

4. Juni 2007

Übung zur Vorlesung

Data Mining mit CI-Methoden
(SS 2007)

Blatt 4

Aufgabe 4 (10 Punkte)

Gegeben seien N Paare jeweils bestehend aus einem Attribut-Tupel und einer zugehörigen Klassenbezeichnung. Erzeugen Sie einen Fuzzy Entscheidungsbaum auf Basis von N_T Datenpaaren, um anschließend die Güte des Klassifikators mit den verbleibenden $N - N_T$ Datenpaaren zu testen. Jedes Attribut soll durch eine linguistische Variable repräsentiert werden. Legen Sie zunächst die jeweiligen linguistischen Terme fest.

Verwenden Sie das Verfahren aus der Vorlesung. Die Originalliteratur befindet sich auf der Webseite der Übung.

- a) Verwenden Sie den Datensatz `iris.data` ($N = 150$). Die *letzte* Spalte enthält den Klassennamen. Wählen Sie $N_T = 100$ Datenpaare zufällig für die Baumerzeugung aus. Dokumentieren Sie die Güte des Klassifikators.
- b) Verwenden Sie den Datensatz `wine.data` ($N = 178$). Die *erste* Spalte enthält den Klassennamen. Wählen Sie $N_T = 120$ Datenpaare zufällig für die Baumerzeugung aus. Dokumentieren Sie die Güte des Klassifikators.

Die Datensätze sind dem *UCI Machine Learning Repository* entnommen. Weitere Informationen finden sich in den zugehörigen Dateien mit der Endung `*.names`, die Sie über den Link `Info` auf der Webseite der Übung erhalten.

Abgabetermin: Donnerstag, 14. Juni 2007, spätestens zur Übung.

Bitte dokumentieren Sie (für sich), wieviel Zeit Sie für die Aufgaben benötigt haben.