

Maschinelles Lernen und Data Mining

Übungsblatt für den 7.12.2004

Aufgabe 1

Gegeben sei folgende Regelmenge, die mit dem one-against-all Verfahren gelernt wurde, mit der Anzahl der positiven und negativen Trainings-Beispiele die von jeder Regel abgedeckt werden.

$x :- A = t. \quad [6, 1]$

$y :- B = t. \quad [3, 1]$

$z :- A = t, B = t. \quad [2, 0]$

$z :- C = t. \quad [3, 0]$

- Wie würden Sie die folgenden Beispiele klassifizieren?

A	B	C	
t	f	t	?
f	t	t	?
t	t	f	?
t	t	t	?
f	f	f	?

- Skizzieren Sie die Situation auch als Mengendiagramm, und bestimmen Sie für alle Teile des Diagramms, welche Klassifikation vorgenommen werden soll.
- Können Sie die Regeln in eine sortierte Liste bringen, sodaß Sie die gewünschte Klassifikation erhalten, wenn Sie die erste Regel, die feuert verwenden?

Aufgabe 2

Wieviele positive Beispiele muß eine Regel, die ein negatives Beispiel abdeckt, abdecken, damit CN2's likelihood ratio statistics zu 95% signifikant wird (threshold 3.84) bzw. zu 99% signifikant wird (threshold 6.64)?

Nehmen Sie an, daß die positiven und negativen Beispiele in der Trainingsmenge a) gleichverteilt sind, b) 90% positive und 10% negative Beispiel vorliegen.