

Knowledge Discovery in Databases II
SoSe 2008

Übungsblatt 6: Verteiltes und Multi-Repräsentiertes Data Mining

Besprechung am Mittwoch, 30.5.2008

Aufgabe 6-1 *Verteilte Multiclass-Klassifikation*

Ihr Datenbestand (Trainings- und Testdaten) ist über n Rechner verteilt und Sie möchten verteilte Multiclass-Klassifikation auf diesen Rechnern betreiben. Welche Vorteile bzw. Nachteile treten auf, wenn Sie dafür

- Entscheidungsbäume,
- Nächste-Nachbarn-Klassifikation,
- Support-Vector-Maschinen oder
- Naive Bayes

verwenden?

Aufgabe 6-2 *Multi-Repräsentierte Klassifikation*

Gegeben sei ein Datensatz mit multiplen Repräsentationen jedes Datenobjekts. Wir möchten Klassenzugehörigkeiten mittels dieser multiplen Repräsentationen ermitteln.

- In welcher Phase des Klassifikationsprozesses können wir die verschiedenen Repräsentationen integrieren?
- Wie können wir die multiplen Repräsentationen beim Trainieren integrieren?
- Wie können wir die multiplen Repräsentationen beim Vorhersagen integrieren?
- Ist in beiden Fällen zuvor eine Normalisierung erforderlich?