

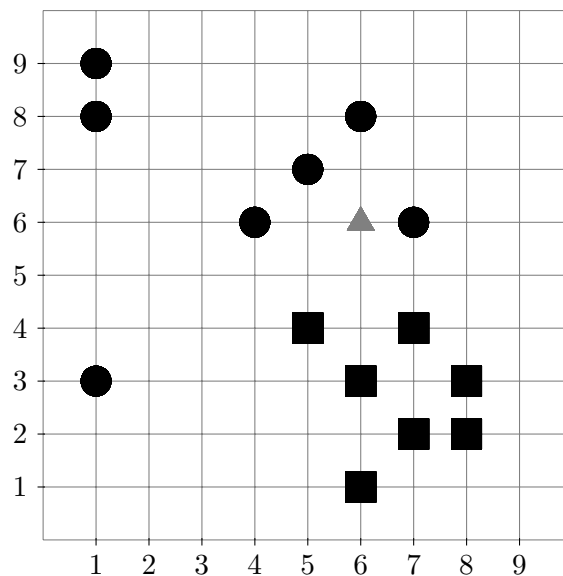
Knowledge Discovery in Databases
SS 2015

Übungsblatt 11: Klassifikation II

Aufgabe 11-1 Nächste-Nachbarn-Klassifikation

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 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) $k = 10$



Aufgabe 11-2 Nächste-Nachbarn-Klassifikation

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 11-3 Entscheidungsbaume

Sie wollen die Risikoklasse eines Autofahrers anhand der folgenden Merkmale vorhersagen:

- Zeit seit Bestehen der Fahrprüfung(1-2 Jahre, 2-7 Jahre, >7 Jahre)
- Geschlecht (männlich, weiblich)
- Wohnort(Stadt, Land)

Für Ihre Analyse stehen Ihnen folgende manuell eingeteilte Testbeispiele zu Verfügung:

Person	Zeit seit der Fahrprüfung	Geschlecht	Wohnort	Risikoklasse
1	1-2	m	Stadt	niedrig
2	2-7	m	Land	hoch
3	>7	w	Land	niedrig
4	1-2	w	Land	hoch
5	>7	m	Land	hoch
6	1-2	m	Land	hoch
7	2-7	w	Stadt	niedrig
8	2-7	m	Stadt	niedrig

- (a) Konstruieren Sie anhand dieser Trainingsdaten einen Entscheidungsbaum. Benutzen Sie beim Split den Informationsgewinn als Maß für die Unreinheit. Erzeugen Sie dabei für jeden Attributwert einen eigenen Ast. Der Entscheidungsbaum soll terminieren, wenn alle Instanzen im Blatt die gleiche Klasse haben. Die Anwendung eines Pruning-Algorithmus ist nicht erforderlich!
- (b) Wenden Sie Ihren Entscheidungsbaum auf folgende Autofahrer an:
Person A: 1-2, w, Land
Person B: 2-7, m, Stadt
Person C: 1-2, w, Stadt