

## Übung 8 zur Vorlesung XML-Technologien (Informationssysteme) Sommersemester 2017

Abgabe bis 8:15h am Donnerstag, 13.07.2017

**Aufgabe 1:** Clustering 28 Punkte

Schreiben Sie ein Java-Programm, benutzen Sie sinnvolle Datenstrukturen:

- (a) Lesen Sie die in der Datei *kMeans.txt* bereitgestellten Punkt-Daten aus dem  $\mathbb{R}^2$  in Ihr Programm ein. Sie liegen zeilenweise als geordnete (x,y)-Paare vor, insgesamt sind es 10000 Stück. Wir arbeiten mit doubles:

```
1 -87.07275507941475 -29.125212864876502
2 3.423920388846696 -26.234365557962235
3 -24.46241872732982 87.59820772522687
4 60.151483084364 71.3750921968498
5 ...
```

- (b) Implementieren Sie den in der Vorlesung vorgestellten **k-Means-Algorithmus** für ein variables  $k$ .
- (c) Geben Sie abschließend die **Qualität des Clusterings** auf der Konsole aus.

**Hinweis:** Nicht lauffähige/kompilierbare Programme führen bereits vor der eigentlichen Korrektur zu einem Abzug der Hälfte der Punkte. Geben Sie das Programm über studip ab.

**Aufgabe 2:** Warenkorbanalyse 17 Punkte

Gegeben seien die folgenden Transaktionen:

- 1:{A, B, C}
- 2:{B, E, F}
- 3:{C, A, E}
- 4:{A}
- 5:{B}
- 6:{D, F}
- 7:{B, E}
- 8:{C, A, E, B, F}
- 9:{D, A}

- 10: {A, F, F, E}

- (a) Stellen Sie die gegebenen Transaktionen als binäre Matrix dar.
- (b) Welche Itemsets resultieren aus einer minfreq von 0.25?
- (c) Berechnen Sie den Support und die Konfidenz von  $\{A\} \rightarrow \{B, C\}$ .
- (d) Benutzen Sie den Apriori-Algorithmus um diejenigen Itemsets zu finden, deren min-sup  $= 2$  ist. Gehen Sie den Algorithmus schrittweise auf Papier durch. Verwenden Sie für die Kandidatengenerierung die  $F_{k-1} \times F_{k-1}$ -Methode.

Diese Aufgaben werden am 17. Juli in einer normalen Übung besprochen.

### Allgemeine Hinweise:

- Begleitend zur Vorlesung findet eine wöchentliche Übung statt. Zwei Übungstermine werden Onlineübungen am Rechner sein, der **19.06** und der **26.06**.
- Bitte beachten Sie, dass es drei Gruppen gibt: Mo 9:10-9:55 Uhr; Mo 10:00-10:45 Uhr und Mo, 10:50-11:35 Uhr (alle H523). Sie haben sich entweder selbst in eine der drei Übungsgruppen eingetragen oder wurden zufällig einer zugewiesen.
- Um die Klausurzulassung zu erhalten, müssen mindestens 50% der möglichen Übungspunkte in Heimübungen und mindestens 25% der möglichen Übungspunkte in Onlineübungen erreicht werden. Um an den Online-Übungen teilzunehmen, benötigen Sie keinen Informatik-CIP-Pool-Account.
- Gruppenabgabe bis zu 3 Studierenden ist in einer festen Konstellation zulässig, pro Gruppe ist nur eine Abgabe notwendig. Vermerken Sie alle Teilnehmer der Gruppe auf der Abgabe. Die Übungen sind fristgerecht abzugeben:
  - Im DBIS-Postkasten im Raum H 540 am Ende des Flurs („Wintergarten“).
  - Über den Dateibereich der Übung im studip als (.zip, o.ä.)-Archiv oder als PDF. Benennen Sie Ihre Abgabe nach folgendem Schema:

**Voller\_Name\_Matr-Nr\_Übungs-Nr.zip**

also bspw. so: **Hans\_Meier\_1234567\_u1.zip**. Möchten Sie eine bereits erfolgte Abgabe korrigieren, so laden Sie ein entsprechendes Archiv aller abzugebender Dateien erneut in den „Hausarbeiten“-Ordner. Achten Sie darauf, Ihre(n) Namen in jeder Datei anzugeben.

- Vor der Vorlesung.

Verspätete Abgaben werden nicht gewertet. Falls Sie Ihre Abgabe teils digital und teils analog abgeben, vermerken Sie dies bitte auf beiden Teilen.

- Korrekturen analoger und digitaler Abgaben werden in der nächsten regulären Übung zurückgegeben.
- Ankündigungen zur Vorlesung oder Übung erfolgen über den entsprechenden Bereich im studip.
- Einen Teilnahmechein können Sie erhalten, wenn Sie 50% der Heim-Übungspunkte und 25% der Online-Übungspunkte erreicht haben.