Clustering Methods

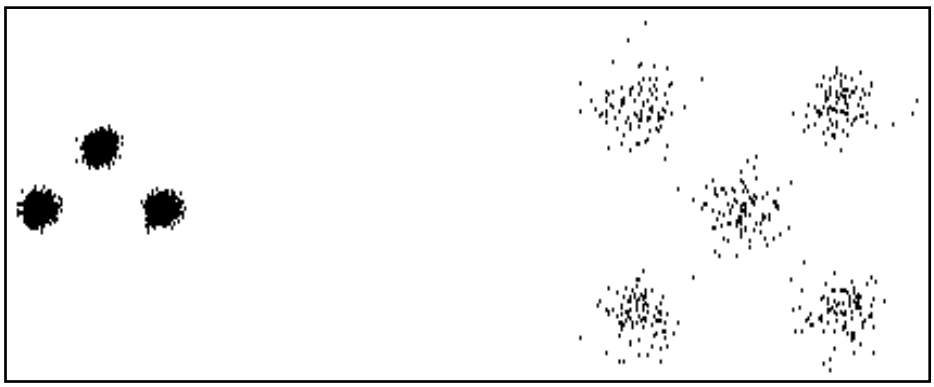
Exercises 4/7

1. Estimating density is the key component in density-based clustering. Suppose three-dimensional data points consisting of attributes name, occupation and language. How would you define density in this case?

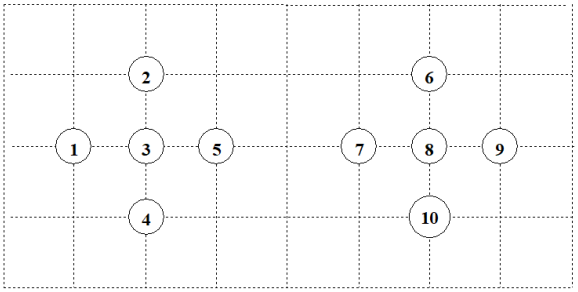
[Zhang, Farmer, Mandarin]

[Malinen, Scientist, Finnish]

2. What are the density values for the points in this data set? Estimate the density values as calculated by DDDE-algorithm?



3. Demonstrate how Density peaks algorithm works for the following data.



4. K-nearest neighbors (KNN) can be used in Density Peaks for estimating density. It can also be used for finding nearest neighbor with bigger density (big brother). Does it work in all situations? If yes, give proof. If not, show example when it does not work. What happens in case the big brother is not reachable via KNN graph?