

Übungsblatt 1

Jutta Mülle, muelle@kit.edu, IPD Böhm, KIT

Dieses Übungsblatt wird nicht bewertet – es müssen keine Bearbeitungen abgegeben werden (keine Relevanz für Klausurbonus).

Ausgabe: 11.05.2017
Besprechung: Montag, 22.05.2017

Aufgabe 1 – Outlier und Clustering

- a) In diesem Aufgabenteil geht es um den Algorithmus *DBSCAN*; verwenden Sie genau die Variante aus der Vorlesung und wenden Sie ihn auf den kleinen Datenbestand gemäß Abbildung 1 an. Um jedes Datenobjekt, gekennzeichnet durch "x" und benannt mit einer Nummer, ist ein Kreis mit Radius ϵ eingezeichnet. Der Algorithmus betrachtet das Objekt mit der kleinsten Nummer, das noch nicht angefasst wurde, als nächstes.

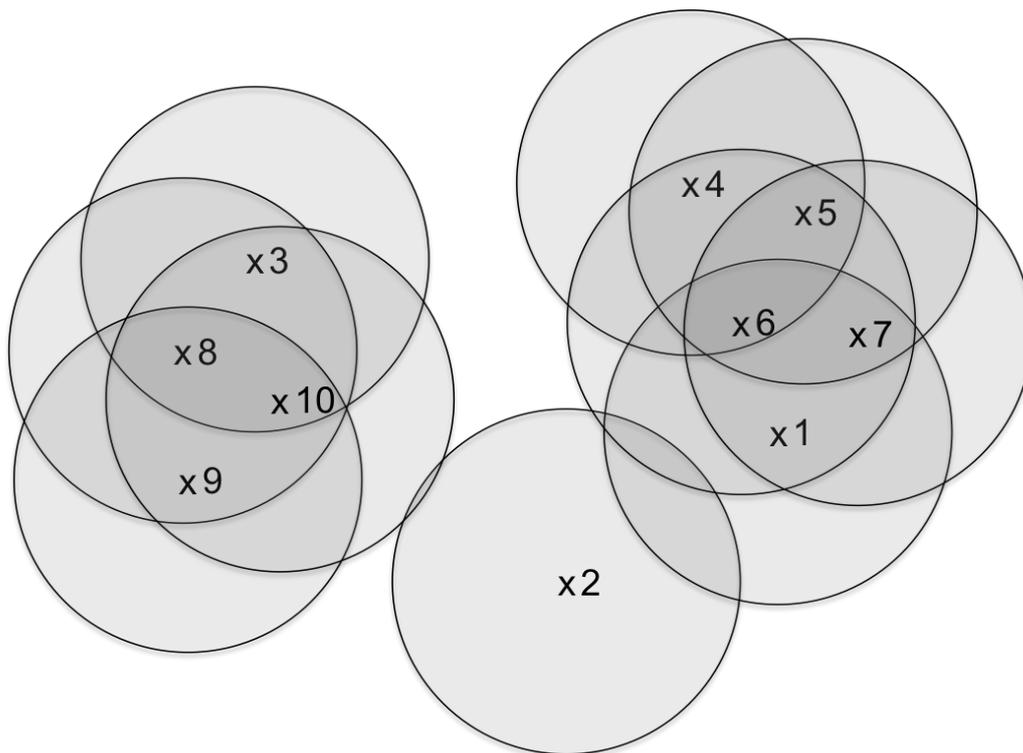


Abbildung 1: Datenbestand mit Objekten 1, ..., 10.

- b) Sei $MinPts = 4$. Welche Objekte sind dicht? Welche Objekte sind Dichte-erreichbar? Wie genau sieht das Clustering-Resultat mit *DBSCAN* aus? Schreiben Sie das Resultat wie folgt hin: $C_1 = \{U, V, \dots\}, C_2 = \{X, Y, \dots\}, \dots$. Dabei sind U, V, X, Y usw.

Nummern von Objekten. Die Reihenfolge der Cluster muss die sein, in der *DBSCAN* sie generiert. Bitte lassen Sie Noise weg, falls *DBSCAN* eine derartige Ausgabe generieren würde.

- c) Sei $MinPts = 2$. Welche Objekte sind dicht? Welche Objekte sind Dichte-erreichbar? Wie genau sieht das Clustering-Resultat mit *DBSCAN* aus? Schreiben Sie das Resultat wie folgt hin: $C_1 = \{U, V, \dots\}, C_2 = \{X, Y, \dots\}, \dots$. Dabei sind U, V, X, Y usw. Nummern von Objekten. Die Reihenfolge der Cluster muss die sein, in der *DBSCAN* sie generiert. Bitte lassen Sie Noise weg, falls *DBSCAN* eine derartige Ausgabe generieren würde.
- d) Für den Fall, dass $MinPts = 1$, lässt sich stets eine Formel angeben für die Anzahl der Dichte-erreichbaren Objekte. Schreiben Sie diese Formel hin. Erklären Sie ggf. neue Parameter. Erklären Sie kurz in natürlicher Sprache, warum Ihre Formel richtig ist.

Aufgabe 2 – Modellierung mit SQL-DDL

Es soll ein Informationssystem für die Fußballweltmeisterschaft entwickelt werden. Für die Umsetzung wird ein SQL-basiertes Datenbanksystem gewählt.

- a) Überlegen Sie zunächst welche Entitäten im Rahmen einer Fußballweltmeisterschaft von Bedeutung sind, und wie diese zueinander in Beziehung stehen.
- b) Welche Informationsbedürfnisse können Sie sich in diesem Kontext vorstellen? Finden Sie einige typische Anfragen an das System. Ergänzen Sie gegebenenfalls die zu modellierenden Entitäten.
- c) Welche Integritätsbedingungen wären in diesem Szenario sinnvoll?
- d) Setzen Sie die Modellierung, ausgehend von Ihren Überlegungen, direkt in SQL-DDL um.

Aufgabe 3 – Modellierung mit SQL-DDL

Verwenden Sie ein Datenbasisschema mit vier Relationen: Lieferant, Produkt, Kunde und Verträge. Sowohl die Lieferant- als auch die Kunde-Relation haben die Attribute Id, Name und Adresse. Ein Id ist eine Zahl mit neun Ziffern. Produkt hat eine Teilnummer (ein Integer zwischen 1 und 999999) und Name. Jedes Tupel in der Verträge-Relation beschreibt einen Vertrag zwischen einem Lieferanten und einem Kunden für ein bestimmtes Produkt mit einer bestimmten Anzahl und einem Preis.

- a) Benutzen Sie SQL-DDL, um das Schema für diese Relationen zu spezifizieren; geben Sie auch die SQL-Domänen an.
- b) Definieren Sie passende Integritätsbedingungen (Primär-, Kandidaten- und Fremdschlüssel).
- c) Geben Sie eine weitere Integritätsbedingung in natürlicher Sprache an und ergänzen Sie, soweit möglich ohne SQL-Query-Kenntnisse, das Schema.

Aufgabe 4: ER-Modellierung - Krankenhausszenario

Gegeben ist ein Anwendungsszenario, eine sogenannte Miniwelt, mit folgenden Zusammenhängen:

Ein Patient leidet an einer oder mehreren Krankheiten; es gibt jedoch Krankheiten, an denen niemand leidet. Jeder Patient wird genau auf einer Station behandelt. Der Patient wird von einem oder mehreren Ärzten behandelt; jeder Arzt behandelt mindestens 10 und höchstens 50 Patienten, die auf verschiedenen Stationen liegen können. Eine Station kann leer sein, jede Station gehört zu exakt einem Krankenhaus, und jedes Krankenhaus hat mindestens 5 Stationen. An einem Krankenhaus sind zwischen 5 und 200 Ärzte angestellt, ein Arzt ist an mindestens einem und an höchstens 3 Krankenhäusern beschäftigt.

- a) Erstellen Sie ein ER-Diagramm dieser Miniwelt. Nutzen Sie die Konzepte des erweiterten ER-Modells (EER-Modellierung aus der Vorlesung). *Hinweis:* Vergessen Sie nicht die Angabe der Kardinalitäten.
- b) Sind alle Angaben im ER-Diagramm darstellbar? Welche nicht und warum ?
- c) Geben Sie ein Beispiel in diesem evtl. von Ihnen ergänzten Anwendungsszenario für eine Eigenschaft, die mit Standard- aber nicht mit Teilnehmerkardinalität (oder umgekehrt) auszudrücken ist.
- d) Ist in Ihrem ER-Diagramm ein abhängiger Entity-Typ enthalten? Falls nicht, ergänzen Sie das ER-Diagramm mit einem solchen. Modellieren Sie den abhängigen Entity-Typ unter Nutzung des erweiterten Schlüsselkonzepts der EER-Modellierung.