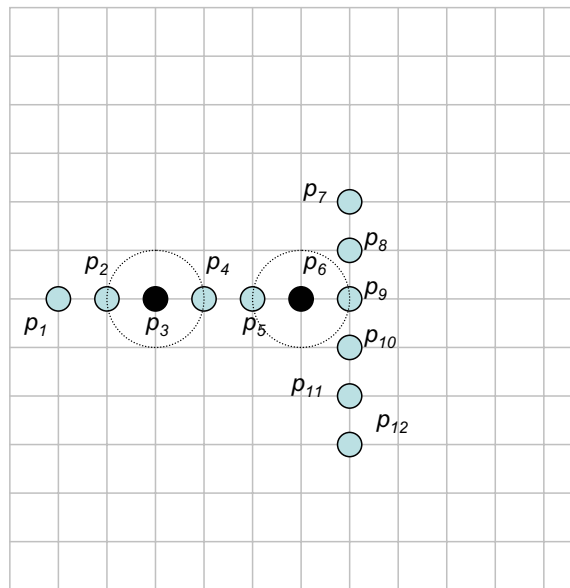


Knowledge Discovery in Databases II
 WS 2013/2014

Übungsblatt 3: Cluster Analysis in High-Dimensional Data

Aufgabe 3-1 Dichte-basiertes Projected-Clustering (PreDeCon)



Gegeben sei obige 2D Datenmenge (der Abstand zwischen den Gitterlinien beträgt 1), die mit euklidischer Distanz verglichen werden soll. Berechnen Sie, ob p_3 und p_6 Kernpunkte im Algorithmus PreDeCon wären. Nehmen Sie hierzu folgende Parameterwerte an: $minPts = 3, \epsilon = 1, \delta = 0.25, \lambda = 1, \kappa = 100$

Aufgabe 3-2 Dichte-basiertes Subspace-Clustering (SubClu)

Beweisen Sie die folgende Aussage (Monotonie der Kernpunkt-Eigenschaft):

Sei D eine Menge von d -dimensionalen Featurevektoren, \mathcal{A} die Menge aller Attribute (Dimensionen/Feature). Sei weiter $p \in D$ und $S \subseteq \mathcal{A}$ ein Unterraum (Attribut-Teilmenge).

Dann gilt für beliebige $\epsilon \in \mathbb{R}^+$ und $minPts \in \mathbb{N}$:

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

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