

2. Übung zur Vorlesung *Mustererkennung*

Institut für Mathematik und Informatik
Freie Universität Berlin, WiSe 2008/2009
Prof. Dr. Raúl Rojas, Marco Block, Dr. Ernesto Tapia

1. Aufgabe (14 Punkte)

In der Vorlesung wurde LBG als unüberwachter Algorithmus zur Klassifizierung von Daten vorgestellt. Implementieren Sie diesen Algorithmus und clustern Sie die Trainingsdaten für

$$k = 10, 11, 12, \dots, 30.$$

Verwenden Sie ein geeignetes Abbruchkriterium. Ermitteln Sie jeweils die Erkennungsraten für die Testdaten und überlegen Sie sich ein Methode, die in Abhängigkeit von k und der Erkennungsraten entscheidet, welches k für den aktuell verwendeten Datensatz das “beste” ist. Begründen Sie, was Sie für das “beste” halten.

Das Zahlenformat wird in den Tutorien besprochen. Die verwendeten Datensätze finden Sie auf der Veranstaltungshomepage.

Abgabe: Freitag 07.11.2008, 14 Uhr