

## 12. Übung „Knowledge Discovery“

Wintersemester 2008/2009

### 1 Preprocessing und Datenbeschaffenheit

1. Diskutieren Sie das Verhältnis von Preprocessing und Datenverständnis.
2. In der Statistik unterscheidet man die Attributausprägungen einer vorgegebenen Menge von Daten mittels Skalen mit unterschiedlichem Skalenniveau. Nennen Sie die drei wichtigsten Skalenniveaus und beschreiben Sie sie kurz; geben Sie jeweils ein Beispiel an. Was bedingt ein Skalenniveau bei der Untersuchung von Daten?
3. Welches Ziel wird in Bezug auf die spätere Anwendung von Algorithmen mit dem Preprocessing der Daten im Data Mining verfolgt? Nennen Sie in diesem Zusammenhang zwei Beispiele, in denen Algorithmen bestimmte Preprocessingschritte erzwingen.
4. Nennen Sie je fünf Preprocessingschritte zu Data Cleansing und Data Manipulation und beschreiben Sie drei davon genauer.

### 2 Vergleich verschiedener Verfahren

1. Überlegen Sie sich einen Datensatz, der sich gut für ein Clustering mit OPTICS eignet, aber schlecht für  $k$ Means. Finden Sie auch ein Beispiel, bei dem die Umkehrung gilt?
2. Finden Sie einen Datensatz, der schlecht für eine Klassifikation mit einem Entscheidungsbaum geeignet ist, dafür aber gute Ergebnisse bei  $k$ NN bringt. Finden Sie auch ein Beispiel, bei dem die Umkehrung gilt?