Algorithmic Data Analysis

Esther Galbrun
Autumn 2022
Q7.1: Outliers in few dimensions

Large red dots are more likely outliers. Match methods to plots

z-number, extreme values of $x_1$
depth-based, peeling layers
density-based, 2D histogram
density-based, nb. of points within fixed radius
clustering model,
minimum Mahalanobis distance
distance-based, distance to third nearest neighbor ($k$-NN)
Q7.2: Further outlier problems

Arrange these task characteristics and tools into four clusters

- high dimensionality
- subspace outlier
- shape outlier
- contextual outlier

- discrete sequences
- angle-based method
- cluster-based method
- grid-based sparsity
- next-element prediction

- time-series
- HOTSAX
- genetic algorithm
- isolation tree
- suffix tree