

Übung zur Vorlesung „Data Mining in Datenbanken“

# Übungsblatt 1

Abgabe am 20.10.2005

**Aufgabe 1** Testen Sie die Entscheidungsliste auf Folie 9 an den Wetterdaten. Wieviele Beispiele werden korrekt klassifiziert? Wenn Sie die Regeln isoliert betrachten, wieviele Beispiele werden dann noch richtig klassifiziert, wieviele Fehler gibt es, wieviele Widersprüche?

**Aufgabe 2** Die Regeln auf Folie 12 sind Assoziationsregeln, die auf mindestens zwei Beispiele der Wetterdaten anwendbar sind (Voraussetzung paßt) und zu 100% richtig sind. Finden Sie 5 weitere solche Regeln.

**Aufgabe 3** Wieviele Beispiele der Kontaktlinsendaten klassifiziert der Entscheidungsbaum auf Folie 16 richtig?

**Aufgabe 4** Installieren Sie WEKA 3.4 und machen Sie sich mit dem Programm vertraut. Laden Sie mal die mitgelieferten Iris-Daten (Verzeichnis data) und zeichnen Sie alle paarweisen 2D Ansichten, der Attribute petalwidth, petallength, sepalwidth, sepallength wobei die Beispiele farbig nach Klassenzugehörigkeit markiert sind.

Zeigen Sie die Diagramme auf dem Übungsblatt und bereiten Sie eine Powerpoint oder pdf Folie vor. Bringen Sie die Folie elektronisch zur Übung mit. Diskutieren Sie kurz (ein paar Stichpunkte) wie eine Entscheidungsliste und ein Entscheidungsbaum die Datenverteilung ausnutzen kann. Was wären Unterschiede?