

Knowledge Discovery in Databases  
 SS 2013

Übungsblatt 8: Klassifikation II

**Aufgabe 8-1 Naive Bayes**

Die Ski-Saison ist eröffnet. Um zuverlässig zu entscheiden, wann Sie Skifahren gehen können 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 Schnee.

Die Schneehöhe: Das Attribut `Schneehöhe` kann die folgenden zwei Werte annehmen:  $\geq 50$  (Es liegen mindestens 50 cm Schnee) und  $< 50$  (Es liegen weniger als 50 cm Schnee).

Angenommen, Sie wollten letztes Jahr 8-mal zum Skifahren gehen. Die folgende Tabelle gibt Ihre jeweiligen Entscheidungen wieder:

Wetter	Schneehöhe	Skifahren ?
Sonne	$< 50$	nein
Regen	$< 50$	nein
Regen	$\geq 50$	nein
Schnee	$\geq 50$	ja
Schnee	$< 50$	nein
Sonne	$\geq 50$	ja
Schnee	$\geq 50$	ja
Regen	$< 50$	ja

- (a) Berechnen Sie die *a priori* Wahrscheinlichkeiten für die beiden Klassen `Skifahren = ja` und `Skifahren = nein` (auf den Trainingsdaten)!
- (b) Berechnen Sie für die Klassen die Werteverteilungen aller Attribute.
- (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	$\geq 50$
Tag B	Regen	$< 50$
Tag C	Schnee	$< 50$

### Aufgabe 8-2 Nächste-Nachbarn-Klassifikation

Die 2D Featurevektoren in der nachfolgenden Abbildung seien mit zwei unterschiedlichen Klassenlabels (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)  $k = 10$

