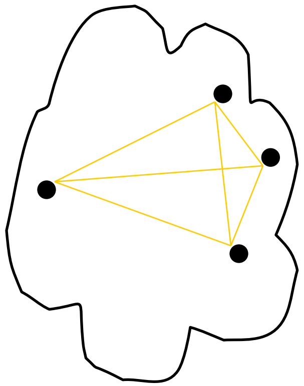
CLUSTERING

CLUSTERING ALGORITHMS

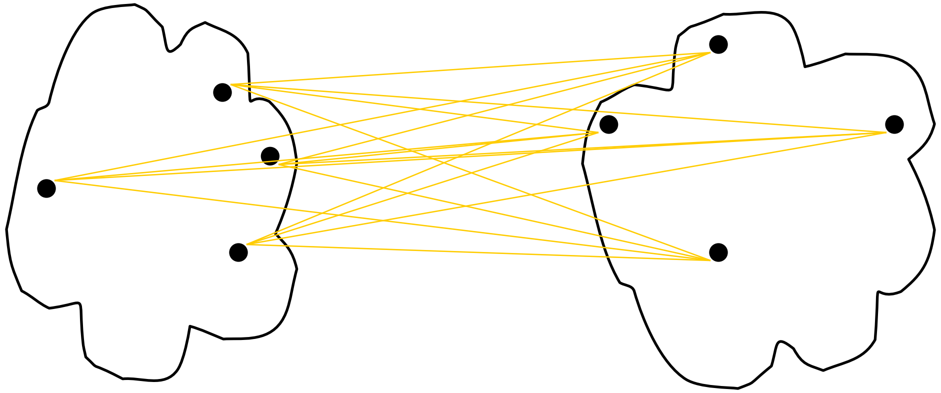
- K-means and its variants
- Hierarchical clustering
- Density-based clustering

UNSUPERVISED MEASURES: COHESION AND SEPARATION

- A proximity graph-based approach can also be used for cohesion and separation.
  - Cluster cohesion is the sum of the weight of all links within a cluster.
  - Cluster separation is the sum of the weights between nodes in the cluster and nodes outside the cluster.



cohesion

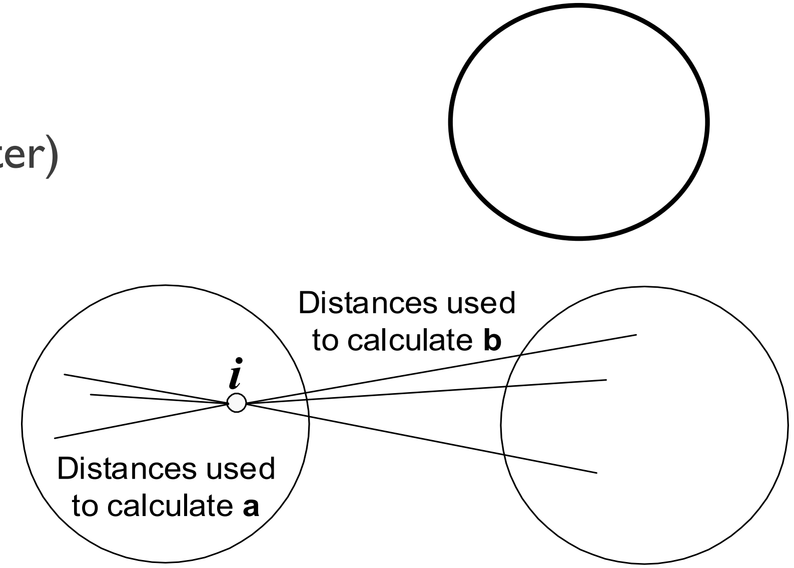


separation

UNSUPERVISED MEASURES: SILHOUETTE COEFFICIENT

- Silhouette coefficient combines ideas of both cohesion and separation, but for individual points, as well as clusters and clusterings
- For an individual point, ***i***
  - Calculate ***a*** = average distance of ***i*** to the points in its cluster
  - Calculate ***b*** = min (average distance of ***i*** to points in another cluster)
  - The silhouette coefficient for a point is then given by

    s = (b – a) / max(a,b)
  - Value can vary between -1 and 1
  - Typically ranges between 0 and 1.
  - The closer to 1 the better.



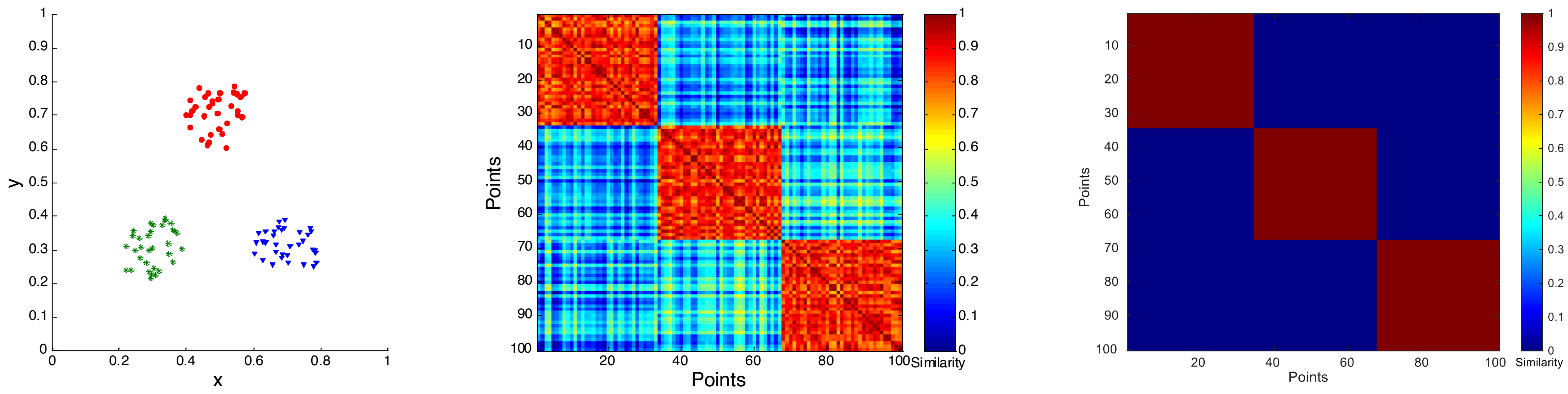
- Can calculate the average silhouette coefficient for a cluster or a clustering

MEASURING CLUSTER VALIDITY VIA CORRELATION

- Two matrices
  - Proximity Matrix
  - Ideal Similarity Matrix
    - One row and one column for each data point
    - An entry is 1 if the associated pair of points belong to the same cluster
    - An entry is 0 if the associated pair of points belongs to different clusters
- Compute the correlation between the two matrices
  - Since the matrices are symmetric, only the correlation between $n(n-1) / 2$ entries needs to be calculated.
- High magnitude of correlation indicates that points that belong to the same cluster are close to each other.
  - Correlation may be positive or negative depending on whether the similarity matrix is a similarity or dissimilarity matrix
- Not a good measure for some density or contiguity based clusters.

MEASURING CLUSTER VALIDITY VIA CORRELATION

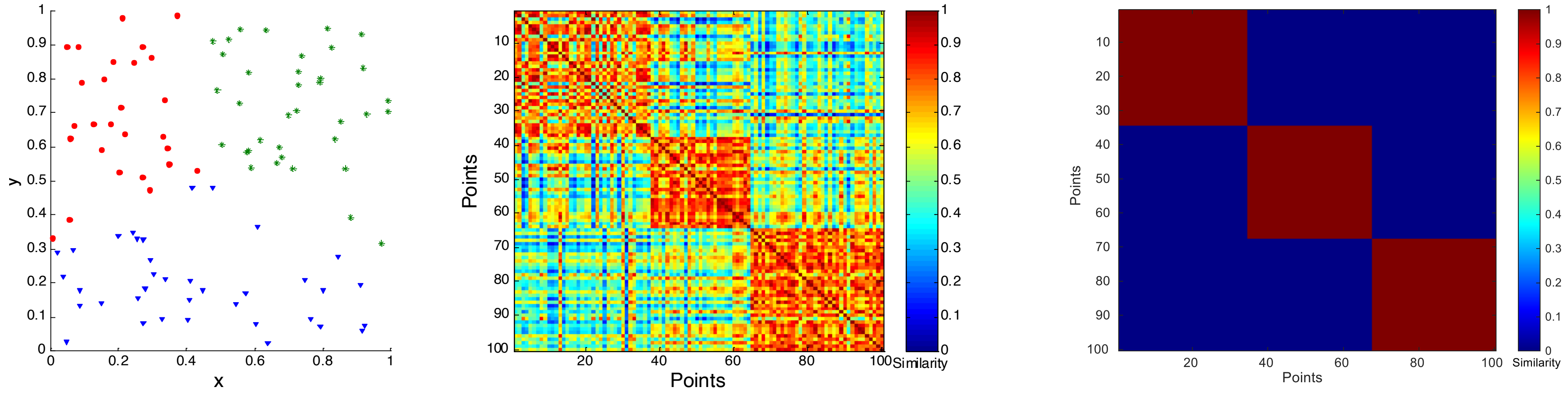
- Correlation of ideal similarity and proximity matrices for the K-means clusterings (partition cluster) of the following well-clustered data set.



Corr = 0.9235

MEASURING CLUSTER VALIDITY VIA CORRELATION

- Correlation of ideal similarity and proximity matrices for the K-means clusterings of the following random data set.

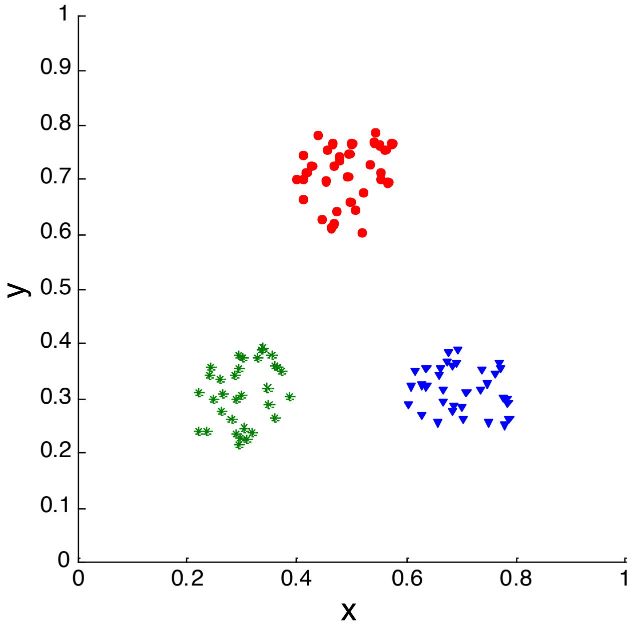


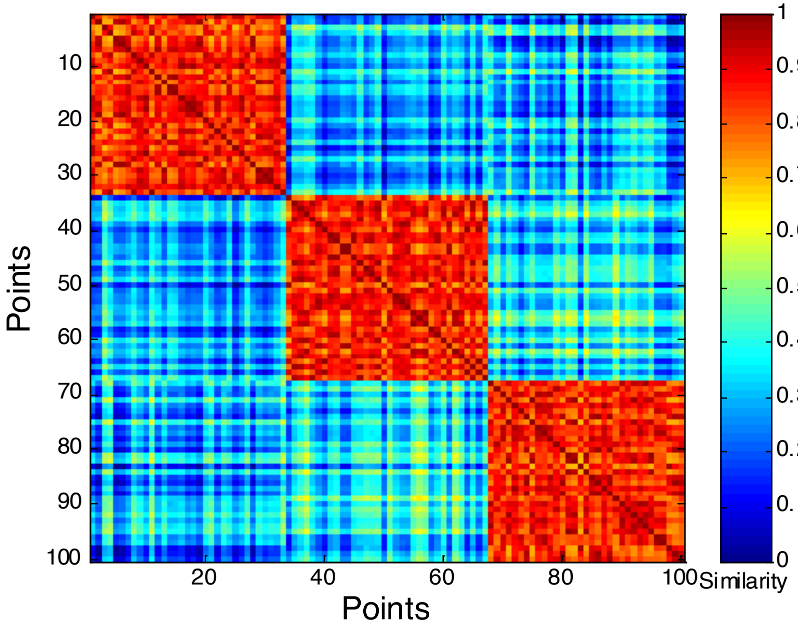
K-means

Corr = 0.5810

JUDGING A CLUSTERING VISUALLY BY ITS SIMILARITY MATRIX

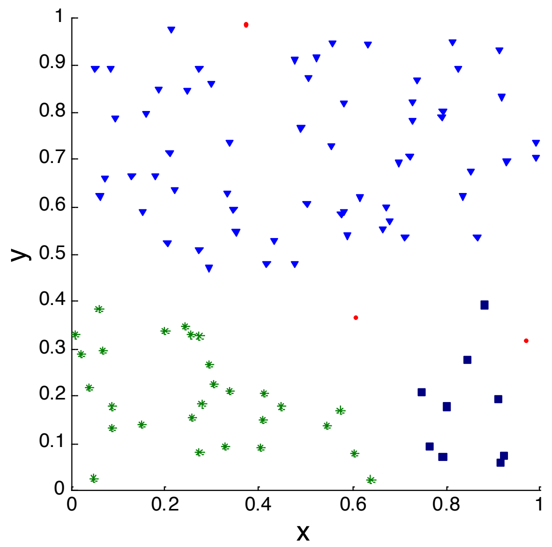
- Order the similarity matrix with respect to cluster labels and inspect visually.

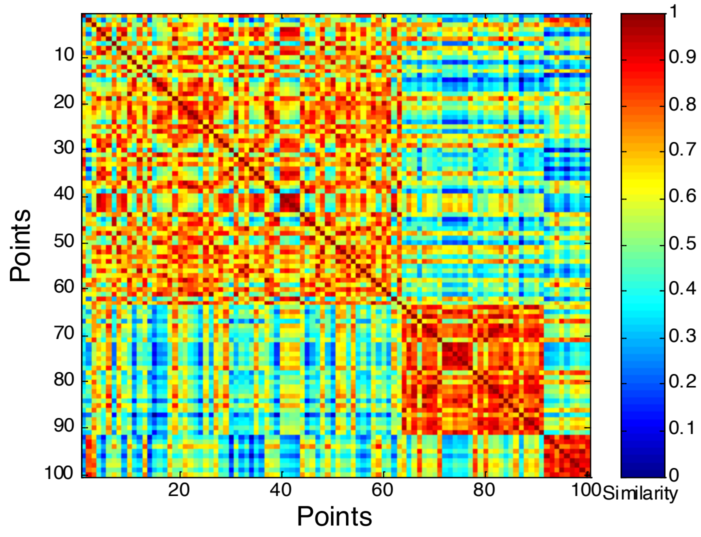




JUDGING A CLUSTERING VISUALLY BY ITS SIMILARITY MATRIX

- Clusters in random data are not so crisp

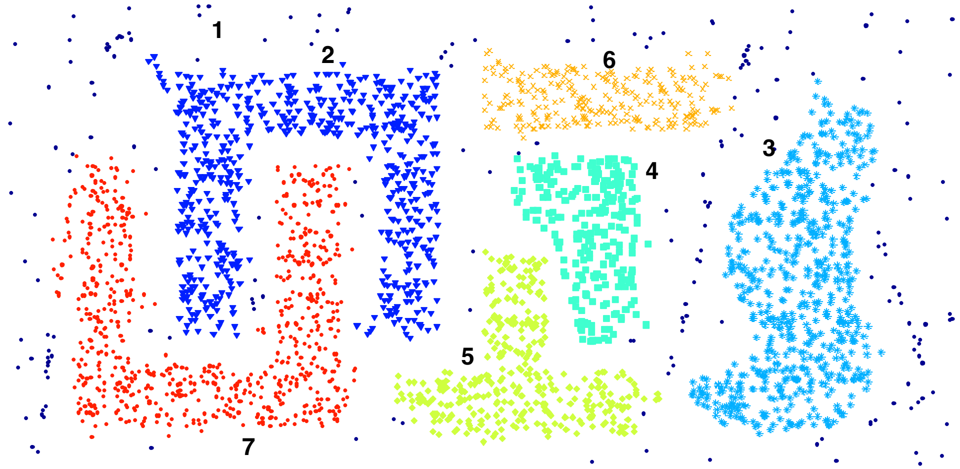


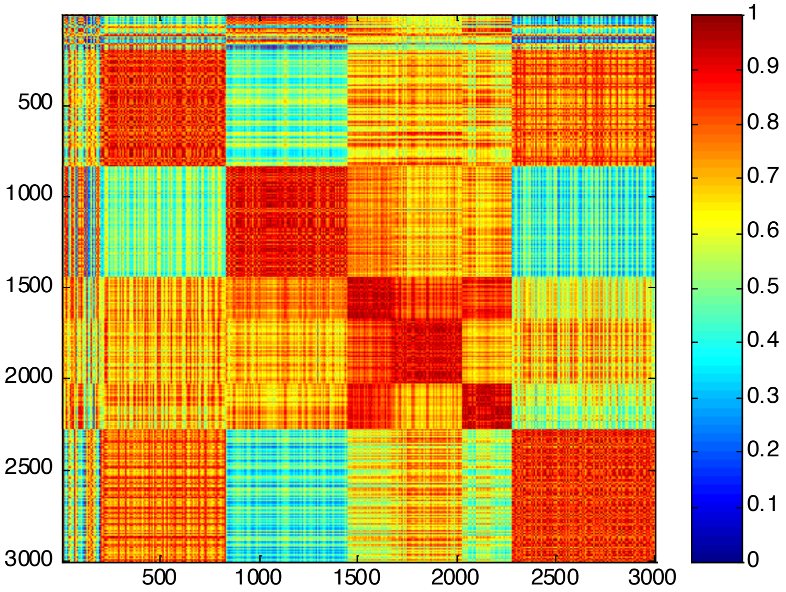


DBSCAN (density based clustering)

Correlation may be not a good measure for some density-based clusters.

JUDGING A CLUSTERING VISUALLY BY ITS SIMILARITY MATRIX





DBSCAN

DETERMINING THE CORRECT NUMBER OF CLUSTERS

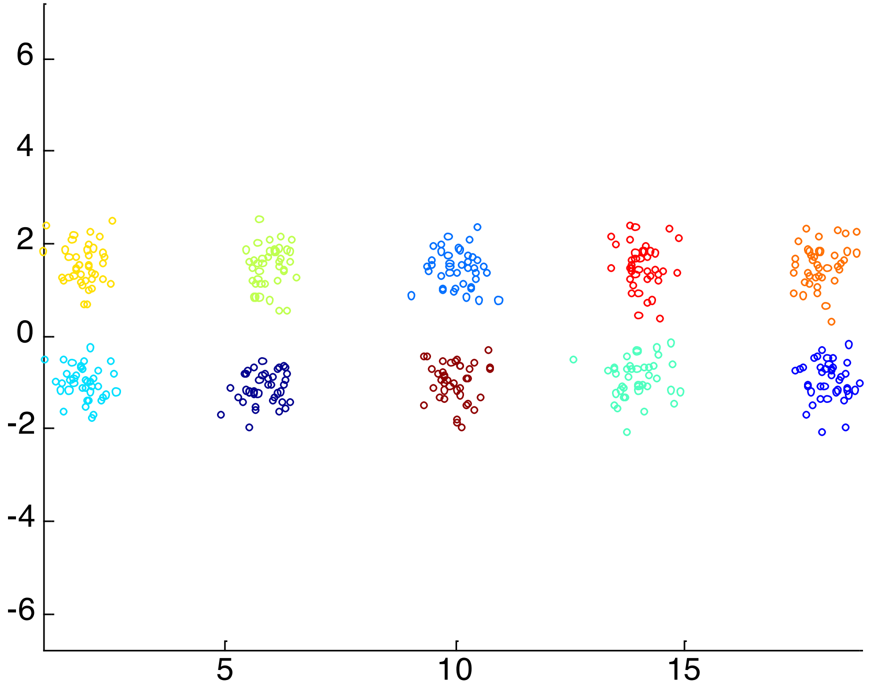
- SSE is good for comparing two clusterings or two clusters
- SSE can also be used to estimate the number of clusters

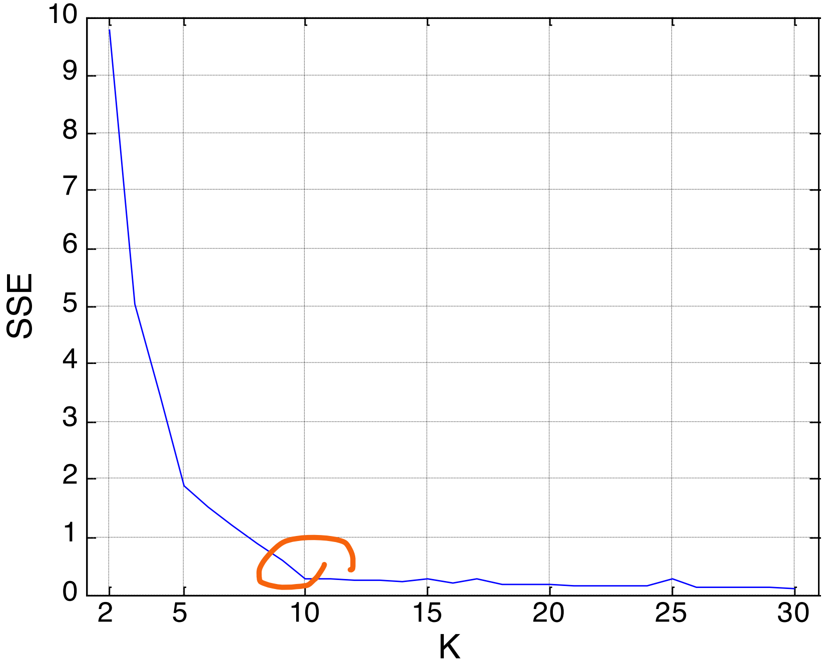
Elbow: after that point, the values of s
Do not change dramastically

K-means

elbow

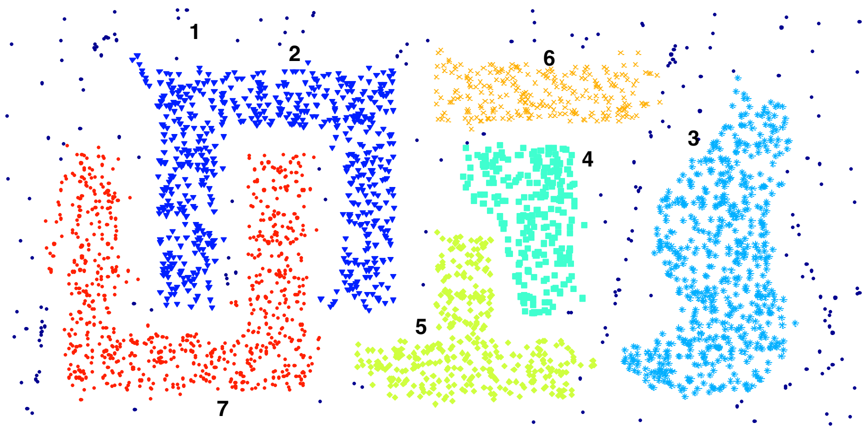
K = 10

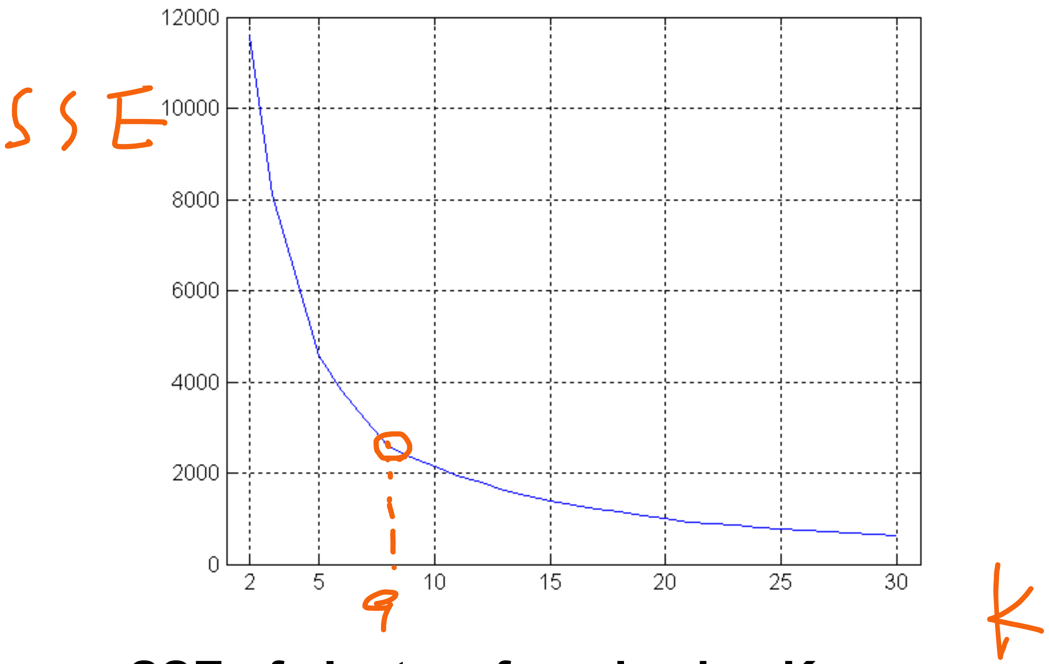




DETERMINING THE CORRECT NUMBER OF CLUSTERS

- SSE curve for a more complicated data set





SSE of clusters found using K-means

ASSESSING THE SIGNIFICANCE OF CLUSTER VALIDITY MEASURES

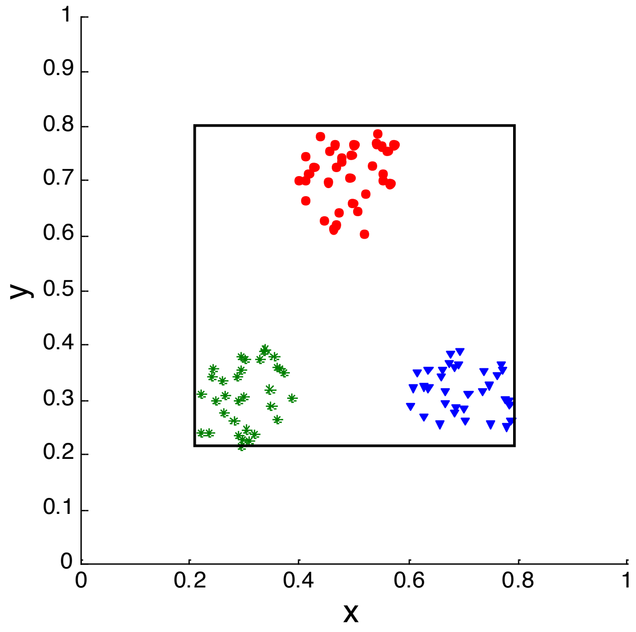
- Need a framework to interpret any measure.
  - For example, if our measure of evaluation has the value, 10, is that good, fair, or poor?
- Statistics provide a framework for cluster validity
  - The more "atypical" a clustering result is, the more likely it represents valid structure in the data
  - Compare the value of an index obtained from the given data with those resulting from random data.
    - If the value of the index is unlikely, then the cluster results are valid

① curve / base on the curr data

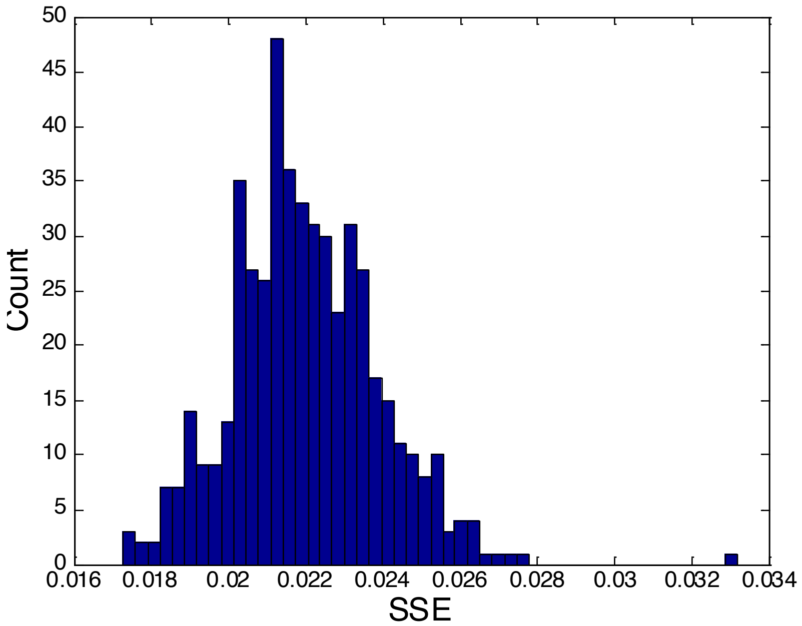
② → on several datasets

STATISTICAL FRAMEWORK FOR SSE

- Example
  - Compare SSE of three cohesive clusters against three clusters in random data

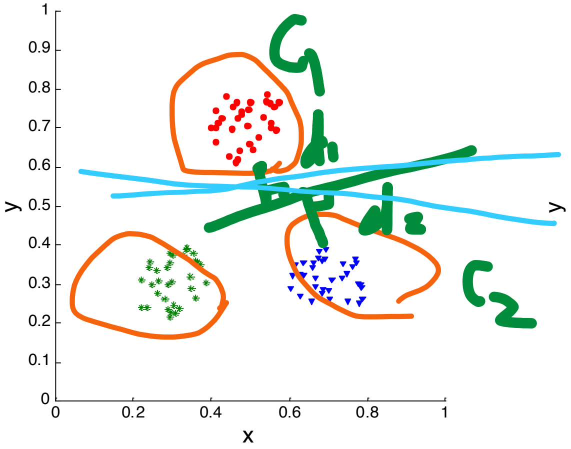


SSE = 0.005

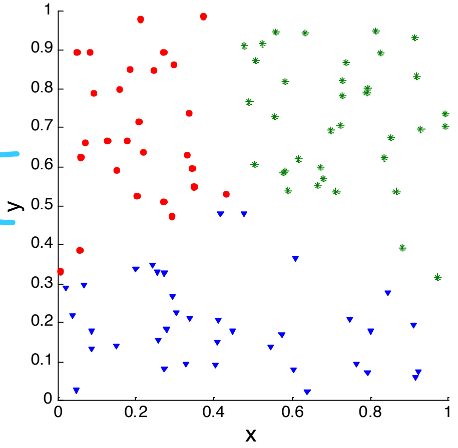


Histogram shows SSE of three clusters in 500 sets of random data points of size 100 distributed over the range 0.2 – 0.8 for x and y values

STATISTICAL FRAMEWORK FOR CORRELATION

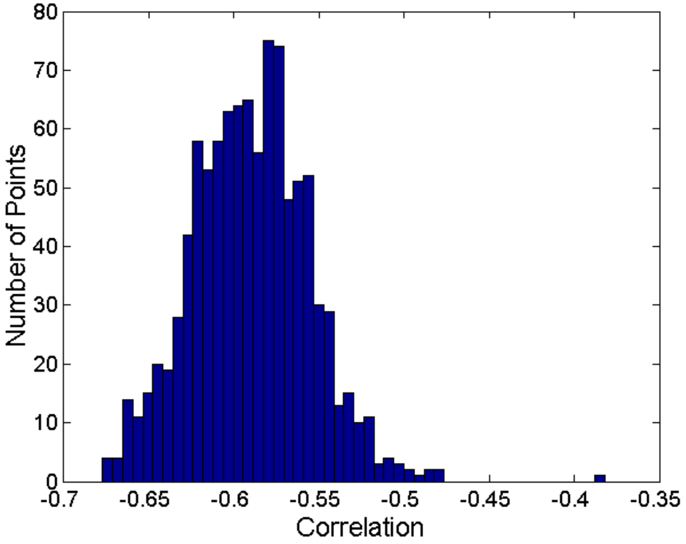


Corr = -0.9235



Corr = -0.5810

Correlation is negative because it is calculated between a distance matrix and the ideal similarity matrix. Higher magnitude is better.



Histogram of correlation for 500 random data sets of size 100 with *x* and *y* values of points between 0.2 and 0.8.

OTHER CLUSTER METHODS

1. Partitioning Methods K-mean
2. Hierarchical Methods
3. Density-Based Methods
4. Grid-Based Methods
5. Model-Based Methods
6. Clustering High-Dimensional Data
7. Constraint-Based Clustering
8. Outlier Analysis

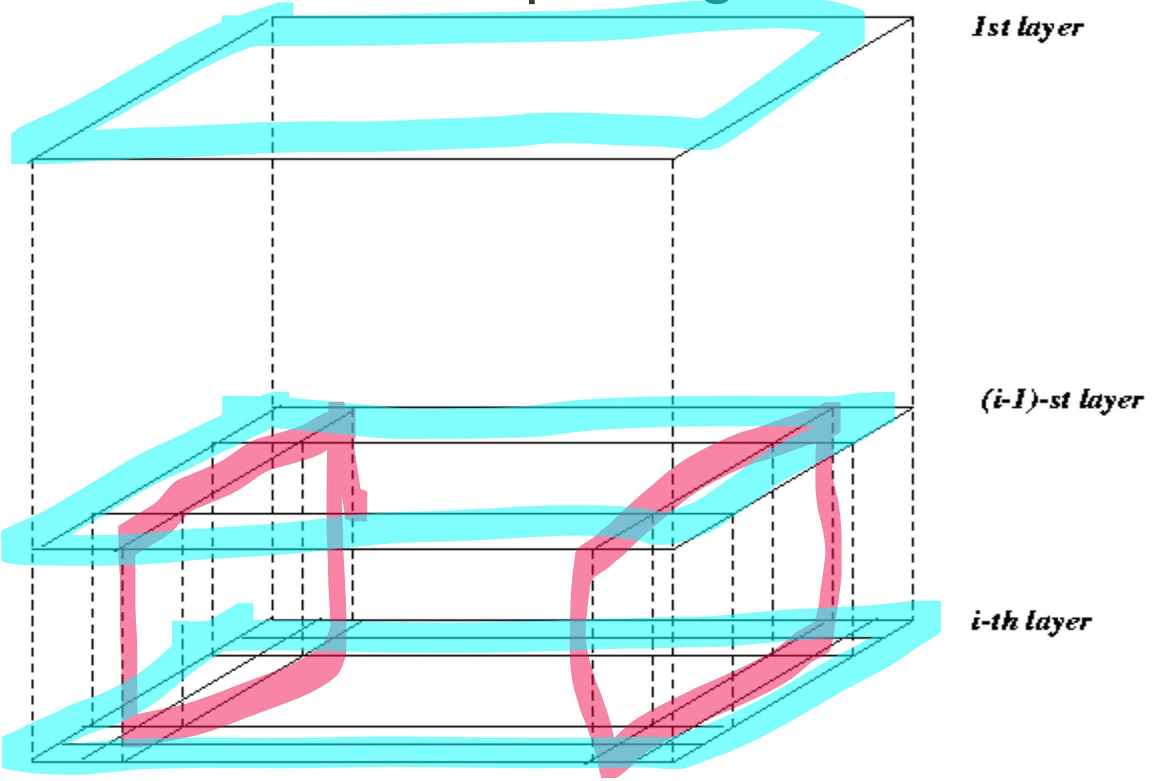
SUMMARY

Clustering Algorithm

Partitioning-Based	Hierarchical-Based	Density-Based	Grid-Based	Model-Based
1. K-means	1. BIRCH	1. DBSCAN	1. Wave-Cluster	1. EM
2. K-medoids	2. CURE	2. OPTICS	2. STING	2. COBWEB
3. K-modes	3. ROCK	3. DBCLASD	3. CLIQUE	3. CLASSIT
4. PAM	4. Chameleon	4. DENCLUE	4. OptiGrid	4. SOMs
5. CLARANS	5. Echidna			
6. CLARA				
7. FCM				

STING: A STATISTICAL INFORMATION GRID APPROACH

- Wang, Yang and Muntz (VLDB'97)
- The spatial area is divided into rectangular cells
- There are several levels of cells corresponding to different levels of resolution

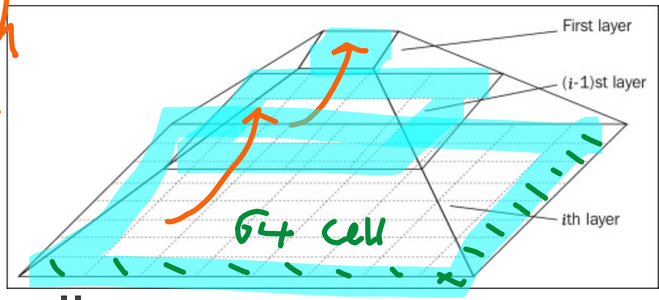


THE STING CLUSTERING METHOD

- Each cell at a high level is partitioned into a number of smaller cells in the next lower level
- Statistical info of each cell is calculated and stored beforehand and is used to answer queries
- Parameters of higher level cells can be easily calculated from parameters of lower level cell
  - *count*, *mean*, *s*, *min*, *max*
  - type of distribution—normal, *uniform*, etc.
- Use a top-down approach to answer spatial data queries

high

Low



- Start from a pre-selected layer—typically with a small number of cells
- For each cell in the current level compute the confidence interval

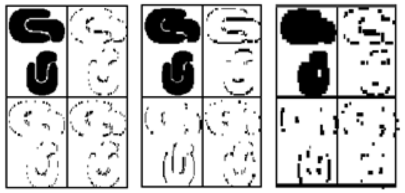
COMMENTS ON STING

- Remove the irrelevant cells from further consideration
- When finish examining the current layer, proceed to the next lower level
- Repeat this process until the bottom layer is reached
- Advantages:
  - Query-independent, easy to parallelize, incremental update
  - $O(K)$, where K is the number of grid cells at the lowest level
- Disadvantages:
  - All the cluster boundaries are either horizontal or vertical, and no diagonal boundary is detected

WAVE CLUSTER: CLUSTERING BY WAVELET ANALYSIS

- Sheikholeslami, Chatterjee, and Zhang
- A multi-resolution clustering approach which applies wavelet transform to the feature space
- How to apply wavelet transform to find clusters
  - Summarizes the data by imposing a multidimensional grid structure onto data space
  - These multidimensional spatial data objects are represented in a n-dimensional feature space
  - Apply wavelet transform on feature space to find the dense regions in the feature space
  - Apply wavelet transform multiple times which result in clusters at different scales from fine

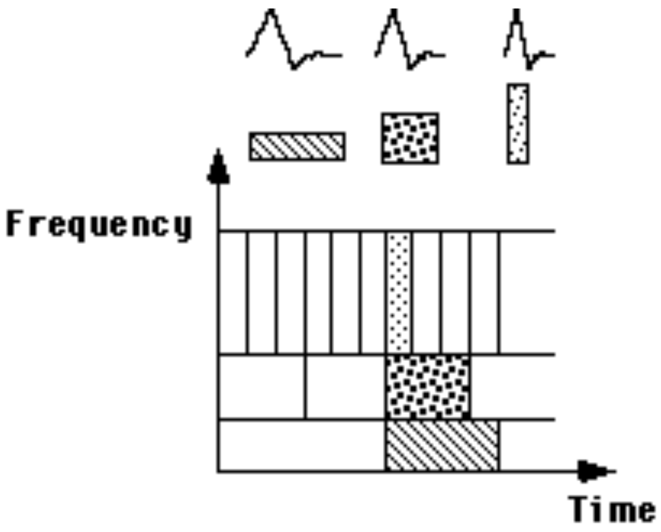
Figure 7.16 A sample of two-dimensional feature space. From [SCZ98].

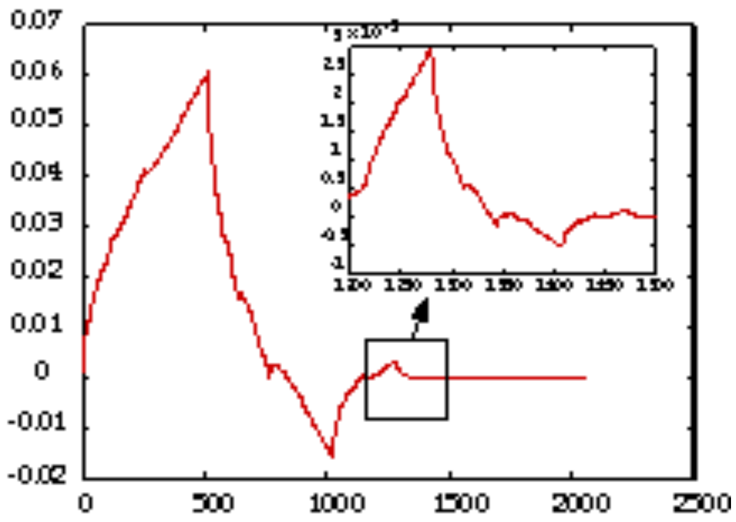


(a) (b) (c)

WAVELET TRANSFORM

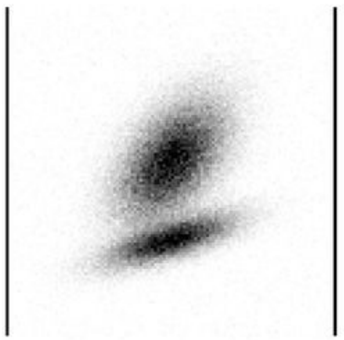
- Wavelet transform: A signal processing technique that decomposes a signal into different frequency sub-band (can be applied to n-dimensional signals)
- Data are transformed to preserve relative distance between objects at different levels of resolution
- Allows natural clusters to become more distinguishable



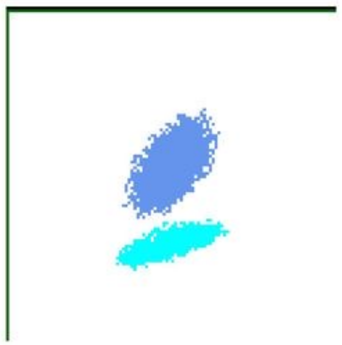


THE WAVECLUSTER ALGORITHM

- Input parameters
  - # of grid cells for each dimension
  - the wavelet, and the # of applications of wavelet transform
- Why is wavelet transformation useful for clustering?
  - Use hat-shape filters to emphasize region where points cluster, but simultaneously suppress weaker information in their boundary
  - Effective removal of outliers, multi-resolution, cost effective
- Major features:
  - Complexity O(N)
  - Detect arbitrary shaped clusters at different scales
  - Not sensitive to noise, not sensitive to input order
  - Only applicable to low dimensional data
- Both grid-based and density-based



a)



b)

QUANTIZATION & TRANSFORMATION

- First, quantize data into m-D grid structure, then wavelet transform
  - a) scale 1: high resolution
  - b) scale 2: medium resolution
  - c) scale 3: low resolution

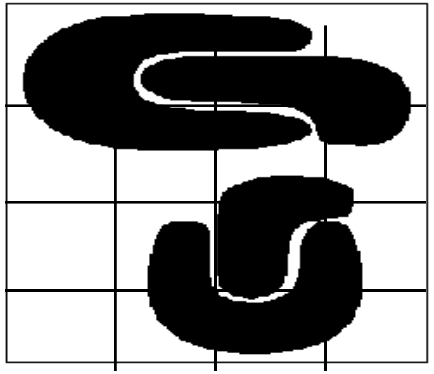


Figure 1: A sample 2-dimensional feature space.

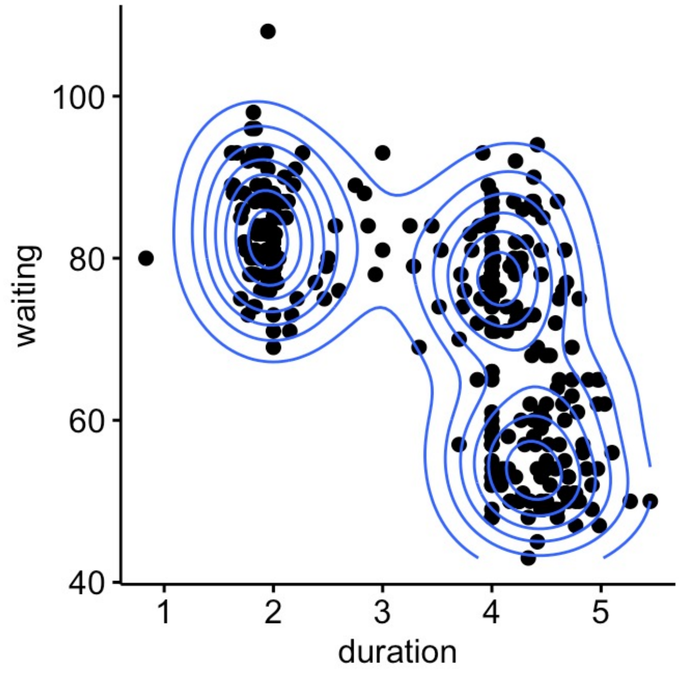
a) b) c)

OTHER CLUSTER METHODS

1. Partitioning Methods
2. Hierarchical Methods
3. Density-Based Methods
4. Grid-Based Methods (sting; wave)
5. Model-Based Methods
6. Clustering High-Dimensional Data
7. Constraint-Based Clustering
8. Outlier Analysis

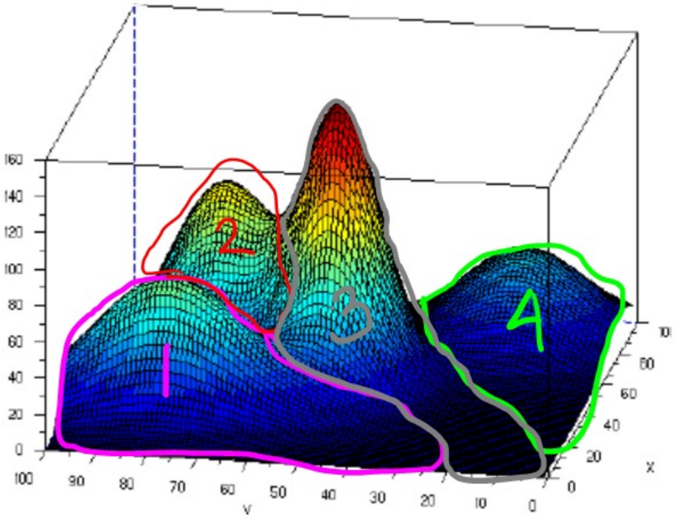
MODEL-BASED CLUSTERING

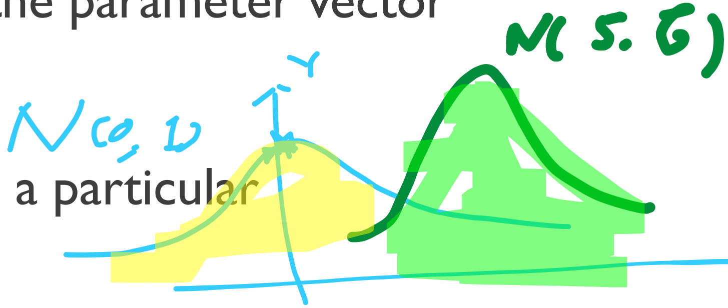
- What is model-based clustering?
  - Attempt to optimize the fit between the given data and some mathematical model
  - Based on the assumption: Data are generated by a mixture of underlying probability distribution
- Typical methods
  - Statistical approach
    - EM (Expectation maximization), AutoClass
  - Machine learning approach
    - COBWEB, CLASSIT
  - Neural network approach
    - SOM (Self-Organizing Feature Map)



EM — EXPECTATION MAXIMIZATION

- EM — A popular iterative refinement algorithm
- An extension to k-means
  - Assign each object to a cluster according to a weight (prob. distribution)
  - New means are computed based on weighted measures
- General idea
  - Starts with an initial estimate of the parameter vector
  - Iteratively rescores the patterns against the mixture density produced by the parameter vector
  - The rescored patterns are used to update the parameter updates
  - Patterns belonging to the same cluster, if they are placed by their scores in a particular component
- Algorithm converges fast but may not be in global optima





THE EM (EXPECTATION MAXIMIZATION) ALGORITHM

- Initially, randomly assign k cluster centers
- Iteratively refine the clusters based on two steps
  - Expectation step: assign each data point X_i to cluster C_i with the following probability

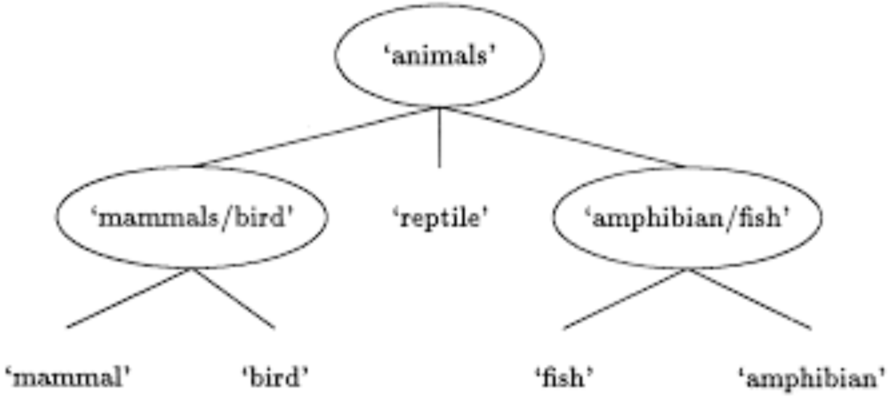
$$P(X_i \in C_k) = p(C_k|X_i) = \frac{p(C_k)p(X_i|C_k)}{p(X_i)},$$

  - Maximization step:
    - Estimation of model parameters

$$m_k = \frac{1}{N}\Sigma_{i=1}^{N}\frac{X_i P(X_i \in C_k)}{\Sigma_j P(X_i \in C_j)}.$$

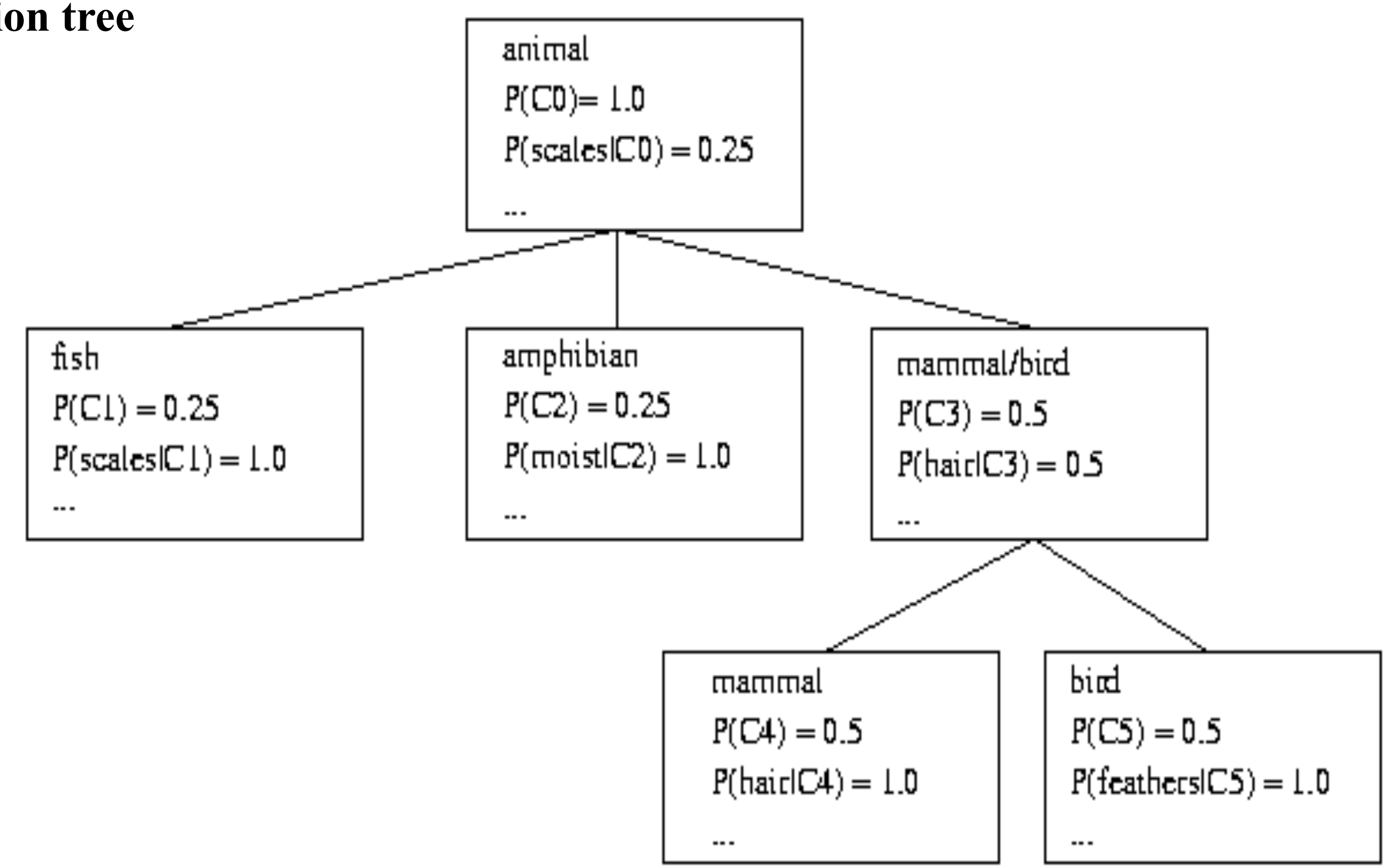
CONCEPTUAL CLUSTERING

- Conceptual clustering
  - A form of clustering in machine learning
  - Produces a classification scheme for a set of unlabeled objects
  - Finds characteristic description for each concept (class)
- COBWEB
  - A popular a simple method of incremental conceptual learning
  - Creates a hierarchical clustering in the form of a classification tree
  - Each node refers to a concept and contains a probabilistic description of that concept



COBWEB CLUSTERING METHOD

A classification tree



MORE ON CONCEPTUAL CLUSTERING

- Limitations of COBWEB
  - The assumption that the attributes are independent of each other is often too strong because correlation may exist
  - Not suitable for clustering large database data – skewed tree and expensive probability distributions
- CLASSIT
  - an extension of COBWEB for incremental clustering of continuous data
  - suffers similar problems as COBWEB
- AutoClass (Cheeseman and Stutz, 1996)
  - Uses Bayesian statistical analysis to estimate the number of clusters
  - Popular in industry

NEURAL NETWORK APPROACH

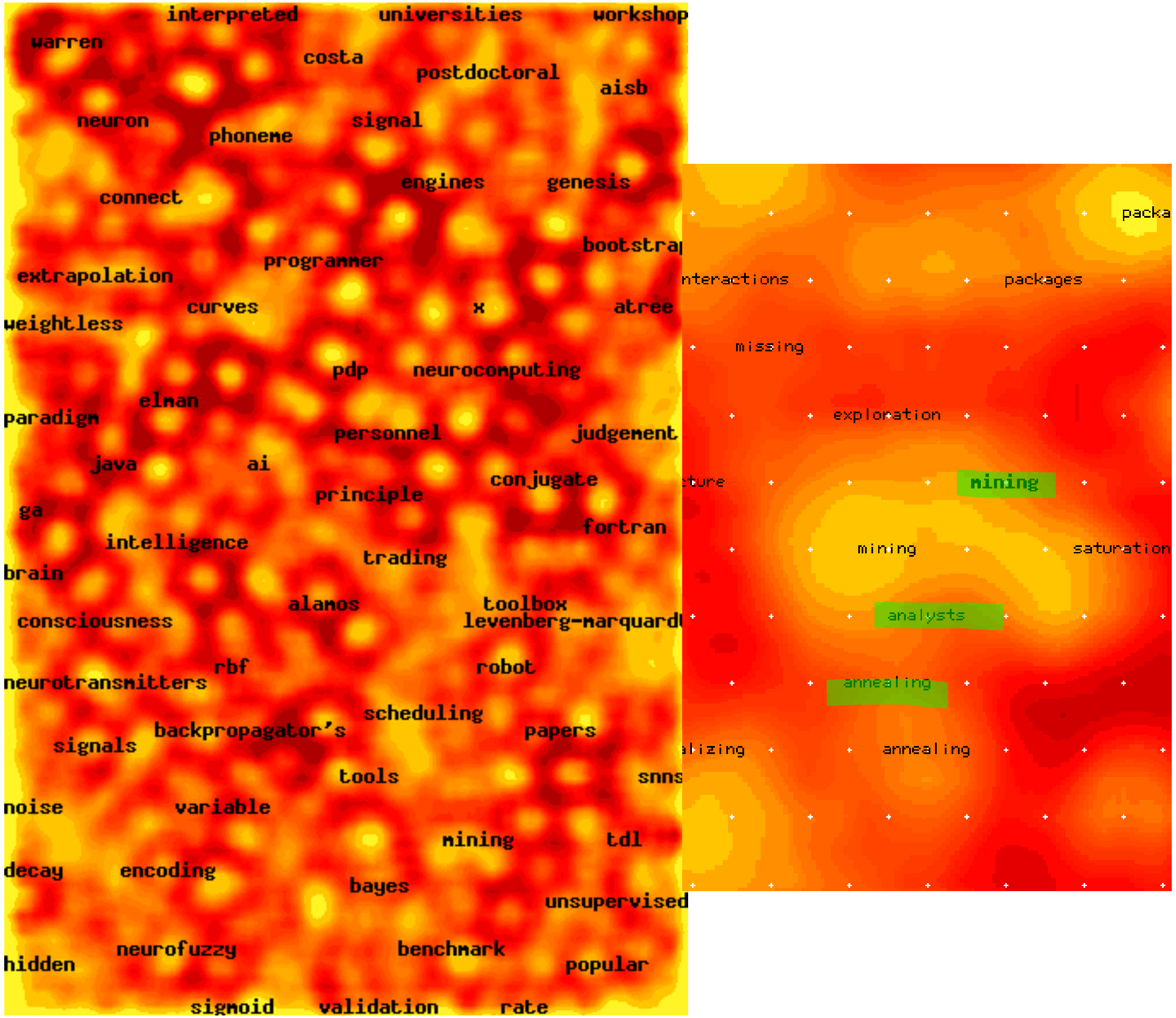
- Neural network approaches
  - Represent each cluster as an exemplar, acting as a "prototype" of the cluster
  - New objects are distributed to the cluster whose exemplar is the most similar according to some distance measure
- Typical methods
  - SOM (Soft-Organizing feature Map)
  - Competitive learning
    - Involves a hierarchical architecture of several units (neurons)
    - Neurons compete in a "winner-takes-all" fashion for the object currently being presented

SELF-ORGANIZING FEATURE MAP (SOM)

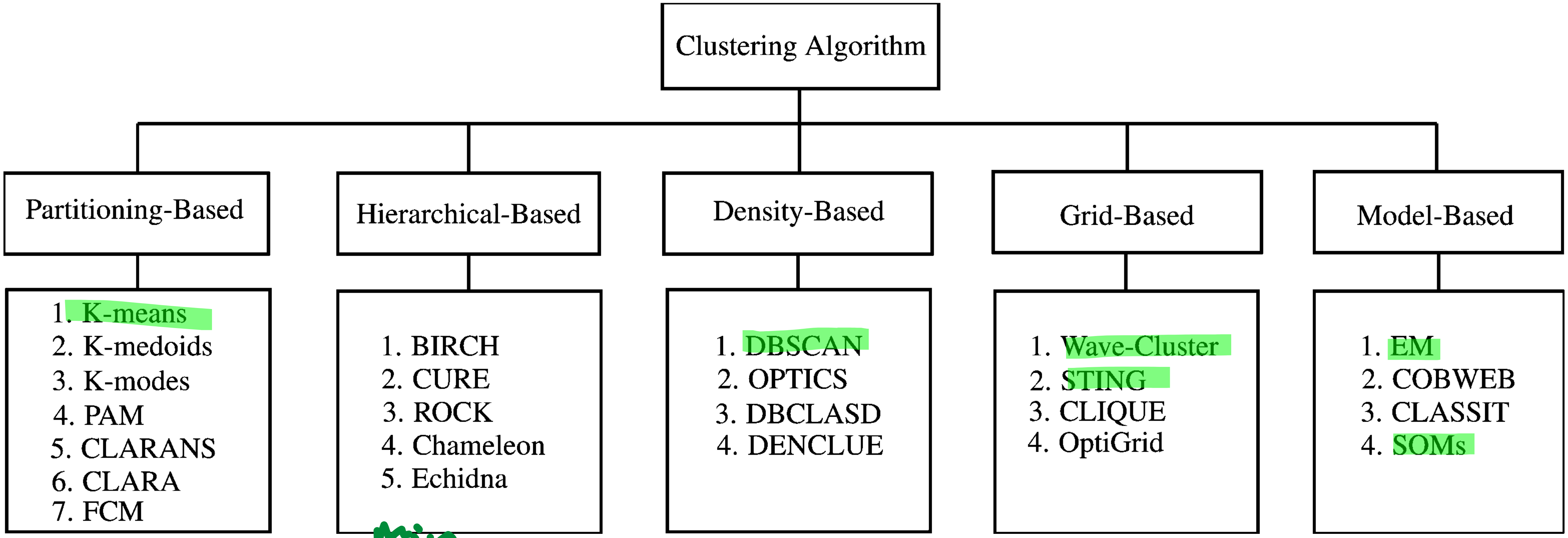
- SOMs, also called topological ordered maps, or Kohonen Self-Organizing Feature Map (KSOMs)
- It maps all the points in a high-dimensional source space into a 2 to 3-d target space, s.t., the distance and proximity relationship (i.e., topology) are preserved as much as possible
- Similar to k-means: cluster centers tend to lie in a low-dimensional manifold in the feature space
- Clustering is performed by having several units competing for the current object
  - The unit whose weight vector is closest to the current object wins
  - The winner and its neighbors learn by having their weights adjusted
- SOMs are believed to resemble processing that can occur in the brain
- Useful for visualizing high-dimensional data in 2- or 3-D space

WEB DOCUMENT CLUSTERING USING SOM

- The result of SOM clustering of 12088 Web articles
- The picture on the right: drilling down on the keyword "mining"
- Based on websom.hut.fi Web page



SUMMARY



min
max

CHAPTER 6. CLUSTER ANALYSIS

1. Partitioning Methods
2. Hierarchical Methods
3. Density-Based Methods
4. Grid-Based Methods
5. Model-Based Methods
6. Clustering High-Dimensional Data
7. Constraint-Based Clustering
8. Outlier Analysis

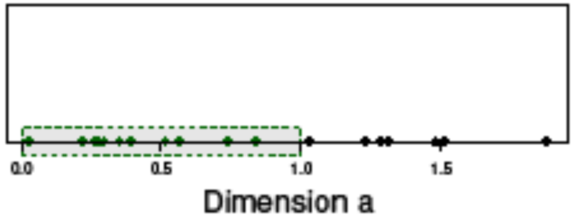
CLUSTERING HIGH-DIMENSIONAL DATA

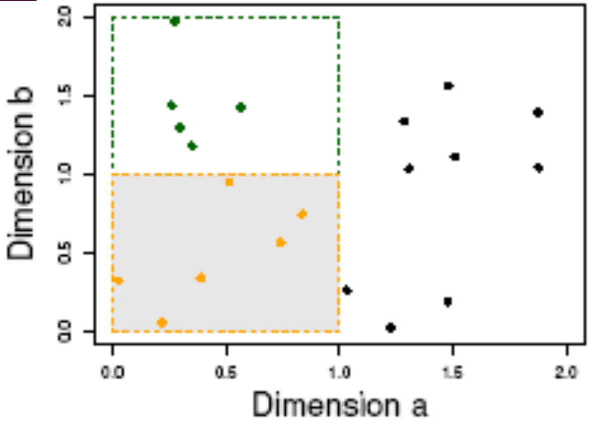
- Clustering high-dimensional data
  - Many applications: text documents, DNA micro-array data
  - Major challenges:
    - Many irrelevant dimensions may mask clusters
    - Distance measure becomes meaningless—due to equi-distance
    - Clusters may exist only in some subspaces
- Methods
  - Feature transformation: only effective if most dimensions are relevant
    - PCA & SVD useful only when features are highly correlated/redundant
  - Feature selection: wrapper or filter approaches
    - useful to find a subspace where the data have nice clusters
  - Subspace-clustering: find clusters in all the possible subspaces
    - CLIQUE, ProClus, and frequent pattern-based clustering

THE CURSE OF DIMENSIONALITY

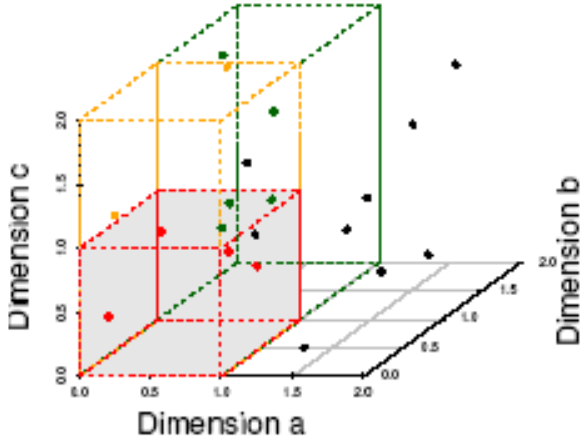
(GRAPHS ADAPTED FROM PARSONS ET AL. KDD EXPLORATIONS 200

- Data in only one dimension is relatively packed
- Adding a dimension “stretch” the points across that dimension, making them further apart
- Adding more dimensions will make the points further apart—high dimensional data is extremely sparse
- Distance measure becomes meaningless—due to equi-distance





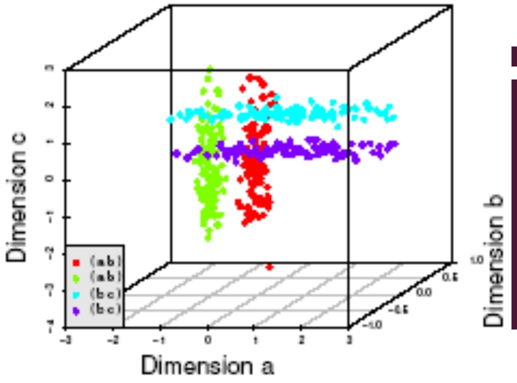
(b) 6 Objects in One Unit Bin



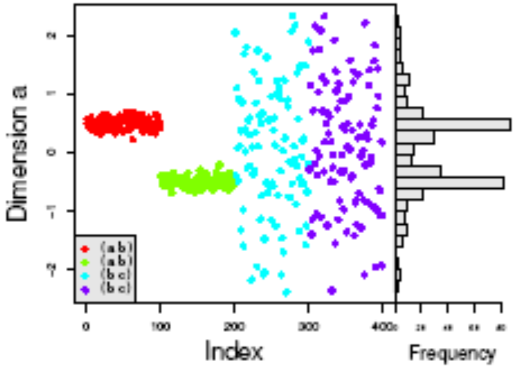
(c) 4 Objects in One Unit Bin

WHY SUBSPACE CLUSTERING?

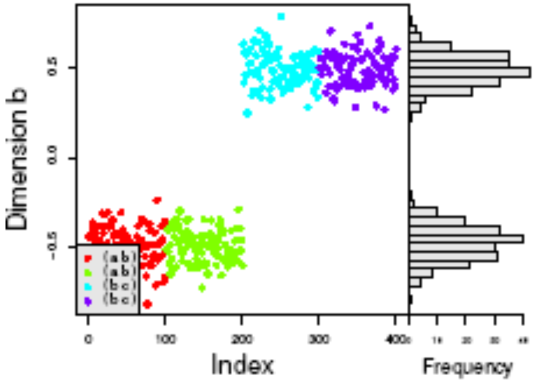
(ADAPTED FROM PARSONS ET AL. SIGKDD EXPLORATIONS 2004)



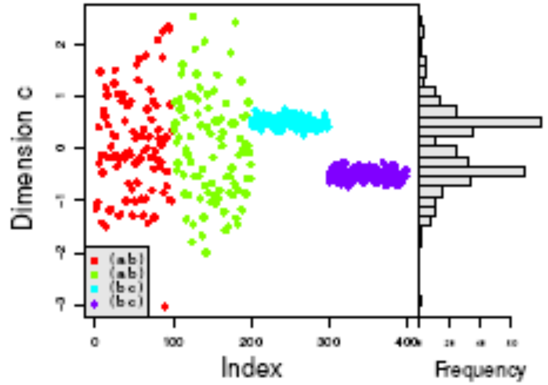
- Clusters may exist only in some subspaces
- Subspace-clustering: find clusters in all the subspaces



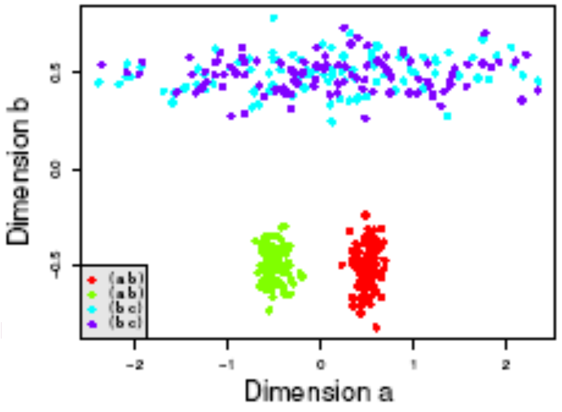
(a) Dimension a



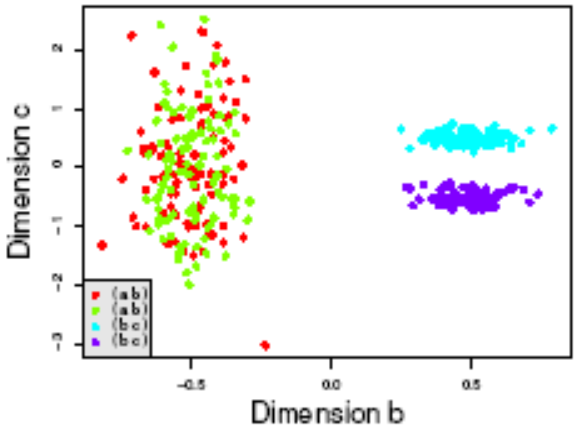
(b) Dimension b



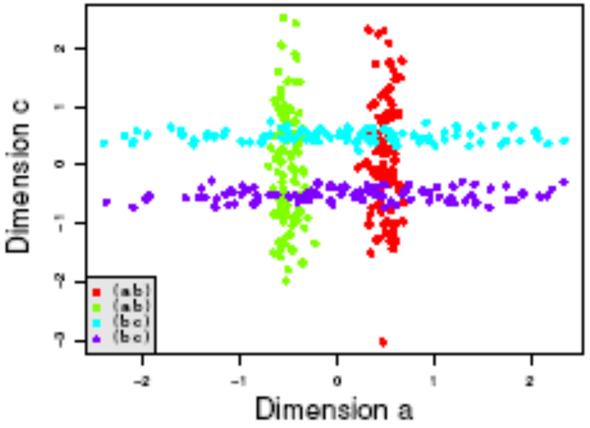
(c) Dimension c



(a) Dims a & b



(b) Dims b & c



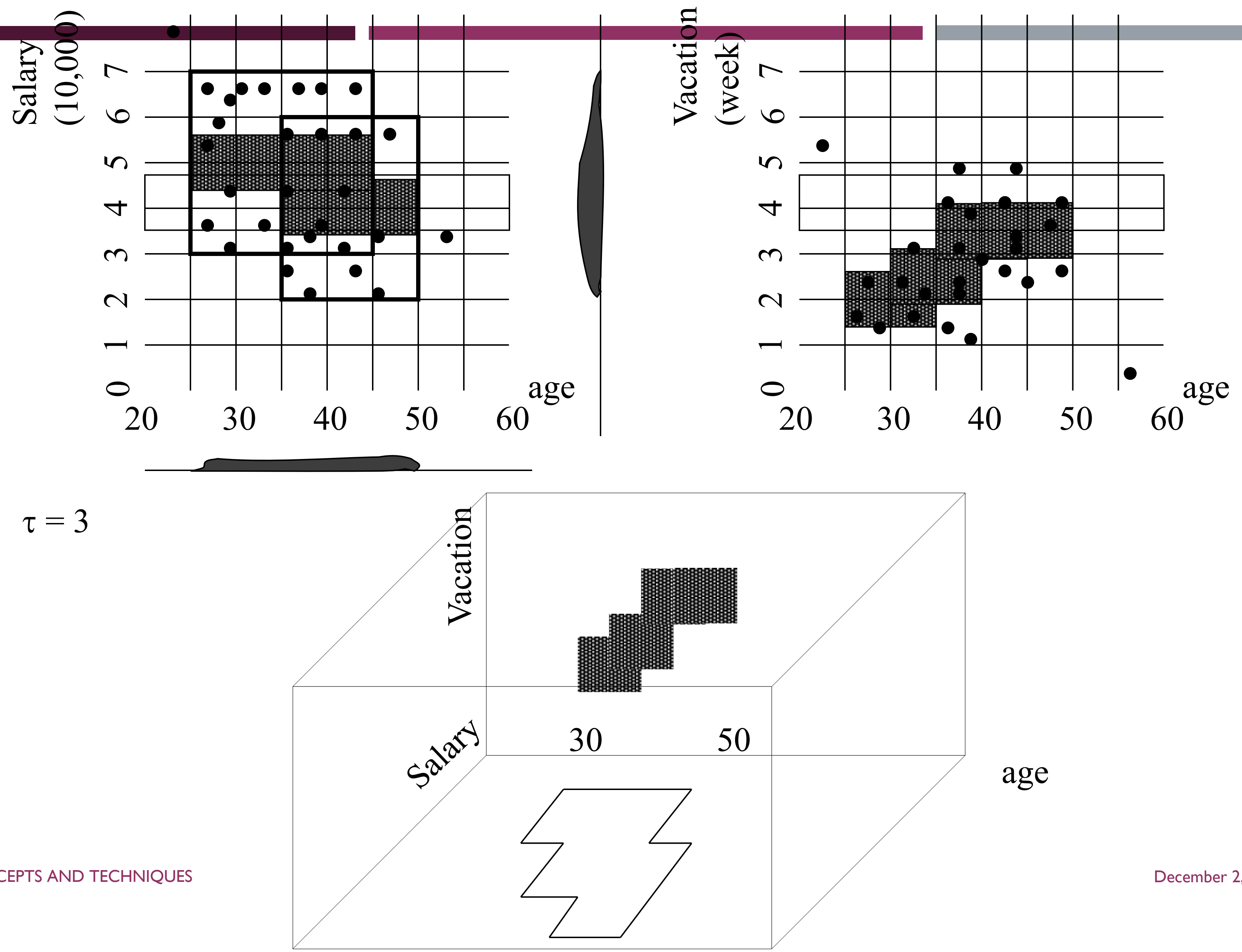
(c) Dims a & c

CLIQUE (CLUSTERING IN QUEST)

- Agrawal, Gehrke, Gunopulos, Raghavan (SIGMOD'98)
- Automatically identifying subspaces of a high dimensional data space that allow better clustering than original space
- CLIQUE can be considered as both density-based and grid-based
  - It partitions each dimension into the same number of equal length interval
  - It partitions an m-dimensional data space into non-overlapping rectangular units
  - A unit is dense if the fraction of total data points contained in the unit exceeds the input model parameter
  - A cluster is a maximal set of connected dense units within a subspace

CLIQUE: THE MAJOR STEPS

- Partition the data space and find the number of points that lie inside each cell of the partition.
- Identify the subspaces that contain clusters using the Apriori principle
- Identify clusters
  - Determine dense units in all subspaces of interests
  - Determine connected dense units in all subspaces of interests.
- Generate minimal description for the clusters
  - Determine maximal regions that cover a cluster of connected dense units for each cluster
  - Determination of minimal cover for each cluster


Salary (10,000)
0
1
2
3
4
5
6
7
20
30
40
50
60
age
Vacation (week)
0
1
2
3
4
5
6
7
20
30
40
50
60
age
Vacation
Salary
30
50
age

$\tau = 3$

STRENGTH AND WEAKNESS OF *CLIQUE*

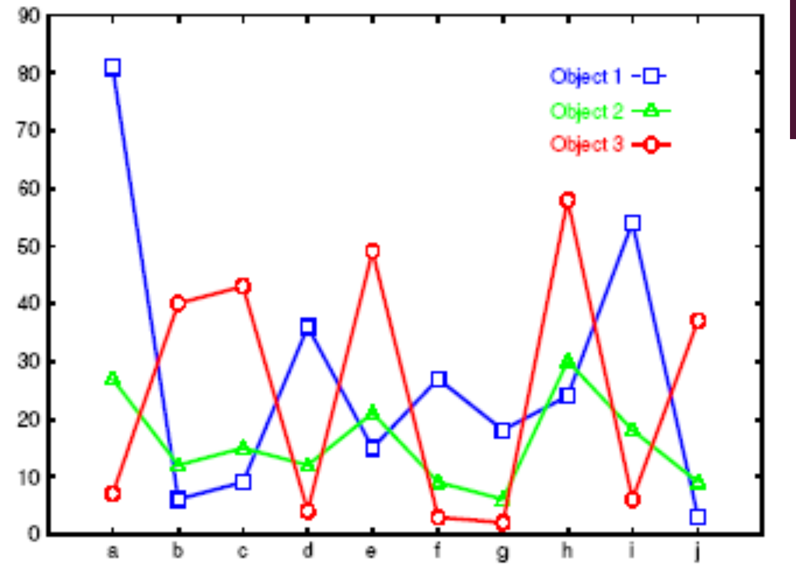
- Strength
  - *automatically* finds subspaces of the highest dimensionality such that high density clusters exist in those subspaces
  - *insensitive* to the order of records in input and does not presume some canonical data distribution
  - scales *linearly* with the size of input and has good scalability as the number of dimensions in the data increases
- Weakness
  - The accuracy of the clustering result may be degraded at the expense of simplicity of the method

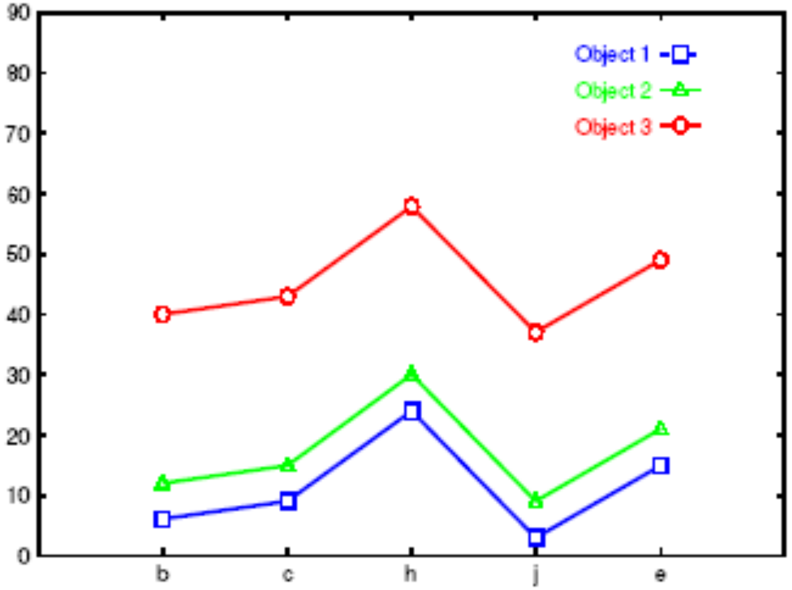
FREQUENT PATTERN-BASED APPROACH

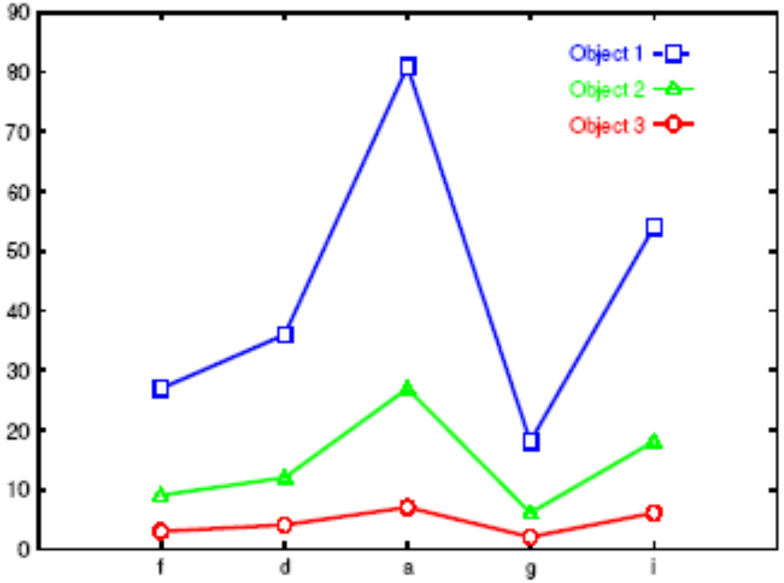
- Clustering high-dimensional space (e.g., clustering text documents, microarray data)
  - Projected subspace-clustering: which dimensions to be projected on?
    - CLIQUE, ProClus
  - Feature extraction: costly and may not be effective?
  - Using frequent patterns as "features"
    - "Frequent" are inherent features
    - Mining freq. patterns may not be so expensive
- Typical methods
  - Frequent-term-based document clustering
  - Clustering by pattern similarity in micro-array data (pClustering)

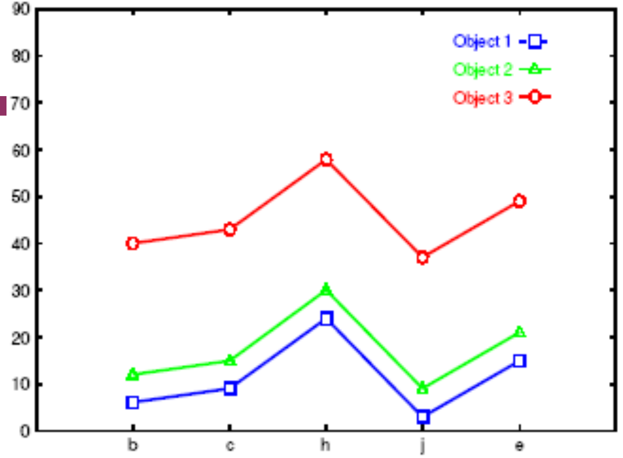
CLUSTERING)

- Right: The micro-array “raw” data shows 3 genes and their values in a multi-dimensional space
  - Difficult to find their patterns
- Bottom: Some subsets of dimensions form nice shift and scaling patterns







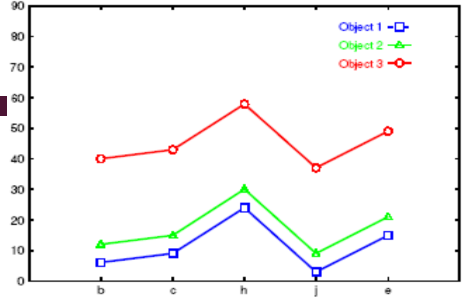


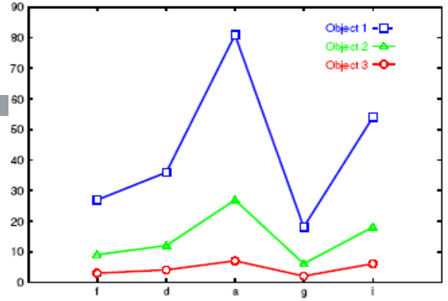
- Microarray data analysis may need to
  - Clustering on thousands of dimensions (attributes)
  - Discovery of both shift and scaling patterns
- Clustering with Euclidean distance measure? — cannot find shift patterns
- Clustering on derived attribute $A_{ij} = a_i - a_j$? — introduces N(N-1) dimensions
- Bi-cluster using transformed mean-squared residue score matrix (I, J)

$$H(IJ) = \frac{1}{|I||J|}\Sigma_{i\in I, j\in J}(d_{ij} - d_{iJ} - d_{Ij} + d_{IJ})^2$$

  - Where $d_{ij} = \frac{1}{|J|}\sum_{j \in J} d_{ij}$ $\quad d_{Ij} = \frac{1}{|I|}\sum_{i \in I} d_{ij}$ $\quad d_{IJ} = \frac{1}{|I||J|}\sum_{i \in I, j \in J} d_{ij}$
  - A submatrix is a δ-cluster if H(I, J) ≤ δ for some δ > 0
- Problems with bi-cluster
  - No downward closure property,
  - Due to averaging, it may contain outliers but still within δ-threshold

CLUSTERING BY





- Given object x, y in O and features a, b in T, pCluster is a 2 by 2 matrix

$$pScore\left(\begin{bmatrix} d_{xa} & d_{xb} \\ d_{ya} & d_{yb} \end{bmatrix}\right) = |(d_{xa} - d_{xb}) - (d_{ya} - d_{yb})|$$

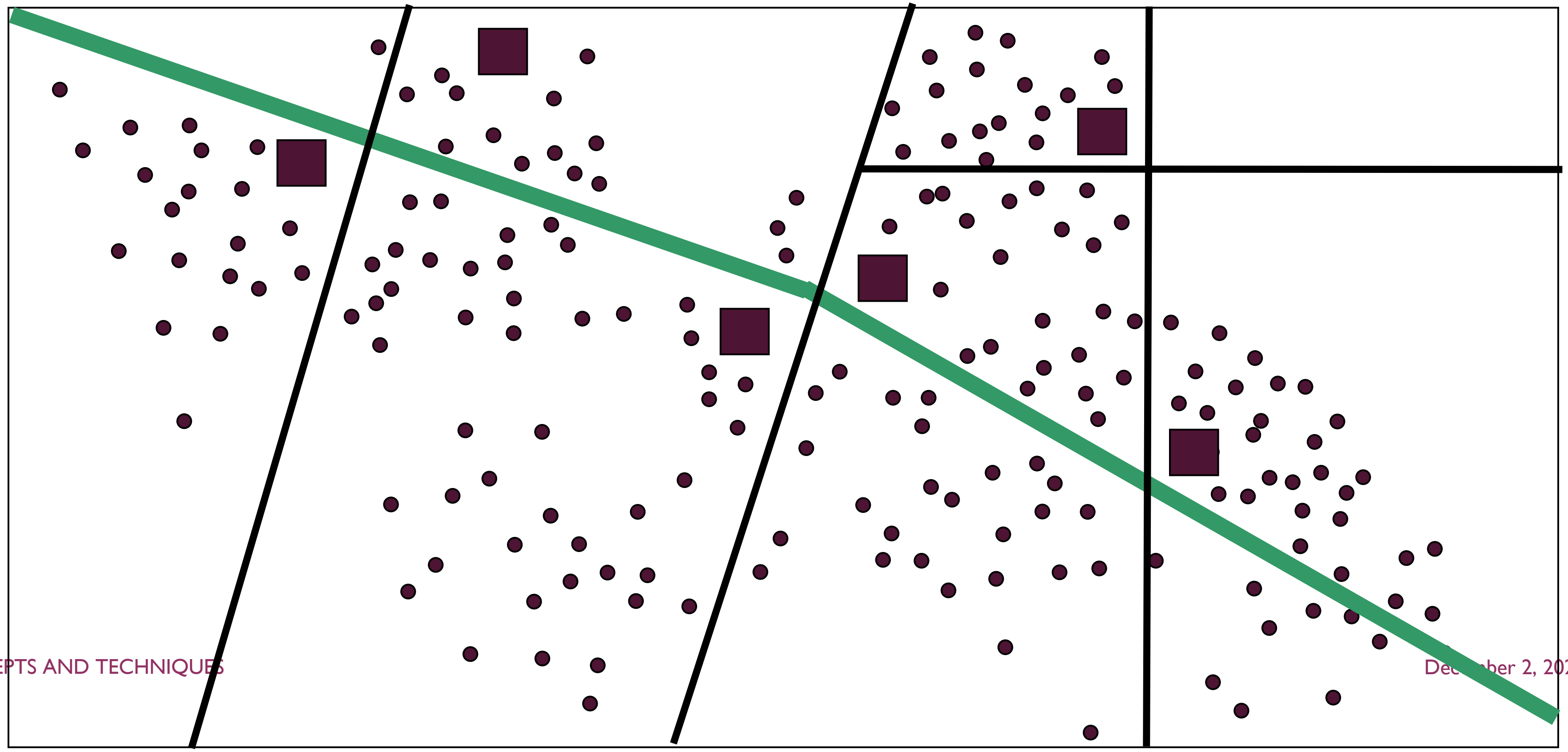
- A pair (O,T) is in δ-pCluster if for any 2 by 2 matrix X in (O,T), pScore(X) ≤ δ for some δ > 0
- Properties of δ-pCluster
  - Downward closure
  - Clusters are more homogeneous than bi-cluster (thus the name: pair-wise Cluster)
- Pattern-growth algorithm has been developed for efficient mining
- For scaling patterns, one can observe, taking logarithmic on $\frac{d_{xa}/d_{ya}}{d_{xb}/d_{yb}} < \delta$ will lead to the pScore form

CHAPTER 6. CLUSTER ANALYSIS

1. What is Cluster Analysis?
2. Types of Data in Cluster Analysis
3. A Categorization of Major Clustering Methods
4. Partitioning Methods
5. Hierarchical Methods
6. Density-Based Methods
7. Grid-Based Methods
8. Model-Based Methods
9. Clustering High-Dimensional Data
10. Constraint-Based Clustering
11. Outlier Analysis
12. Summary

WHY CONSTRAINT-BASED CLUSTER ANALYSIS?

- Need user feedback: Users know their applications the best
- Less parameters but more user-desired constraints, e.g., an ATM allocation problem: obstacle & desired clusters

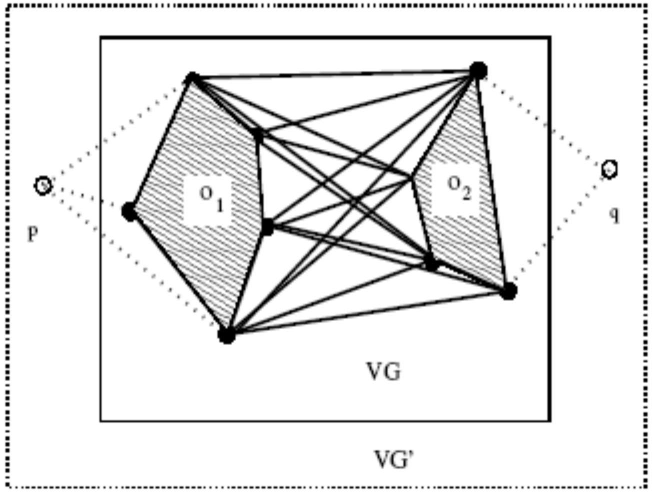


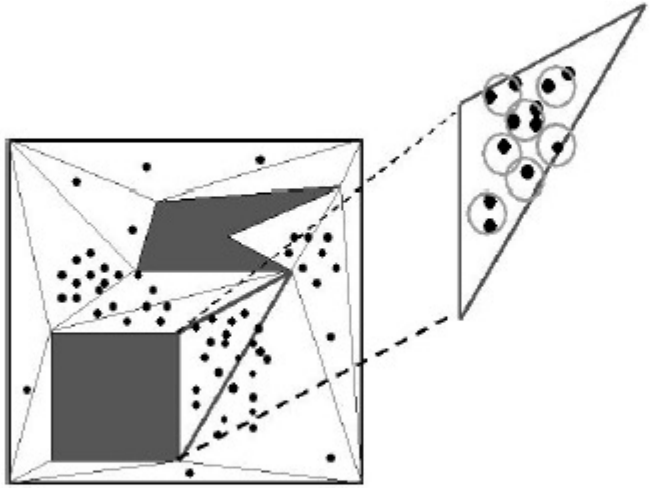
ANALYSIS

- Clustering in applications: desirable to have user-guided (i.e., constrained) cluster analysis
- Different constraints in cluster analysis:
  - Constraints on individual objects (do selection first)
    - Cluster on houses worth over $300K
  - Constraints on distance or similarity functions
    - Weighted functions, obstacles (e.g., rivers, lakes)
  - Constraints on the selection of clustering parameters
    - # of clusters, MinPts, etc.
  - User-specified constraints
    - Contain at least 500 valued customers and 5000 ordinary ones
  - Semi-supervised: giving small training sets as "constraints" or hints

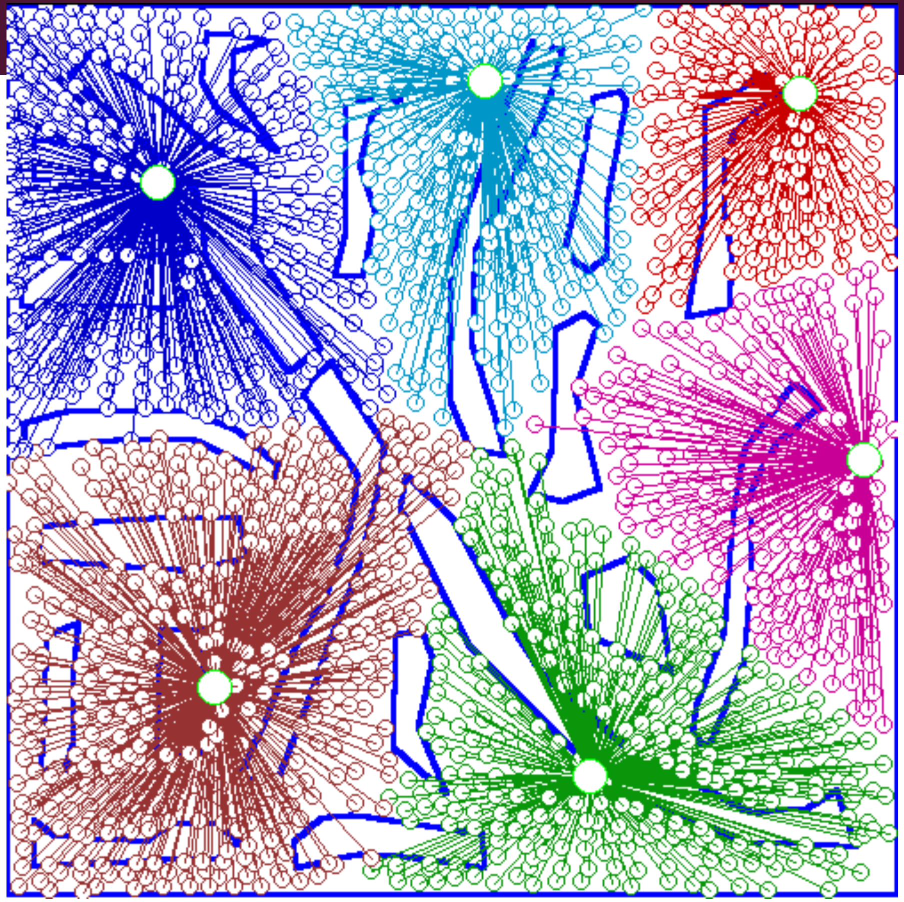
CLUSTERING WITH OBSTACLE OBJECTS

- K-medoids is more preferable since k-means may locate the ATM center in the middle of a lake
- Visibility graph and shortest path
- Triangulation and micro-clustering
- Two kinds of join indices (shortest-paths) worth pre-computation
  - VV index: indices for any pair of obstacle vertices
  - MV index: indices for any pair of micro-cluster and obstacle indices

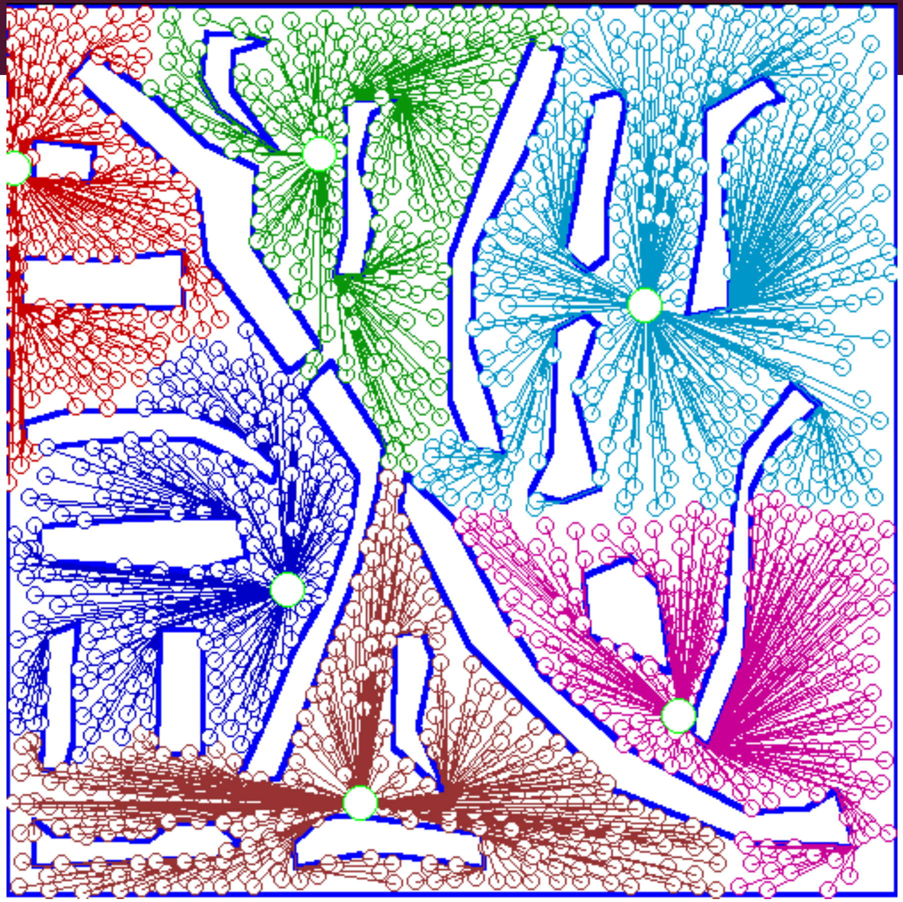




OBJECTS



Not Taking obstacles into account



Taking obstacles into account

CLUSTERING WITH USER-SPECIFIED CONSTRAINTS

- Example: Locating k delivery centers, each serving at least m valued customers and n ordinary ones
- Proposed approach
  - Find an initial "solution" by partitioning the data set into k groups and satisfying user-constraints
  - Iteratively refine the solution by micro-clustering relocation (e.g., moving δ μ-clusters from cluster C_i to C_j) and "deadlock" handling (break the microclusters when necessary)
  - Efficiency is improved by micro-clustering
- How to handle more complicated constraints?
  - E.g., having approximately same number of valued customers in each cluster?! — Can you solve it?

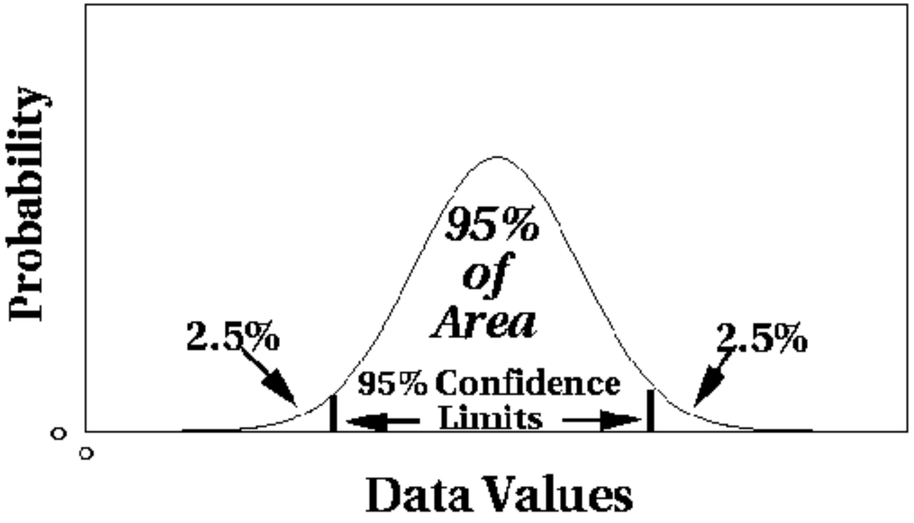
CHAPTER 7. CLUSTER ANALYSIS

1. What is Cluster Analysis?
2. Types of Data in Cluster Analysis
3. A Categorization of Major Clustering Methods
4. Partitioning Methods
5. Hierarchical Methods
6. Density-Based Methods
7. Grid-Based Methods
8. Model-Based Methods
9. Clustering High-Dimensional Data
10. Constraint-Based Clustering
11. Outlier Analysis
12. Summary

WHAT IS OUTLIER DISCOVERY?

- What are outliers?
  - The set of objects are considerably dissimilar from the remainder of the data
  - Example: Sports: Michael Jordon, Wayne Gretzky, ...
- Problem: Define and find outliers in large data sets
- Applications:
  - Credit card fraud detection
  - Telecom fraud detection
  - Customer segmentation
  - Medical analysis

OUTLIER DISCOVERY: STATISTICAL APPROACHES



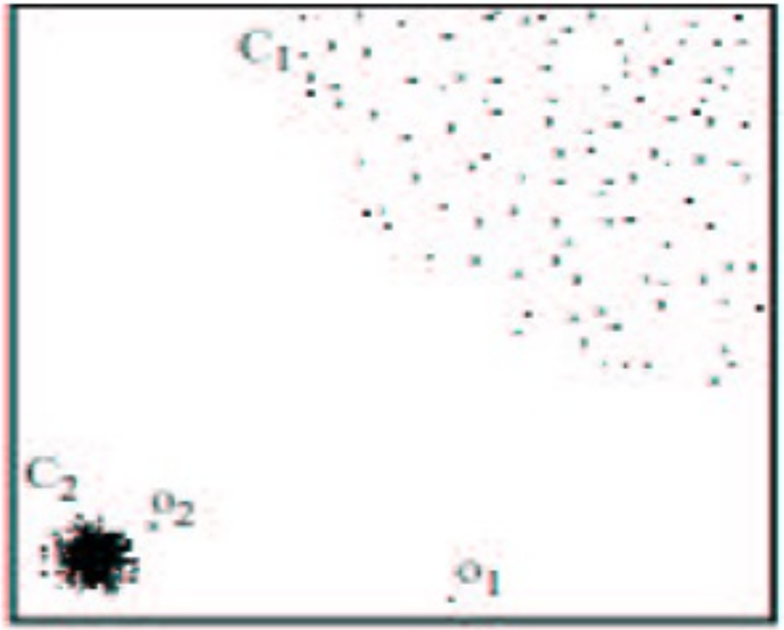
- Assume a model underlying distribution that generates data set (e.g. normal distribution)
- Use discordancy tests depending on
  - data distribution
  - distribution parameter (e.g., mean, variance)
  - number of expected outliers
- Drawbacks
  - most tests are for single attribute
  - In many cases, data distribution may not be known

APPROACH

- Introduced to counter the main limitations imposed by statistical methods
  - We need multi-dimensional analysis without knowing data distribution
- Distance-based outlier: A DB(p, D)-outlier is an object O in a dataset T such that at least a fraction p of the objects in T lies at a distance greater than D from O
- Algorithms for mining distance-based outliers
  - Index-based algorithm
  - Nested-loop algorithm
  - Cell-based algorithm

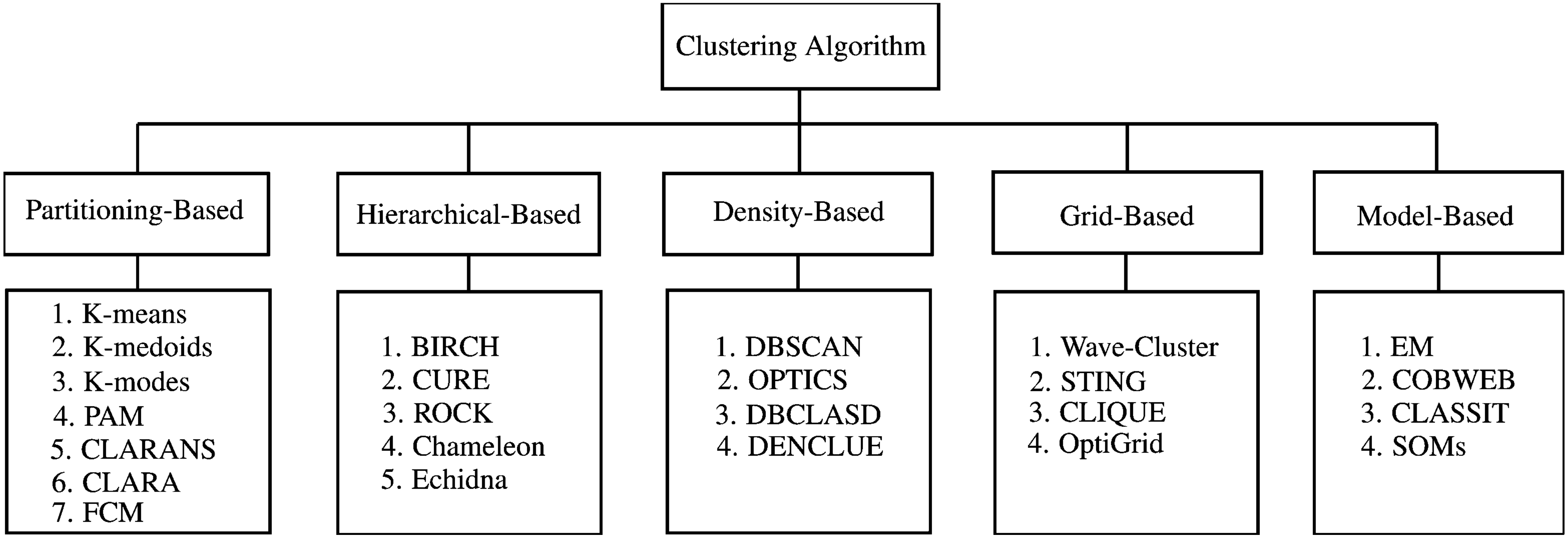
DENSITY-BASED LOCAL OUTLIER DETECTION

- Distance-based outlier detection is based on global distance distribution
- It encounters difficulties to identify outliers if data is not uniformly distributed
- Ex. C_1 contains 400 loosely distributed points, C_2 has 100 tightly condensed points, 2 outlier points o_1, o_2
- Distance-based method cannot identify o_2 as an outlier
- Need the concept of local outlier



- Local outlier factor (LOF)
  - Assume outlier is not crisp
  - Each point has a LOF

SUMMARY



Clustering Algorithm
Partitioning-Based
Hierarchical-Based
Density-Based
Grid-Based
Model-Based
1. K-means
2. K-medoids
3. K-modes
4. PAM
5. CLARANS
6. CLARA
7. FCM
1. BIRCH
2. CURE
3. ROCK
4. Chameleon
5. Echidna
1. DBSCAN
2. OPTICS
3. DBCLASD
4. DENCLUE
1. Wave-Cluster
2. STING
3. CLIQUE
4. OptiGrid
1. EM
2. COBWEB
3. CLASSIT
4. SOMs

APPROACH

- Identifies outliers by examining the main characteristics of objects in a group
- Objects that "deviate" from this description are considered outliers
- Sequential exception technique
  - simulates the way in which humans can distinguish unusual objects from among a series of supposedly like objects
- OLAP data cube technique
  - uses data cubes to identify regions of anomalies in large multidimensional data