

17. April 2007

Übung zur Vorlesung

Data Mining mit CI-Methoden
(SS 2007)

Blatt 1

Aufgabe 1.1 (5 Punkte)

Implementieren Sie den Fuzzy-C-Means Algorithmus (FCM) oder besorgen Sie sich eine lauffähige Version, die für die hier vorgesehenen Tests angepasst werden muss.

- a) Testen Sie den Algorithmus für die Daten in den Dateien `uniform.txt`, `uniform4.txt` und `olympic.txt` für $c = 2, \dots, 6$ Cluster und Fuzzifizier $m = \frac{1}{2}, 1, 2, 4$. Dokumentieren Sie die Positionen der Clusterzentren und interpretieren Sie die Zugehörigkeitsmatrix. Die Dateien sind auf der Webseite der Übung verfügbar. Die 2-dimensionalen Datenpunkte sind zeilenweise im Textformat gespeichert.
- b) Berechnen Sie die Gütemaße namens Partitionskoeffizient, Partitionsentropie und Xie-Beni-Index für die obigen Experimente und dokumentieren Sie die Werte. Lässt sich daraus eine „richtige“ Clustereinteilung ablesen?

Aufgabe 1.2 (5 Punkte)

Was muss am FCM-Algorithmus verändert werden, wenn als Abstandsnorm die ℓ_1 -Norm

$$d(x, y) = \|x - y\|_1 = \sum_{i=1}^d |x_i - y_i|$$

verwendet wird? Begründen Sie Ihre Aussagen.

Abgabetermin: Dienstag, 24. April 2007, 20:00 Uhr per email.

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