

Knowledge Discovery in Databases
WS 2007/08
Übungsblatt 3

Aufgabe 3-1 Bewertung von Klassifikatoren
Übungsaufgabe

Gegeben sei ein Datensatz mit bekannter Klassenzugehörigkeit der Objekte. Um die Qualität eines Klassifikators K zu ermitteln wurden die Objekte mittels K klassifiziert. Die Klassifikationsergebnisse sind in der folgenden Tabelle dargestellt.

ID	Objektklasse	$K(o)$
O_1	A	A
O_2	B	A
O_3	A	C
O_4	C	C
O_5	C	B
O_6	B	B
O_7	A	A
O_8	A	A
O_9	A	A
O_{10}	B	C
O_{11}	B	A
O_{12}	C	A
O_{13}	C	C
O_{14}	C	C
O_{15}	B	B

Ermitteln Sie über eine 3-fache Überkreuzvalidierung die Validierungsparameter:

- Precision und Recall jeder Klasse
- Micro F_1 -Measure, Macro F_1 -Measure (siehe Übungsblatt 02)

Aufgabe 3-2 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 8-mal zum Skifahren gehen. Die folgende Tabelle gibt Ihre jeweiligen Entscheidungen wieder:

Wetter	Schneehöhe	Skifahren ?
Sonne	< 50	nein
Regen	< 50	nein
Regen	≥ 50	nein
Schnee	≥ 50	ja
Schnee	< 50	nein
Sonne	≥ 50	ja
Schnee	≥ 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 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 3-3 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 3-4 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) $k = 10$

