Ludwig-Maximilians-Universität München
Institut für Informatik
PD Dr. Matthias Schubert
Markus Mauder

Knowledge Discovery in Databases II

WS 2014/2015

Übungsblatt 3: High-Dimensional Data

Aufgabe 3-1 Singular Value Decomposition

Another approach to feature reduction is Singular Value Decomposition. Given a Matrix M and its SVD decomposition:

$$M = T * S * D'$$

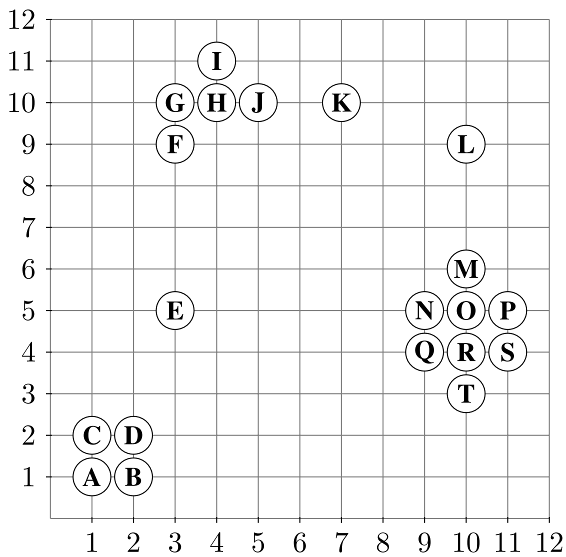
with

$$M = \begin{bmatrix} 1 & 2 \\ 6 & 3 \\ 0 & 2 \end{bmatrix} \qquad T = \begin{bmatrix} -0.2707 & 0.5458 \\ -0.9509 & -0.2797 \\ -0.1497 & 0.7899 \end{bmatrix}$$

$$S = \begin{bmatrix} 7.0257 & 0 \\ 0 & 2.1539 \end{bmatrix} \qquad D = \begin{bmatrix} -0.8507 & -0.5257 \\ -0.5257 & 0.8507 \end{bmatrix}$$

Reduce to one dimension using the approach described in the lecture script.

Aufgabe 3-2 Recapitulating DBSCAN



Compute DBSCAN on the dataset above using Manhattan distance. Indicate core points, border points, and noise points. Use the following parameters:

- Radius $\varepsilon = 1.1$ and $minPts = 3$
- Radius $\varepsilon = 1.1$ and $minPts = 4$
- Radius $\varepsilon = 2.1$ and $minPts = 4$

Aufgabe 3-3 Density-based Subspace-Clustering (SubClu)

Show that the following statement (monotonicity of the core point property) holds:

Let D be a set of d-dimensional feature vectors, $\mathcal{A}$ the set of all attributes (dimensions/features). Further let $p \in D$ and $S \subseteq \mathcal{A}$ be a subspace (attribute subset).

Then the following holds for arbitrary $\epsilon \in \mathbb{R}^+$ and $minPts \in \mathbb{N}$:

$$\forall T \subseteq S \; : \; |\mathcal{N}_\epsilon^S(p)| \geq minPts \; \Rightarrow \; |\mathcal{N}_\epsilon^T(p)| \geq minPts$$

with $|\mathcal{N}_\epsilon^S(p)| := \{q \in D \mid L_P(\pi_S(p), \pi_S(q)) \leq \epsilon\}$.