Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Peer Kröger
Yifeng Lu

Knowledge Discovery in Databases II

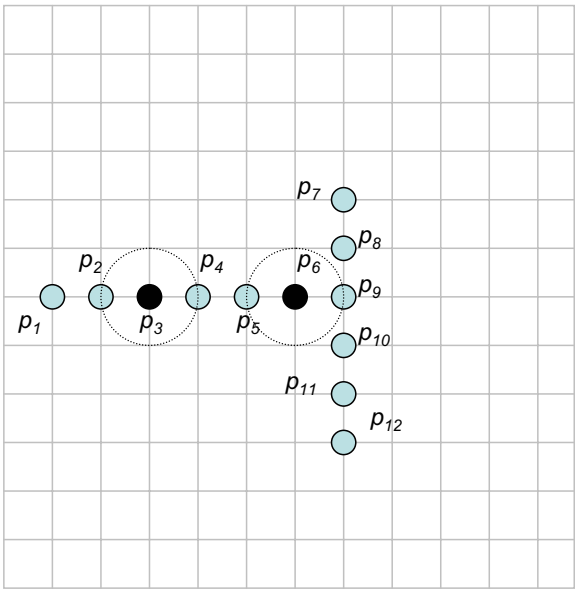
SS 2019

Exercise 5: High Dimensional Data Clustering

Exercise 5-1 Density-based Projected-Clustering (PreDeCon)

PreDeCon uses variance analysis and a weighted Euclidean distance function: For the points in a candidate's ϵ-neighborhood, each dimension whose variance is below δ is weighted more heavily (κ).

Consider the 2D data set shown below. Assume the width of the grid to be 1 unit, use the Euclidean distance function to determine a point's ϵ-neighborhood.



Calculate, if p_3 and p_6 are core points. Assume the following parameter values: $minPts = 3, \epsilon = 1, \delta = 0.25, \lambda = 1, \kappa = 100$

Exercise 5-2 Pattern Based Clustering

Given the following dataset, find out equations (mean and adjustment value) of the following cluster model (one for each):

(a) Biclusters with constant column values

(b) Biclusters with constant row values

(c) Biclusters with coherent values

	a1	a2	a3	a4	a5	a6
p1	1	2	2	2	2	3
p2	1	6	2	2	4	3
p3	1	2	2	2	4	3
p4	2	3	3	3	3	4
p5	3	4	4	3	4	5
p6	4	5	5	3	5	6