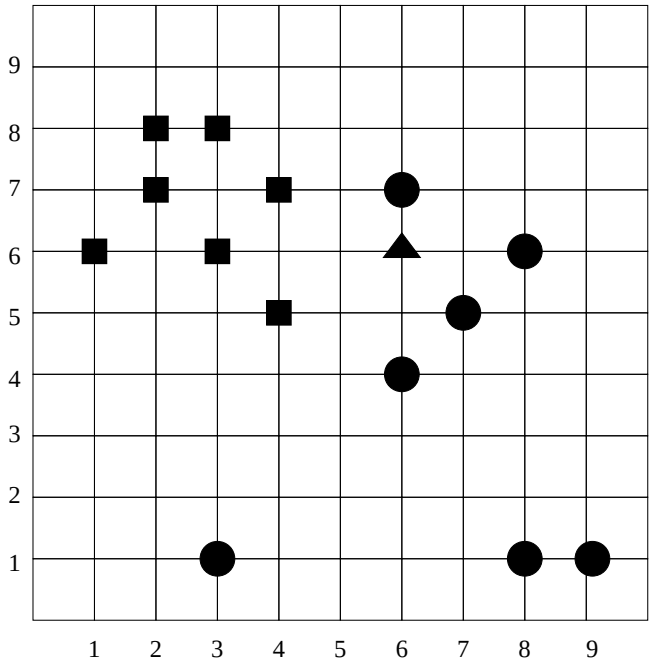
Aufgabe 6-2 Nächste-Nachbarn-Klassifikation I
Hausaufgabe

Geben Sie eine Punktmenge an, bestehend aus mindestens vier 2-dimensionalen Punkten, so dass die Nächste-Nachbarn-Klassifikation ($k = 1$) auf diesen Punkten nur Fehlklassifikationen liefert! Als Distanzfunktion sei die euklidische Distanz gegeben.

Aufgabe 6-3 Nächste-Nachbarn-Klassifikation II

Die 2D Featurevektoren in der nachfolgenden Abbildung seien mit zwei unterschiedlichen Klassenlabeln (Quadrate und Kreise) versehen. Klassifizieren Sie den Punkt (6,6) — im Bild dargestellt durch ein Dreieck — mit einem k-nächsten Nachbarn Klassifikator. Distanzfunktion soll wieder die L_1-Norm (Manhattan-Distanz) sein. Verwenden Sie dabei als Entscheidungsregel die ungewichtete Anzahl der einzelnen Klassen in der k-nächsten Nachbarn Menge, d.h. der Punkt wird der Klasse zugewiesen, die die meisten k-nächsten Nachbarn stellt. Führen Sie die Klassifikation für folgende Werte für k durch und vergleichen Sie die Ergebnisse mit ihrem eigenen intuitiven Ergebnis:

(a) $k = 4$

(b) $k = 7$

(c) $k = 10$