Outline

1. Introduction
2. Axis-parallel Subspace Clustering
3. Pattern-based Clustering
4. Arbitrarily-oriented Subspace Clustering
5. Summary

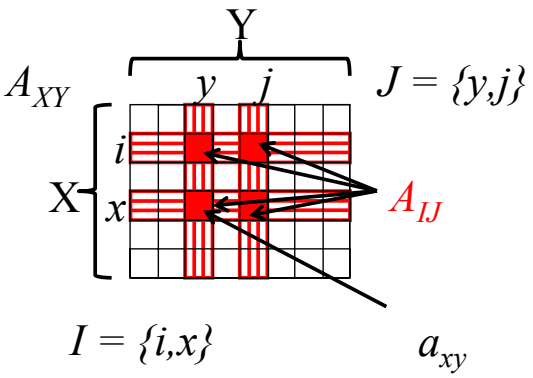
Outline: Pattern-based Clustering

- Challenges and Approaches, Basic Models for
  - Constant Biclusters
  - Biclusters with Constant Values in Rows or Columns
  - Pattern-based Clustering: Biclusters with Coherent Values
  - Biclusters with Coherent Evolutions
- Algorithms for
  - Constant Biclusters
  - Pattern-based Clustering: Biclusters with Coherent Values
- Summary

Challenges and Approaches, Basic Models

Pattern-based clustering relies on patterns in the data matrix.

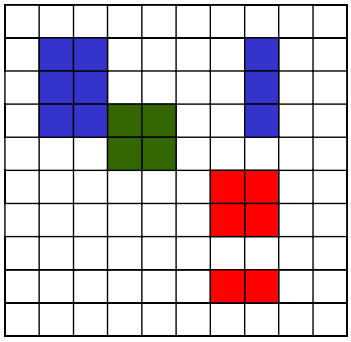
- Simultaneous clustering of rows and columns of the data matrix (hence *bi*clustering).
  - Data matrix $\boldsymbol{A}$ = (X,Y) with set of rows X and set of columns Y
  - a_{xy} is the element in row x and column y.
  - submatrix A_{IJ} = (I,J) with subset of rows I ⊆ X and subset of columns J ⊆ Y contains those elements a_{ij} with $i \in$ I und $j \in$ J



Challenges and Approaches, Basic Models

General aim of biclustering approaches:

Find a set of submatrices $\{(I_1,J_1),(I_2,J_2),...,(I_k,J_k)\}$ of the matrix $\boldsymbol{A}$=(X,Y) (with $I_i \subseteq$ X and $J_i \subseteq$ Y for i = 1,...,k) where each submatrix (= bicluster) meets a given homogeneity criterion.



Challenges and Approaches, Basic Models

- Some values often used by bicluster models:
  - mean of row *i*:

$$a_{iJ} = \frac{1}{|J|}\sum_{j \in J} a_{ij}$$

  - mean of column *j*:

$$a_{Ij} = \frac{1}{|I|}\sum_{i \in I} a_{ij}$$

  - mean of all elements:

$$a_{IJ} = \frac{1}{|I||J|}\sum_{i \in I, j \in J} a_{ij}$$

$$= \frac{1}{|J|}\sum_{j \in J} a_{Ij}$$

$$= \frac{1}{|I|}\sum_{i \in I} a_{iJ}$$

Challenges and Approaches, Basic Models

Different types of biclusters (cf. [MO04]):

- constant biclusters
- biclusters with
  - constant values on columns
  - constant values on rows
- biclusters with coherent values (aka. pattern-based clustering)
- biclusters with coherent evolutions

Challenges and Approaches, Basic Models

Constant biclusters

- all points share identical value in selected attributes.
- The constant value μ is a typical value for the cluster.
- Cluster model:

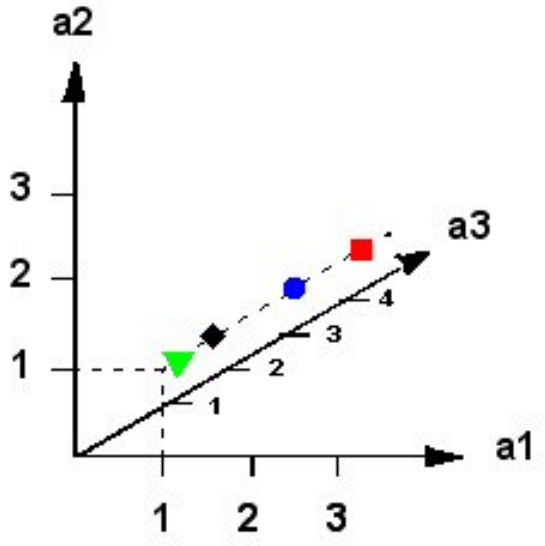
$$a_{ij} = \mu$$

- Obviously a special case of an axis-parallel subspace cluster.

Challenges and Approaches, Basic Models

- example – embedding 3-dimensional space:

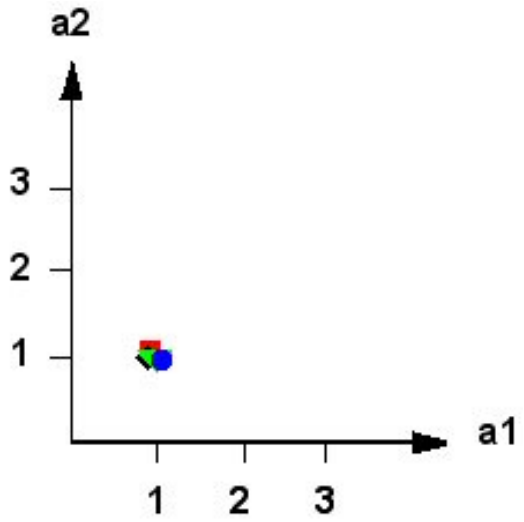
	a1	a2	a3
P1	1	1	3.5
P2	1	1	2.3
P3	1	1	0.2
P4	1	1	0.7



Challenges and Approaches, Basic Models

- example – 2-dimensional subspace:

	a1	a2
P1	1	1
P2	1	1
P3	1	1
P4	1	1

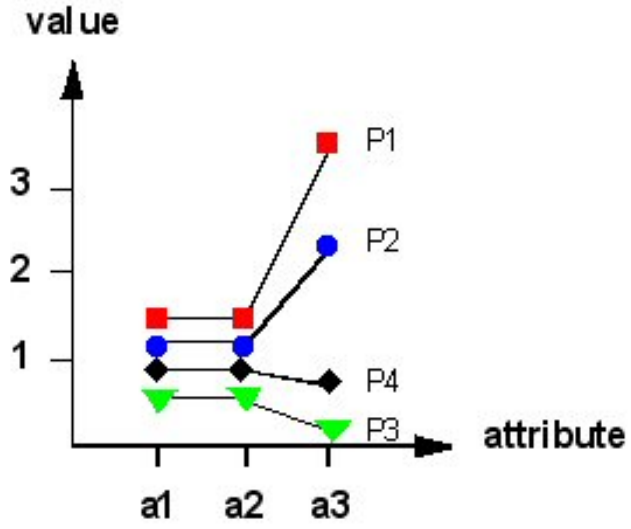


- points located on the bisecting line of participating attributes

Challenges and Approaches, Basic Models

- example – transposed view of attributes:

	a1	a2	a3
P1	1	1	3.5
P2	1	1	2.3
P3	1	1	0.2
P4	1	1	0.7



- pattern: identical constant lines

Challenges and Approaches, Basic Models

- real-world constant biclusters will not be perfect
- cluster model relaxes to:

$$a_{ij} \approx \mu$$

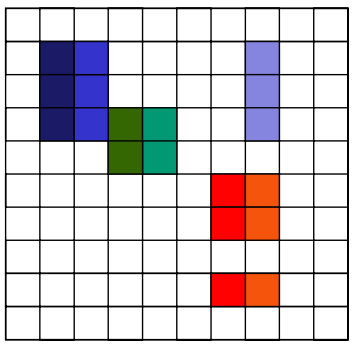
- Optimization on matrix $\boldsymbol{A}$ = (X,Y) may lead to |X|·|Y| singularity-biclusters each containing one entry.
- Challenge: Avoid this kind of overfitting.

Challenges and Approaches, Basic Models

Biclusters with constant values on columns

- Cluster model for $\boldsymbol{A}_{IJ}$ = (I,J):

$$a_{ij} = \mu + c_j$$
$$\forall i \in I, j \in J$$

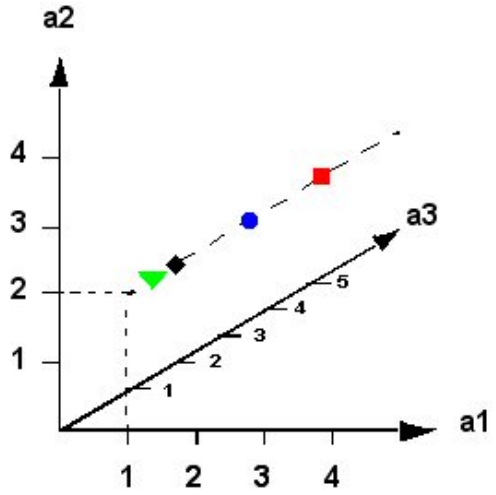


- adjustment value c_j for column $j \in J$
- results in axis-parallel subspace clusters

Challenges and Approaches, Basic Models

- example – 3-dimensional embedding space:

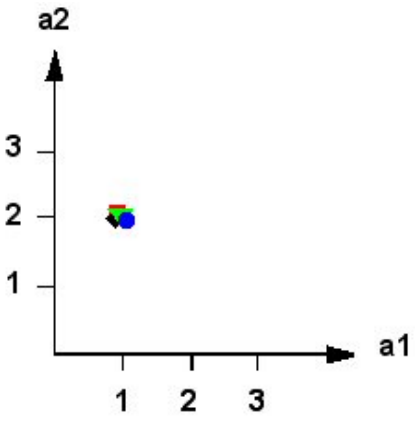
	a1	a2	a3
P1	1	2	3.5
P2	1	2	2.3
P3	1	2	0.2
P4	1	2	0.7



Challenges and Approaches, Basic Models

- example – 2-dimensional subspace:

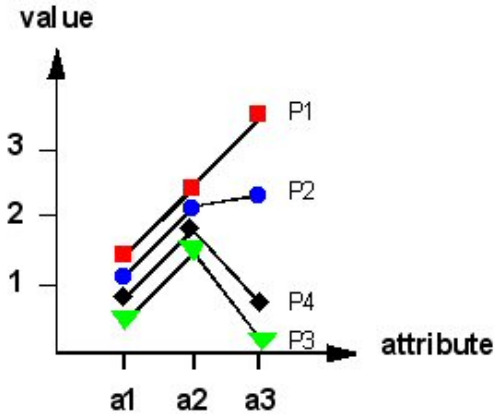
	a1	a2
P1	1	2
P2	1	2
P3	1	2
P4	1	2



Challenges and Approaches, Basic Models

- example – transposed view of attributes:

	a1	a2	a3
P1	1	2	3.5
P2	1	2	2.3
P3	1	2	0.2
P4	1	2	0.7



- pattern: identical lines

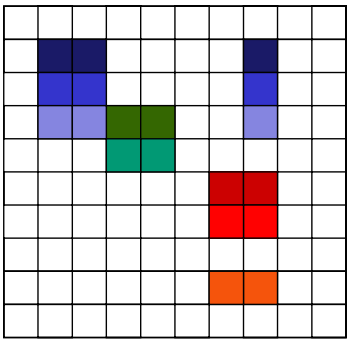
Challenges and Approaches, Basic Models

Biclusters with constant values on rows

- Cluster model for $A_{IJ} = (I,J)$:

$$a_{ij} = \mu + r_i$$

$$\forall i \in I, j \in J$$

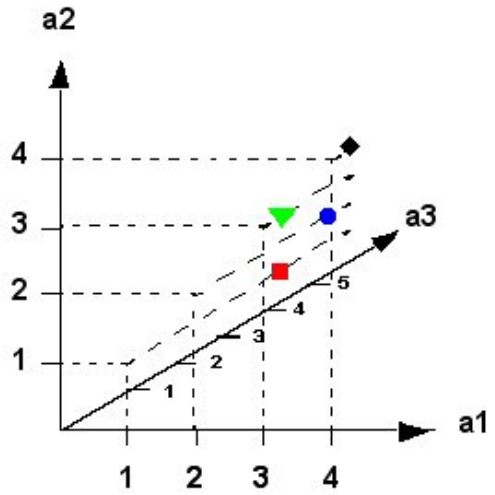


- adjustment value r_i for row $i \in I$

Challenges and Approaches, Basic Models

- example – 3-dimensional embedding space:

	a1	a2	a3
P1	1	1	3.5
P2	2	2	2.3
P3	3	3	0.2
P4	4	4	0.7

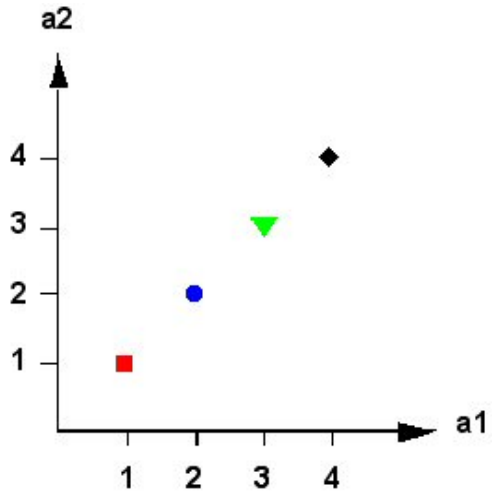


- in the embedding space, points build a sparse hyperplane parallel to irrelevant axes

Challenges and Approaches, Basic Models

- example – 2-dimensional subspace:

	a1	a2
P1	1	1
P2	2	2
P3	3	3
P4	4	4

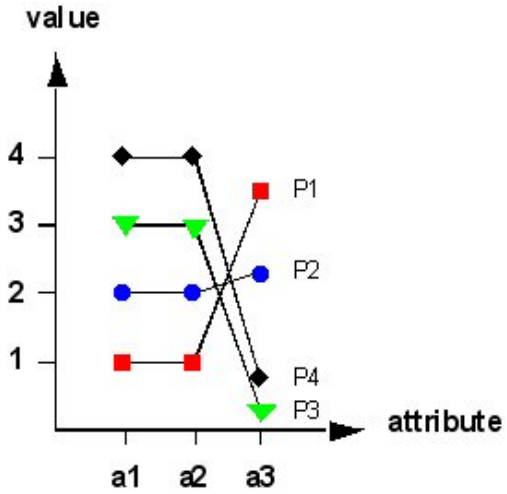


- points are accommodated on the bisecting line of participating attributes

Challenges and Approaches, Basic Models

- example – transposed view of attributes:

	a1	a2	a3
P1	1	1	3.5
P2	2	2	2.3
P3	3	3	0.2
P4	4	4	0.7



- pattern: parallel constant lines

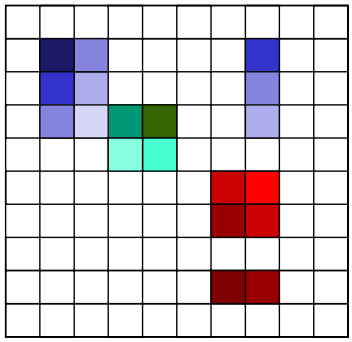
Challenges and Approaches, Basic Models

Biclusters with coherent values

- based on a particular form of covariance between rows and columns

$$a_{ij} = \mu + r_i + c_j$$

$$\forall i \in I, j \in J$$

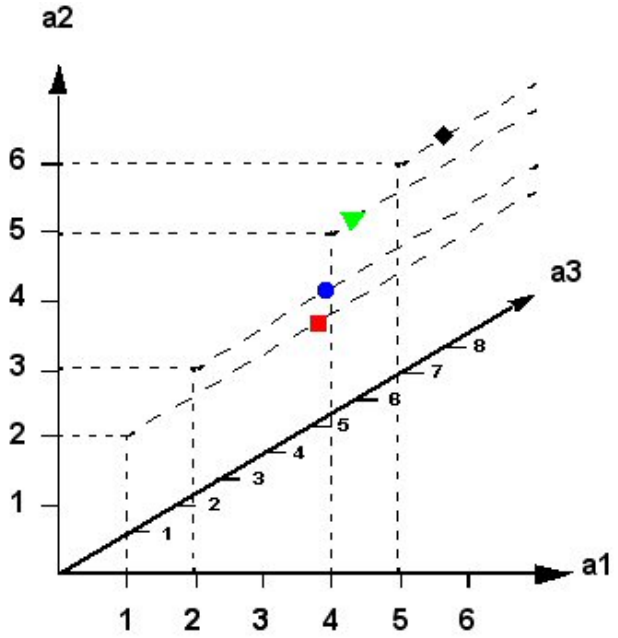


- special cases:
  - $c_j = 0$ for all j → constant values on rows
  - $r_i = 0$ for all i → constant values on columns

Challenges and Approaches, Basic Models

- embedding space: sparse hyperplane parallel to axes of irrelevant attributes

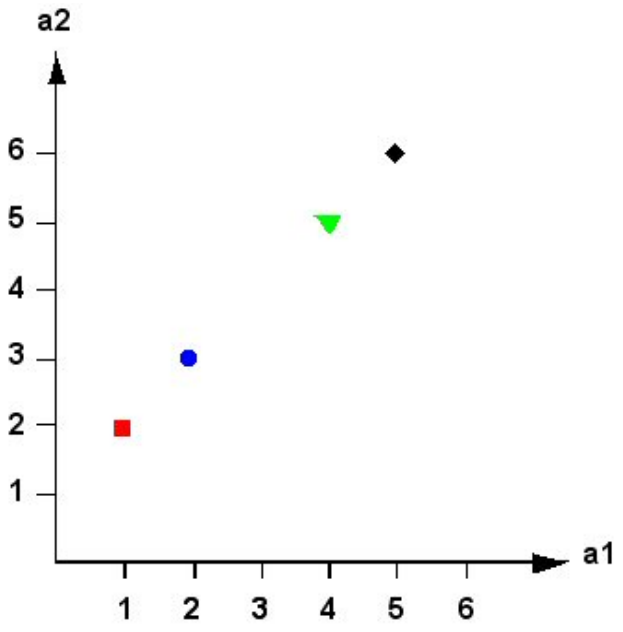
	a1	a2	a3
P1	1	2	3.5
P2	2	3	2.3
P3	4	5	0.2
P4	5	6	0.7



Challenges and Approaches, Basic Models

- subspace: increasing one-dimensional line

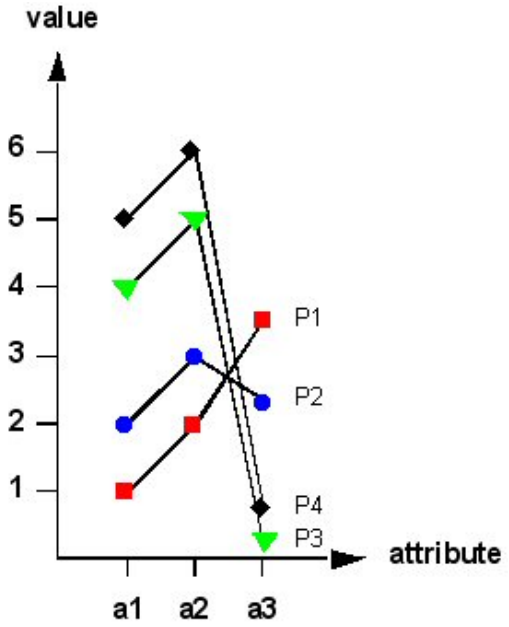
	a1	a2
P1	1	2
P2	2	3
P3	4	5
P4	5	6



Challenges and Approaches, Basic Models

- transposed view of attributes:

	a1	a2	a3
P1	1	2	3.5
P2	2	3	2.3
P3	4	5	0.2
P4	5	6	0.7

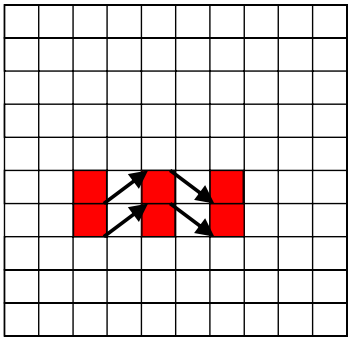


- pattern: parallel lines

Challenges and Approaches, Basic Models

Biclusters with coherent evolutions

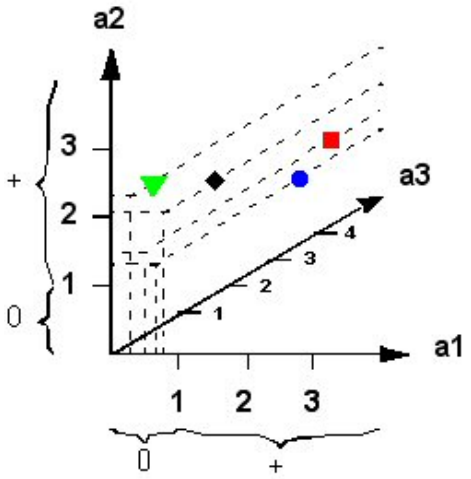
- for all rows, all pairs of attributes change simultaneously
  - discretized attribute space: coherent state-transitions
  - change in same direction irrespective of the quantity



Challenges and Approaches, Basic Models

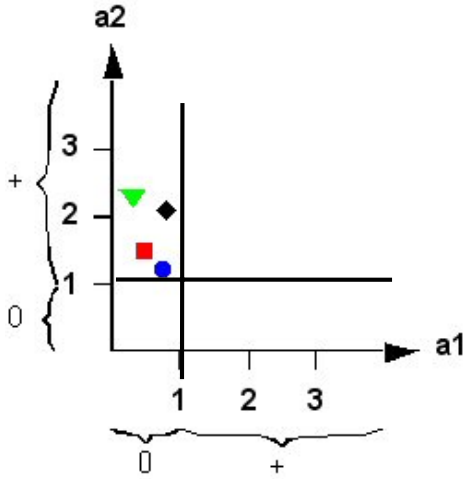
- Approaches with coherent state-transitions: [TSS02,MK03]
- reduces the problem to grid-based axis-parallel approach:

	a1	a2	a3
P1	0.5	1.5	3.5
P2	0.7	1.3	2.3
P3	0.3	2.3	0.2
P4	0.8	2.1	0.7

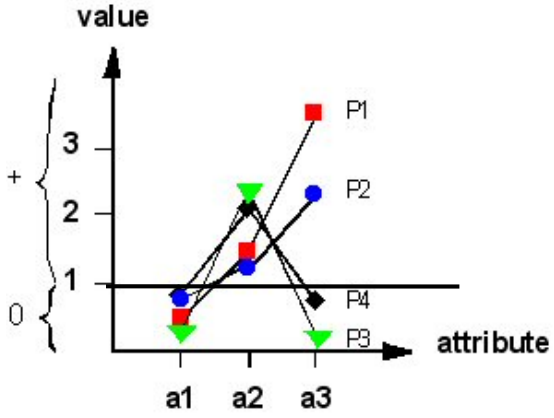


Challenges and Approaches, Basic Models

	a1	a2
P1	0	+
P2	0	+
P3	0	+
P4	0	+



	a1	a2	a3
P1	0.5	1.5	3.5
P2	0.7	1.3	2.3
P3	0.3	2.3	0.2
P4	0.8	2.1	0.7



pattern: all lines cross border between states (in the same direction)

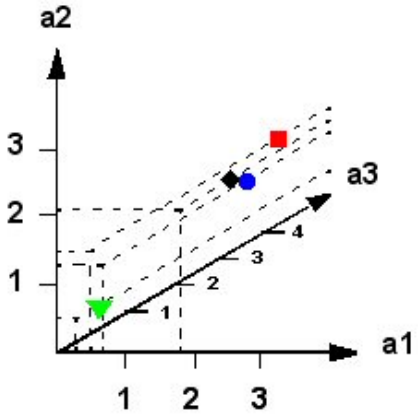
Challenges and Approaches, Basic Models

- change in same direction – general idea: find a subset of rows and columns, where a permutation of the set of columns exists such that the values in every row are increasing
- clusters do not form a subspace but rather half-spaces
- related approaches:
  - quantitative association rule mining [Web01,RRK04,GRRK05]
  - adaptation of formal concept analysis [GW99] to numeric data [Pfa07]

Challenges and Approaches, Basic Models

- example – 3-dimensional embedding space

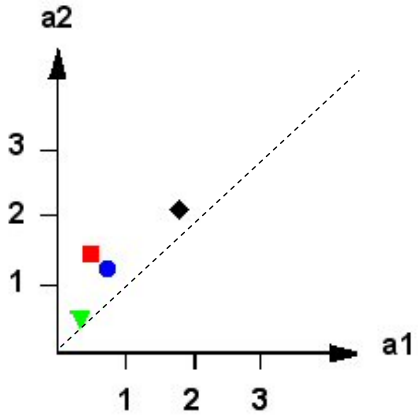
	a1	a2	a3
P1	0.5	1.5	3.5
P2	0.7	1.3	2.3
P3	0.3	0.5	0.2
P4	1.8	2.1	0.7



Challenges and Approaches, Basic Models

- example – 2-dimensional subspace

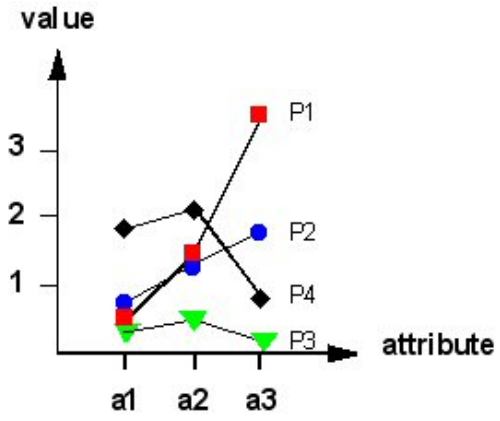
	a1	a2
P1	0.5	1.5
P2	0.7	1.3
P3	0.3	0.5
P4	1.8	2.1



Challenges and Approaches, Basic Models

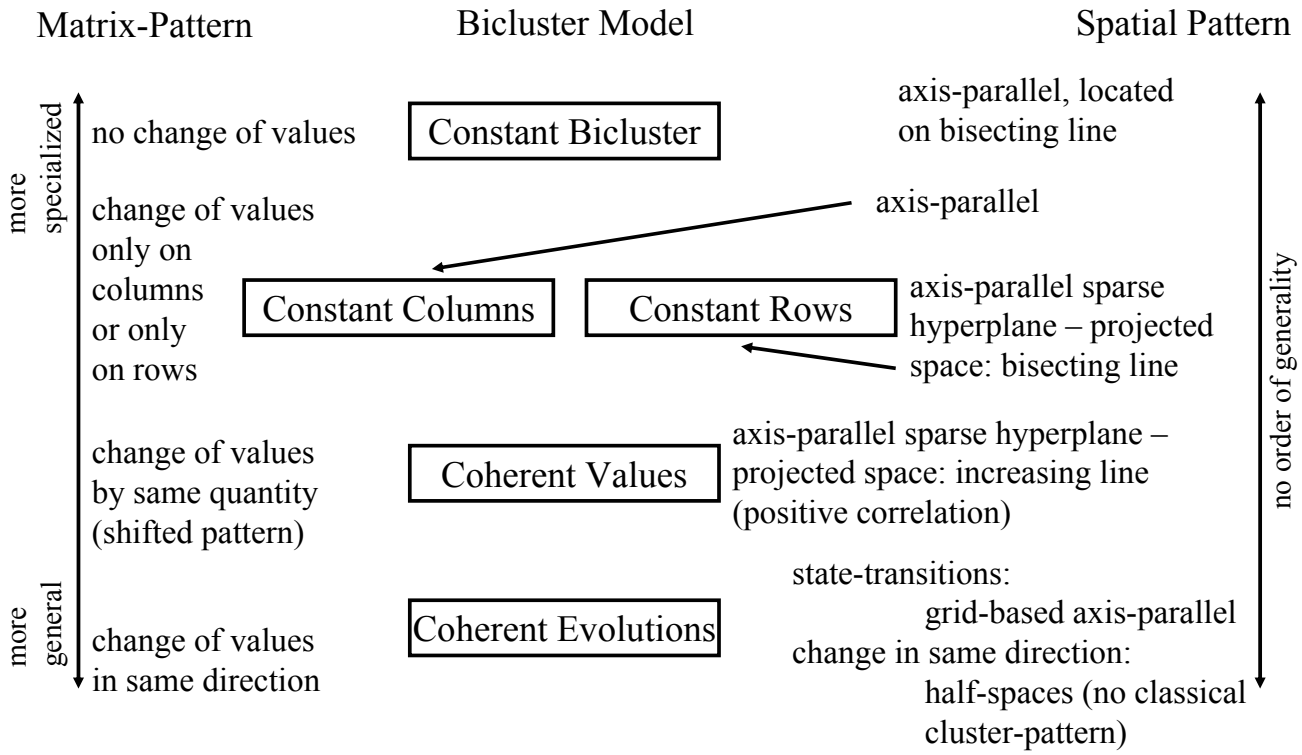
- example – transposed view of attributes

	a1	a2	a3
P1	0.5	1.5	3.5
P2	0.7	1.3	2.3
P3	0.3	0.5	0.2
P4	1.8	2.1	0.7



- pattern: all lines increasing

Challenges and Approaches, Basic Models



Algorithms for Constant Biclusters

- classical problem statement by Hartigan [Har72]
- quality measure for a bicluster: variance of the submatrix $\boldsymbol{A}_{IJ}$:

$$VAR\left(A_{IJ}\right) = \sum_{i \in I,\, j \in J} \left(a_{ij} - a_{IJ}\right)^2$$

- avoids partitioning into |X|·|Y| singularity-biclusters (optimizing the sum of squares) by comparing the reduction with the reduction expected by chance
- recursive split of data matrix into two partitions
- each split chooses the maximal reduction in the overall sum of squares for all biclusters

Biclusters with Constant Values in Rows or Columns

- simple approach: normalization to transform the biclusters into constant biclusters and follow the first approach (e.g. [GLD00])
- some application-driven approaches with special assumptions in the bioinformatics community (e.g. [CST00,SMD03,STG+01])
- constant values on columns: general axis-parallel subspace/projected clustering
- constant values on rows: special case of general correlation clustering
- both cases special case of approaches to biclusters with coherent values

Algorithms for Biclusters with Coherent Values

classical approach: Cheng&Church [CC00]

- introduced the term biclustering to analysis of gene expression data
- quality of a bicluster: *mean squared residue* value *H*

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

- submatrix (I,J) is considered a bicluster, if H(I,J) < δ

Algorithms for Biclusters with Coherent Values

- $\delta = 0 \rightarrow$ *perfect* bicluster:
  - each row and column exhibits absolutely consistent bias
  - bias of row *i* w.r.t. other rows:

$$a_{iJ} - a_{IJ}$$

- the model for a perfect bicluster predicts value a_{ij} by a row-constant, a column-constant, and an overall cluster-constant:

$$a_{ij} = a_{iJ} + a_{Ij} - a_{IJ}$$

$$\Updownarrow \mu = a_{IJ}, r_i = a_{iJ} - a_{IJ}, c_j = a_{Ij} - a_{IJ}$$

$$a_{ij} = \mu + r_i + c_j$$

Algorithms for Biclusters with Coherent Values

- for a non-perfect bicluster, the prediction of the model deviates from the true value by a residue:

$$a_{ij} = \text{res}(a_{ij}) + a_{iJ} + a_{Ij} - a_{IJ}$$

$$\Updownarrow$$

$$\text{res}(a_{ij}) = a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}$$

- This residue is the optimization criterion:

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

Algorithms for Biclusters with Coherent Values

- The optimization is also possible for the row-residue of row *i* or the column-residue of column *j*.
- Algorithm:
  1. find a δ -bicluster: greedy search by removing the row or column (or the set of rows/columns) with maximal mean squared residue until the remaining submatrix (I,J) satisfies H(I,J)< δ.
  2. find a maximal δ -bicluster by adding rows and columns to (I,J) unless this would increase *H*.
  3. replace the values of the found bicluster by random numbers and repeat the procedure until *k* δ -biclusters are found.

Algorithms for Biclusters with Coherent Values

Weak points in the approach of Cheng&Church:

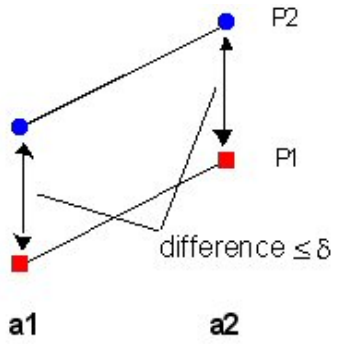
1. One cluster at a time is found, the cluster needs to be masked in order to find a second cluster.
2. This procedure bears an inefficient performance.
3. The masking may lead to less accurate results.
4. The masking inhibits simultaneous overlapping of rows and columns.
5. Missing values cannot be dealt with.
6. The user must specify the number of clusters beforehand.

Algorithms for Biclusters with Coherent Values

p-cluster model [WWYY02]

- p-cluster model: deterministic approach
- specializes δ -bicluster-property to a pairwise property of two objects in two attributes:

$$\left|\left(a_{i_1 j_1} - a_{i_1 j_2}\right) - \left(a_{i_2 j_1} - a_{i_2 j_2}\right)\right| \leq \delta$$



- submatrix (I,J) is a δ -p-cluster if this property is fulfilled for any 2x2 submatrix ($\{i_1, i_2\}$, $\{j_1, j_2\}$) where $\{i_1, i_2\} \in$ I and $\{j_1, j_2\} \in$ J.

Algorithms for Biclusters with Coherent Values

Algorithm:

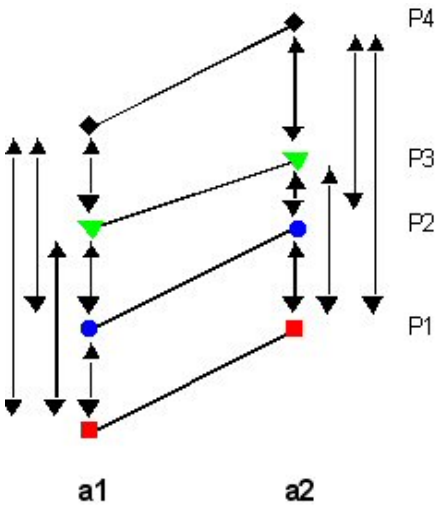
1. create maximal set of attributes for each pair of objects forming a δ -p-cluster
2. create maximal set of objects for each pair of attributes forming a δ -p-cluster
3. pruning-step
4. search in the set of submatrices

Problem: complete enumeration approach

Addressed issues:

1. multiple clusters simultaneously
4. allows for overlapping rows and columns
6. allows for arbitrary number of clusters

Related approaches: FLOC [YWWY02], MaPle [PZC+03]



Summary

- Biclustering models do not fit exactly into the spatial intuition behind subspace, projected, or correlation clustering.
- Models make sense in view of a data matrix.
- Strong point: the models generally do not rely on the locality assumption.
- Models differ substantially → fair comparison is a non-trivial task.
- Comparison of five methods: [PBZ+06]
- Rather specialized task – comparison in a broad context (subspace/projected/correlation clustering) is desirable.
- Biclustering performs generally well on microarray data – for a wealth of approaches see [MO04].

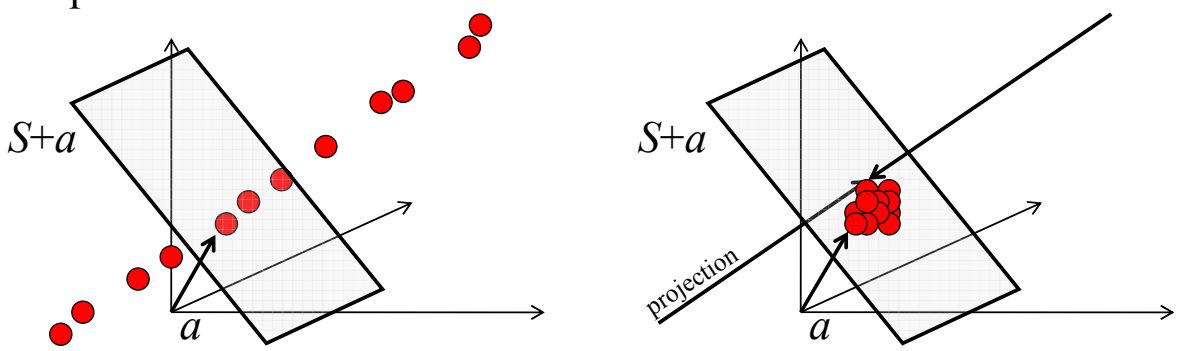
Outline

Outline: Arbitrarily-oriented Subspace Clustering

- Challenges and Approaches
- Correlation Clustering Algorithms
- Summary and Perspectives

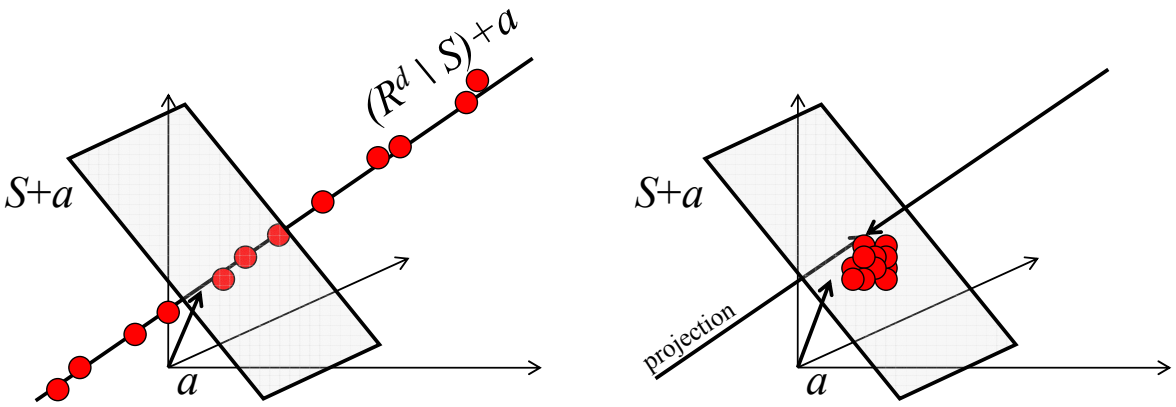
Challenges and Approaches

- Pattern-based approaches find simple positive correlations
- More general approach: oriented clustering aka. generalized subspace/projected clustering aka. correlation clustering
  - Note: different notion of "Correlation Clustering" in machine learning community, e.g. cf. [BBC04]
- Assumption: any cluster is located in an arbitrarily oriented affine subspace $S+a$ of R^d



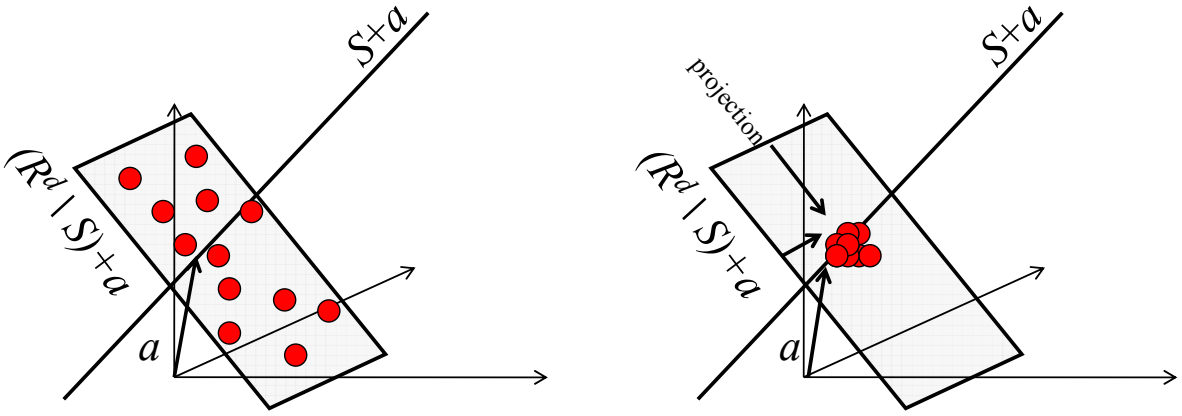
Challenges and Approaches

- Affine subspace $S+a$, $S \subset R^d$, *affinity* $a \in R^d$ is interesting if a set of points clusters within this subspace
- Points may exhibit high variance in perpendicular subspace $(R^d \setminus S)+a$



Challenges and Approaches

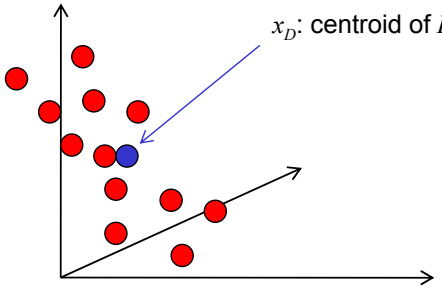
- high variance in perpendicular subspace $(R^d \setminus S)+a$ → points form a hyperplane within R^d located in this subspace $(R^d \setminus S)+a$
- Points on a hyperplane appear to follow linear dependencies among the attributes participating in the description of the hyperplane



Challenges and Approaches

- Directions of high/low variance: PCA (local application)
- locality assumption: local selection of points sufficiently reflects the hyperplane accommodating the points
- general approach: build covariance matrix Σ_D for a selection D of points (e.g. k nearest neighbors of a point)

$$\Sigma_D = \frac{1}{|D|}\sum_{x \in D}(x - x_D)(x - x_D)^T$$



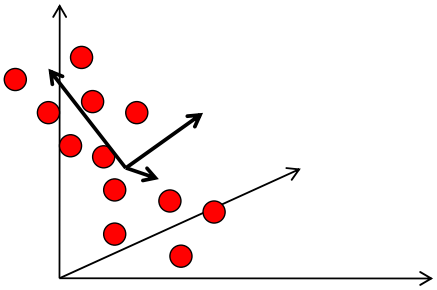
properties of Σ_D:
- d x d
- symmetric
- positive semidefinite
- $\sigma_{D_{ij}}$ (value at row i, column j) = covariance between dimensions i and j
- $\sigma_{D_{ii}}$ = variance in ith dimension

Challenges and Approaches

- decomposition of Σ_D to eigenvalue matrix E_D and eigenvector matrix V_D:

$$\Sigma_D = V_D E_D V_D^T$$

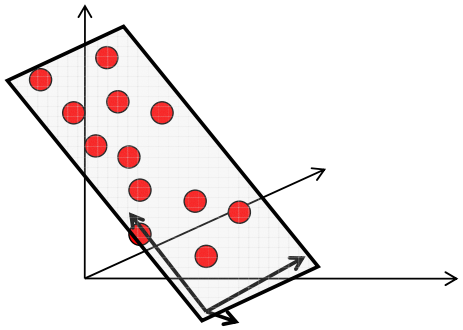
- E_D : diagonal matrix, holding eigenvalues of Σ_D in decreasing order in its diagonal elements
- V_D : orthonormal matrix with eigenvectors of Σ_D ordered correspondingly to the eigenvalues in E_D



- V_D : new basis, first eigenvector = direction of highest variance
- E_D : covariance matrix of D when represented in new axis system V_D

Challenges and Approaches

- points forming λ-dimensional hyperplane → hyperplane is spanned by the first λ eigenvectors (called "strong" eigenvectors – notation: $\breve{V}_D$)
- subspace where the points cluster densely is spanned by the remaining d-λ eigenvectors (called "weak" eigenvectors – notation: $\hat{V}_D$)



for the eigensystem, the sum of the smallest d-λ eigenvalues $\sum_{i=\lambda+1}^{d} e_{D_{ii}}$ is minimal under all possible transformations → points cluster optimally dense in this subspace

Challenges and Approaches

model for correlation clusters [ABK+06]:

- λ-dimensional hyperplane accommodating the points of a correlation cluster $C \subset R^d$ is defined by an equation system of d-λ equations for d variables and the affinity (e.g. the mean point x_C of all cluster members):

$$\hat{V}_C^{\mathrm{T}} x = \hat{V}_C^{\mathrm{T}} x_C$$

- equation system approximately fulfilled for all points $x \in C$
- quantitative model for the cluster allowing for probabilistic prediction (classification)
- Note: correlations are observable, linear dependencies are merely an assumption to explain the observations – predictive model allows for evaluation of assumptions and experimental refinements

Correlation Clustering Algorithms

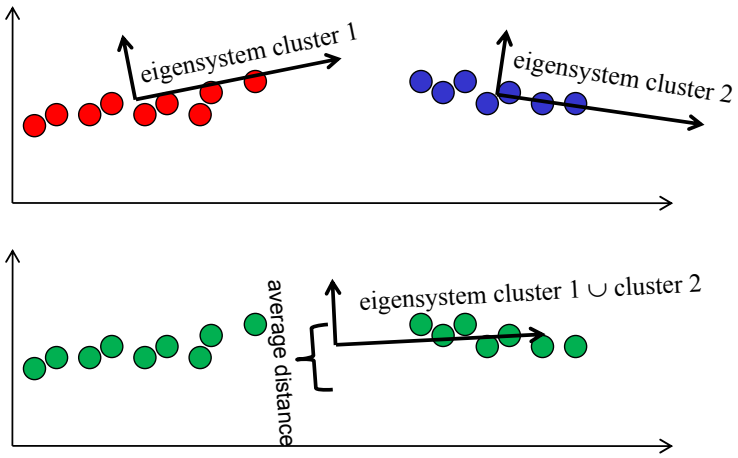
ORCLUS [AY00]:

first approach to *generalized projected clustering*

- similar ideas to PROCLUS [APW+99]
- k-means like approach
- start with $k_c > k$ seeds
- assign cluster members according to distance function based on the eigensystem of the current cluster (starting with axes of data space, i.e. Euclidean distance)
- reduce k_c in each iteration by merging best-fitting cluster pairs

Correlation Clustering Algorithms

- best fitting pair of clusters: least average distance in the projected space spanned by weak eigenvectors of the merged clusters



- assess average distance in all merged pairs of clusters and finally merge the best fitting pair

Correlation Clustering Algorithms

- adapt eigensystem to the updated cluster
- new iteration: assign points according to updated eigensystems (distance along weak eigenvectors)
- dimensionality gradually reduced to a user-specified value l
- initially exclude only eigenvectors with very high variance

Correlation Clustering Algorithms

properties:

- finds k correlation clusters (user-specified)
- higher initial k_c → higher runtime, probably better results
- biased to average dimensionality l of correlation clusters (user specified)
- cluster-based locality assumption: subspace of each cluster is learned from its current members (starting in the full dimensional space)

Correlation Clustering Algorithms

4C [BKKZ04]

- density-based cluster-paradigm (cf. DBSCAN [EKSX96])
- extend a cluster from a seed as long as a density-criterion is fulfilled – otherwise pick another seed unless all data base objects are assigned to a cluster or noise
- density criterion: minimal required number of points in the neighborhood of a point
- neighborhood: distance between two points ascertained based on the eigensystems of both compared points

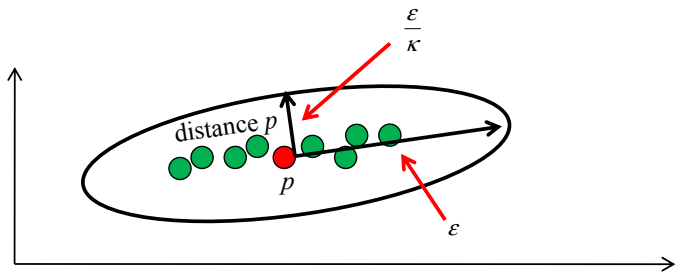
Correlation Clustering Algorithms

- eigensystem of a point p based on its ε-neighborhood in Euclidean space
- threshold δ discerns large from small eigenvalues
- in eigenvalue matrix E_p replace large eigenvalues by 1, small eigenvalues by $\kappa >> 1$
- adapted eigenvalue matrix yields a correlation similarity matrix for point p:

$$V_p E'_p V_p^{\mathrm{T}}$$

Correlation Clustering Algorithms

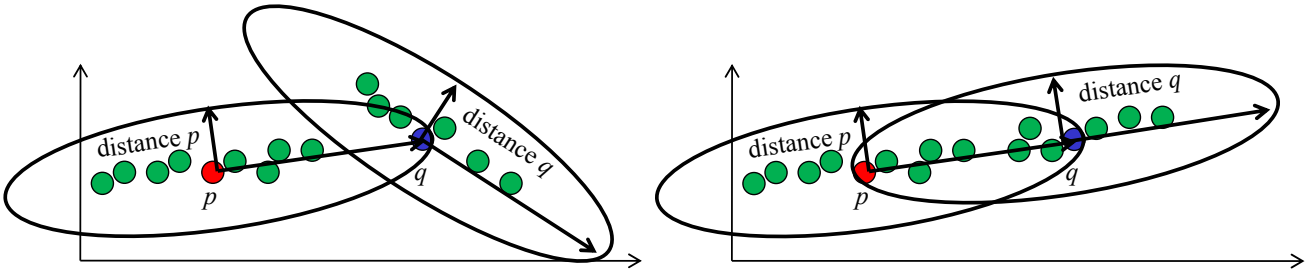
- effect on distance measure:



- distance of p and q w.r.t. p: $\sqrt{(p-q)\cdot V_p \cdot E'_p \cdot V_p^{\mathrm{T}} \cdot (p-q)^{\mathrm{T}}}$
- distance of p and q w.r.t. q: $\sqrt{(q-p)\cdot V_q \cdot E'_q \cdot V_q^{\mathrm{T}} \cdot (q-p)^{\mathrm{T}}}$

Correlation Clustering Algorithms

- symmetry of distance measure by choosing the maximum:



- p and q are correlation-neighbors if

$$\max\left\{\sqrt{(p-q)\cdot V_p \cdot E'_p \cdot V_p^{\mathrm{T}} \cdot (p-q)^{\mathrm{T}}},\ \sqrt{(q-p)\cdot V_q \cdot E'_q \cdot V_q^{\mathrm{T}} \cdot (q-p)^{\mathrm{T}}}\right\} \leq \varepsilon$$

Correlation Clustering Algorithms

properties:

- finds arbitrary number of clusters
- requires specification of density-thresholds
  - μ (minimum number of points): rather intuitive
  - ε (radius of neighborhood): hard to guess
- biased to maximal dimensionality λ of correlation clusters (user specified)
- instance-based locality assumption: correlation distance measure specifying the subspace is learned from local neighborhood of each point in the *d*-dimensional space

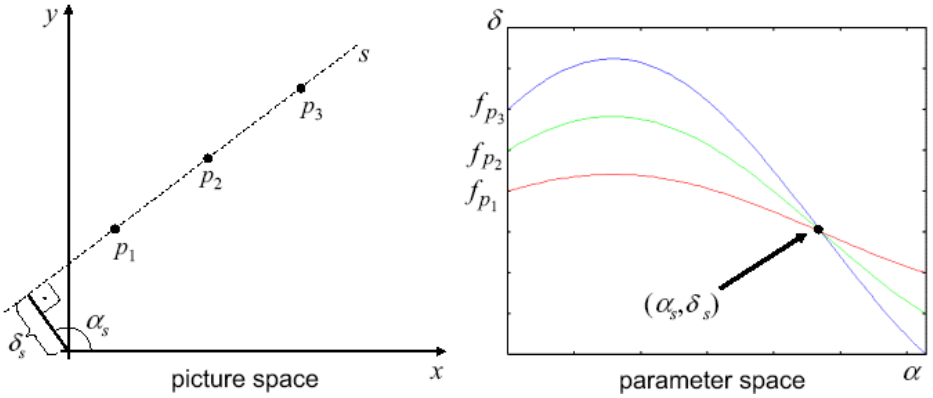
enhancements also based on PCA:

- COPAC [ABK+07c] and
- ERiC [ABK+07b]

Correlation Clustering Algorithms

different correlation primitive: Hough-transform

- points in data space are mapped to functions in the parameter space

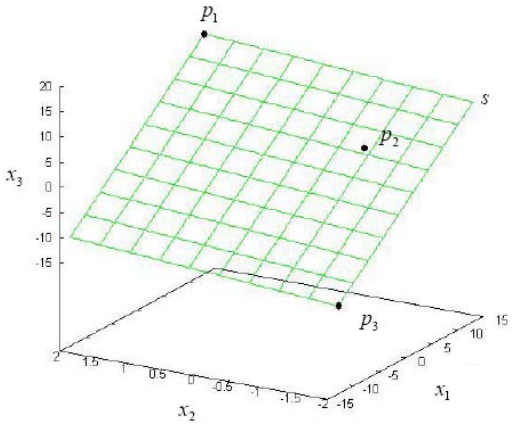


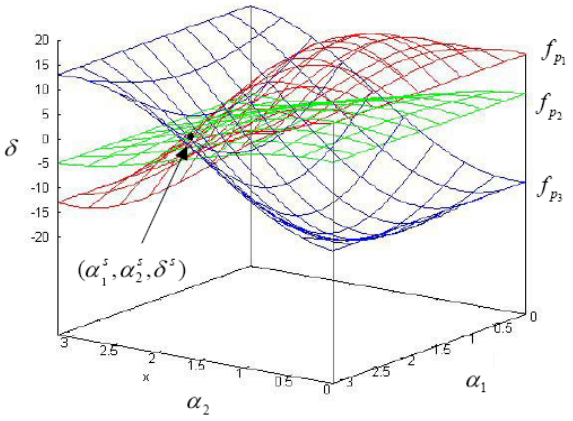
$$f_p(\alpha_1, \ldots, \alpha_{d-1}) = \langle p, n \rangle = \sum_{i=1}^{d} p_i \cdot \left(\prod_{j=1}^{i-1} \sin(\alpha_j) \right) \cdot \cos(\alpha_i)$$

- functions in the parameter space define all lines possibly crossing the point in the data space

Correlation Clustering Algorithms

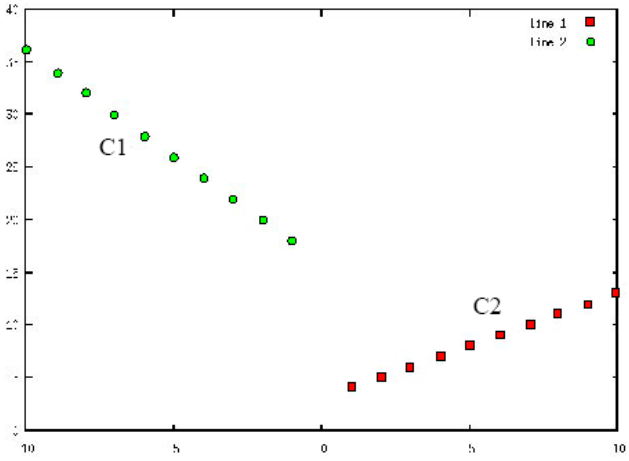
- Properties of the transformation
  - Point in the data space = sinusoidal curve in parameter space
  - Point in parameter space = hyper-plane in data space
  - Points on a common hyper-plane in data space = sinusoidal curves through a common point in parameter space
  - Intersections of sinusoidal curves in parameter space = hyper-plane through the corresponding points in data space

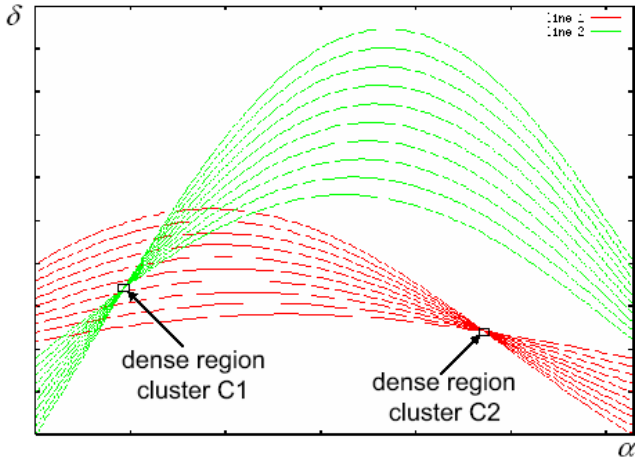




Correlation Clustering Algorithms

Algorithm based on the Hough-transform: CASH [ABD+08]





dense regions in parameter space correspond to linear structures in data space

Correlation Clustering Algorithms

Idea: find dense regions in parameter space

- construct a grid by recursively splitting the parameter space (best-first-search)
- identify dense grid cells as intersected by many parametrization functions
- dense grid represents (*d-1*)-dimensional linear structure
- transform corresponding data objects in corresponding (*d-1*)-dimensional space and repeat the search recursively

Correlation Clustering Algorithms

properties:

- finds arbitrary number of clusters
- requires specification of depth of search (number of splits per axis)
- requires minimum density threshold for a grid cell
- Note: this minimum density does not relate to the locality assumption: CASH is a global approach to correlation clustering
- search heuristic: linear in number of points, but $\sim d^4$
- But: complete enumeration in worst case (exponential in d)

Summary and Perspectives

- PCA: mature technique, allows construction of a broad range of similarity measures for local correlation of attributes
- drawback: all approaches suffer from locality assumption
- successfully employing PCA in correlation clustering in “really” high-dimensional data requires more effort henceforth
- new approach based on Hough-transform:
  - does not rely on locality assumption
  - but worst case again complete enumeration

Summary and Perspectives

- some preliminary approaches base on concept of self-similarity (intrinsic dimensionality, fractal dimension): [BC00,PTTF02,GHPT05]
- interesting idea, provides quite a different basis to grasp correlations in addition to PCA
- drawback: self-similarity assumes locality of patterns even by definition

Summary and Perspectives

comparison: correlation clustering – biclustering:

- model for correlation clusters more general and meaningful
- models for biclusters rather specialized
- in general, biclustering approaches do not rely on locality assumption
- non-local approach and specialization of models may make biclustering successful in many applications
- correlation clustering is the more general approach but the approaches proposed so far are rather a first draft to tackle the complex problem

Outline

Summary

- Let's take a global view:
  - Traditional clustering in high dimensional spaces is most likely meaningless with increasing dimensionality (curse of dimensionality)
  - Clusters may be found in (generally arbitrarily oriented) subspaces of the data space
  - So the general problem of clustering high dimensional data is:

    "find a partitioning of the data where each cluster may exist in its own subspace"
    - The partitioning need not be unique (clusters may overlap)
    - The subspaces may be axis-parallel or arbitrarily oriented
  - Analysis of this general problem:
    - A naïve solution would examine all possible subspaces to look for clusters
    - The search space of all possible arbitrarily oriented subspaces is infinite
    - We need assumptions and heuristics to develop a feasible solution

Summary

- What assumptions did we get to know here?
  - The search space is restricted to certain subspaces
  - A clustering criterion that implements the downward closure property enables efficient search heuristics
  - The locality assumption enables efficient search heuristics
  - Assuming simple additive models ("patterns") enables efficient search heuristics
  - …
- Remember: also the clustering model may rely on further assumptions that have nothing to do with the infinite search space
  - Number of clusters need to be specified
  - Results are not deterministic e.g. due to randomized procedures
  - …
- We can classify the existing approaches according to the assumptions they made to conquer the infinite search space

Summary

- The global view
  - Subspace clustering/projected clustering:
    - The search space is restricted to axis-parallel subspaces
    - A clustering criterion that implements the downward closure property is defined (usually based on a global density threshold)
    - The locality assumption enables efficient search heuristics
  - Bi-clustering/pattern-based clustering:
    - The search space is restricted to special forms and locations of subspaces or half-spaces
    - Over-optimization (e.g. singularity clusters) is avoided by assuming a predefined number of clusters
  - Correlation clustering:
    - The locality assumption enables efficient search heuristics

- Any of the proposed methods is based on at least one assumption because otherwise, it would not be applicable

Summary

Algorithm	complex correlations	simple positive correlation	simple negative correlation	axis parallel	not relying on locality assumption	adaptive density threshold	independent w.r.t. order of attributes	independent w.r.t. order of objects	deterministic	arbitrary number of clusters	overlapping clusters	overlapping subspaces	simultaneously overlapping clusters and subspaces	arbitrary subspace dimensionality	hierarchical structure	avoiding complete enumeration	noise robust
CLIQUE [AGGR98]				√	√		√	√	√	√	√	√	√	√			√
ENCLUS [CFZ99]				√	√		√	√	√	√	√	√	√	√			√
MAFIA [NGC01]				√	√		√	√	√	√	√	√	√	√			√
SUBCLU [KKK04]				√	√		√	√	√	√	√	√	√	√			√
PROCLUS [APW+99]				√		√						√				√	
PreDeCon [BKKK04]				√			√	√	√	√		√				√	√
P3C [MSE06]				√	√	√	√	√	√	√	√	√	√	√			√
COSA [FM04]				√			√	√	√			√		√		√	√
DOC [PJAM02]				√	√		√	√		√	√	√	√	√			
DiSH [ABK+07a]				√	√	√		√	√	√		√		√	√	√	√
FIRES [KKRW05]				√	√	√		√	√	√	√	√	√	√		√	√

Summary

Algorithm	complex correlations	simple positive correlation	simple negative correlation	axis parallel	not relying on locality assumption	adaptive density threshold	independent w.r.t. order of attributes	independent w.r.t. order of objects	deterministic	arbitrary number of clusters	overlapping clusters	overlapping subspaces	simultaneously overlapping clusters and subspaces	arbitrary subspace dimensionality	hierarchical structure	avoiding complete enumeration	noise robust
Block clustering [Har72]					√	*n a*	√	√	√					√	√		√
δ-bicluster [CC00]		√	√	√	√	*n a*	√	√	√		√	√		√		√	√
FLOC [YWWY02]		√		√	√	*n a*					√	√	√	√		√	√
p-Cluster [WWYY02]		√		√	√	*n a*	√	√	√	√	√	√	√	√			√
MaPle [PZC+03]		√		√	√	*n a*	√	√	√	√	√	√	√	√			√
CoClus [CDGS04]		√		√	√	*n a*								√		√	
OP-Cluster [LW03]					√	*n a*	√	√	√	√	√	*n a*	*n a*	*n a*			√

Summary

Algorithm	complex correlations	simple positive correlation	simple negative correlation	axis parallel	not relying on locality assumption	adaptive density threshold	independent w.r.t. order of attributes	independent w.r.t. order of objects	deterministic	arbitrary number of clusters	overlapping clusters	overlapping subspaces	simultaneously overlapping clusters and subspaces	arbitrary subspace dimensionality	hierarchical structure	avoiding complete enumeration	noise robust
ORCLUS [AY00]	√	√	√	√			√					√				√	
4C [BKKZ04]	√	√	√	√			√	√	√	√		√				√	√
COPAC [ABK+07c]	√	√	√	√			√	√	√	√		√		√		√	√
ERiC [ABK+07b]	√	√	√	√			√	√	√	√		√		√	√	√	√
CASH [ABD+08]	√	√	√	√	√	*n a*		√	√	√		√		√	√		√

List of References

Literature

[ABD+08] E. Achtert, C. Böhm, J. David, P. Kröger, and A. Zimek.
Robust clustering in arbitrarily oriented subspaces.
In Proceedings of the 8th SIAM International Conference on Data Mining (SDM), Atlanta, GA, 2008

[ABK+06] E. Achtert, C. Böhm, H.-P. Kriegel, P. Kröger, and A. Zimek.
Deriving quantitative models for correlation clusters.
In Proceedings of the 12th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Philadelphia, PA, 2006.

[ABK+07a] E. Achtert, C. Böhm, H.-P. Kriegel, P. Kröger, I. Müller-Gorman, and A. Zimek.
Detection and visualization of subspace cluster hierarchies.
In Proceedings of the 12th International Conference on Database Systems for Advanced Applications (DASFAA), Bangkok, Thailand, 2007.

[ABK+07b] E. Achtert, C. Böhm, H.-P. Kriegel, P. Kröger, and A. Zimek.
On exploring complex relationships of correlation clusters.
In Proceedings of the 19th International Conference on Scientific and Statistical Database Management (SSDBM), Banff, Canada, 2007.

[ABK+07c] E. Achtert, C. Böhm, H.-P. Kriegel, P. Kröger, and A. Zimek.
Robust, complete, and efficient correlation clustering.
In Proceedings of the 7th SIAM International Conference on Data Mining (SDM), Minneapolis, MN, 2007.

Literature

[AGGR98] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan.
Automatic subspace clustering of high dimensional data for data mining applications.
In Proceedings of the ACM International Conference on Management of Data (SIGMOD), Seattle, WA, 1998.

[AHK01] C. C. Aggarwal, A. Hinneburg, and D. Keim.
On the surprising behavior of distance metrics in high dimensional space.
In Proceedings of the 8th International Conference on Database Theory (ICDT), London, U.K., 2001.

[APW+99] C. C. Aggarwal, C. M. Procopiuc, J. L. Wolf, P. S. Yu, and J. S. Park.
Fast algorithms for projected clustering.
In Proceedings of the ACM International Conference on Management of Data (SIGMOD), Philadelphia, PA, 1999.

[AS94] R. Agrawal and R. Srikant. **Fast algorithms for mining association rules**.
In Proceedings of the ACM International Conference on Management of Data (SIGMOD), Minneapolis, MN, 1994.

[AY00] C. C. Aggarwal and P. S. Yu.
Finding generalized projected clusters in high dimensional space.
In Proceedings of the ACM International Conference on Management of Data (SIGMOD), Dallas, TX, 2000.

Literature

[BBC04] N. Bansal, A. Blum, and S. Chawla.
Correlation clustering.
Machine Learning, 56:89–113, 2004.

[BC00] D. Barbara and P. Chen.
Using the fractal dimension to cluster datasets.
In Proceedings of the 6th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Boston, MA, 2000.

[BDCKY02] A. Ben-Dor, B. Chor, R. Karp, and Z. Yakhini.
Discovering local structure in gene expression data: The order-preserving submatrix problem.
In Proceedings of the 6th Annual International Conference on Computational Molecular Biology (RECOMB), Washington, D.C., 2002.

[BGRS99] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft.
When is "nearest neighbor" meaningful?
In Proceedings of the 7th International Conference on Database Theory (ICDT), Jerusalem, Israel, 1999.

[BKKK04] C. Böhm, K. Kailing, H.-P. Kriegel, and P. Kröger.
Density connected clustering with local subspace preferences.
In Proceedings of the 4th International Conference on Data Mining (ICDM), Brighton, U.K., 2004.

Literature

[BKKZ04] C. Böhm, K. Kailing, P. Kröger, and A. Zimek.
Computing clusters of correlation connected objects.
In Proceedings of the ACM International Conference on Management of Data (SIGMOD), Paris, France, 2004.

[CC00] Y. Cheng and G. M. Church.
Biclustering of expression data.
In Proceedings of the 8th International Conference Intelligent Systems for Molecular Biology (ISMB), San Diego, CA, 2000.

[CDGS04] H. Cho, I. S. Dhillon, Y. Guan, and S. Sra.
Minimum sum-squared residue co-clustering of gene expression data.
In Proceedings of the 4th SIAM International Conference on Data Mining (SDM), Orlando, FL, 2004.

[CFZ99] C. H. Cheng, A. W.-C. Fu, and Y. Zhang.
Entropy-based subspace clustering for mining numerical data.
In Proceedings of the 5th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), San Diego, CA, pages 84–93, 1999.

Literature

[CST00] A. Califano, G. Stolovitzky, and Y. Tu.
Analysis of gene expression microarrays for phenotype classification.
In Proceedings of the 8th International Conference Intelligent Systems for Molecular Biology (ISMB), San Diego, CA, 2000.

[EKSX96] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu.
A density-based algorithm for discovering clusters in large spatial databases with noise.
In Proceedings of the 2nd ACM International Conference on Knowledge Discovery and Data Mining (KDD), Portland, OR, 1996.

[FM04] J. H. Friedman and J. J. Meulman.
Clustering objects on subsets of attributes.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66(4):825–849, 2004.

[GHPT05] A. Gionis, A. Hinneburg, S. Papadimitriou, and P. Tsaparas.
Dimension induced clustering.
In Proceedings of the 11th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Chicago, IL, 2005.

Literature

[GLD00] G. Getz, E. Levine, and E. Domany.
Coupled two-way clustering analysis of gene microarray data.
Proceedings of the National Academy of Sciences of the United States of America, 97(22):12079–12084, 2000.

[GRRK05] E. Georgii, L. Richter, U. Rückert, and S. Kramer.
Analyzing microarray data using quantitative association rules.
Bioinformatics, 21(Suppl. 2):ii1–ii8, 2005.

[GW99] B. Ganter and R. Wille.
Formal Concept Analysis.
Mathematical Foundations. Springer, 1999.

[HAK00] A. Hinneburg, C. C. Aggarwal, and D. A. Keim.
What is the nearest neighbor in high dimensional spaces?
In Proceedings of the 26th International Conference on Very Large Data Bases (VLDB), Cairo, Egypt, 2000.

[Har72] J. A. Hartigan.
Direct clustering of a data matrix.
Journal of the American Statistical Association, 67(337):123–129, 1972.

Literature

[IBB04] J. Ihmels, S. Bergmann, and N. Barkai.
Defining transcription modules using large-scale gene expression data.
Bioinformatics, 20(13):1993–2003, 2004.

[Jol02] I. T. Jolliffe.
Principal Component Analysis.
Springer, 2nd edition, 2002.

[KKK04] K. Kailing, H.-P. Kriegel, and P. Kröger.
Density-connected subspace clustering for highdimensional data.
In Proceedings of the 4th SIAM International Conference on Data Mining (SDM), Orlando, FL, 2004.

[KKRW05] H.-P. Kriegel, P. Kröger, M. Renz, and S. Wurst.
A generic framework for efficient subspace clustering of high-dimensional data.
In Proceedings of the 5th International Conference on Data Mining (ICDM), Houston, TX, 2005.

[LW03] J. Liu and W. Wang.
OP-Cluster: Clustering by tendency in high dimensional spaces.
In Proceedings of the 3th International Conference on Data Mining (ICDM), Melbourne, FL, 2003.

Literature

[MK03] T. M. Murali and S. Kasif.
Extracting conserved gene expression motifs from gene expression data.
In Proceedings of the 8th Pacific Symposium on Biocomputing (PSB), Maui, HI, 2003.

[MO04] S. C. Madeira and A. L. Oliveira.
Biclustering algorithms for biological data analysis: A survey.
IEEE Transactions on Computational Biology and Bioinformatics, 1(1):24–45, 2004.

[MSE06] G. Moise, J. Sander, and M. Ester.
P3C: A robust projected clustering algorithm.
In Proceedings of the 6th International Conference on Data Mining (ICDM), Hong Kong, China, 2006.

[NGC01] H.S. Nagesh, S. Goil, and A. Choudhary.
Adaptive grids for clustering massive data sets.
In Proceedings of the 1st SIAM International Conference on Data Mining (SDM), Chicago, IL, 2001.

[PBZ+06] A. Prelic, S. Bleuler, P. Zimmermann, A. Wille, P. Bühlmann, W. Guissem, L. Hennig, L. Thiele, and E. Zitzler.
A systematic comparison and evaluation of biclustering methods for gene expression data.
Bioinformatics, 22(9):1122–1129, 2006.

Literature

[Pfa07] J. Pfaltz.
What constitutes a scientific database?
In Proceedings of the 19th International Conference on Scientific and Statistical Database Management (SSDBM), Banff, Canada, 2007.

[PHL04] L. Parsons, E. Haque, and H. Liu.
Subspace clustering for high dimensional data: A review.
SIGKDD Explorations, 6(1):90–105, 2004.

[PJAM02] C. M. Procopiuc, M. Jones, P. K. Agarwal, and T. M. Murali.
A Monte Carlo algorithm for fast projective clustering.
In Proceedings of the ACM International Conference on Management of Data (SIGMOD), Madison, WI, 2002.

[PTTF02] E. Parros Machado de Sousa, C. Traina, A. Traina, and C. Faloutsos.
How to use fractal dimension to find correlations between attributes.
In Proc. KDD-Workshop on Fractals and Self-similarity in Data Mining: Issues and Approaches, 2002.

[PZC+03] J. Pei, X. Zhang, M. Cho, H. Wang, and P. S. Yu.
MaPle: A fast algorithm for maximal pattern-based clustering.
In Proceedings of the 3th International Conference on Data Mining (ICDM), Melbourne, FL, 2003.

Literature

[RRK04] U. Rückert, L. Richter, and S. Kramer.
Quantitative association rules based on half-spaces: an optimization approach.
In Proceedings of the 4th International Conference on Data Mining (ICDM), Brighton, U.K., 2004.

[SLGL06] K. Sim, J. Li, V. Gopalkrishnan, and G. Liu.
Mining maximal quasi-bicliques to co-cluster stocks and financial ratios for value investment.
In Proceedings of the 6th International Conference on Data Mining (ICDM), Hong Kong, China, 2006.

[SMD03] Q. Sheng, Y. Moreau, and B. De Moor.
Biclustering microarray data by Gibbs sampling.
Bioinformatics, 19(Suppl. 2):ii196–ii205, 2003.

[STG+01] E. Segal, B. Taskar, A. Gasch, N. Friedman, and D. Koller.
Rich probabilistic models for gene expression.
Bioinformatics, 17(Suppl. 1):S243–S252, 2001.

[SZ05] K. Sequeira and M. J. Zaki.
SCHISM: a new approach to interesting subspace mining.
International Journal of Business Intelligence and Data Mining, 1(2):137–160, 2005.

Literature

[TSS02] A. Tanay, R. Sharan, and R. Shamir.
Discovering statistically significant biclusters in gene expression data.
Bioinformatics, 18 (Suppl. 1):S136–S144, 2002.

[TXO05] A. K. H. Tung, X. Xu, and C. B. Ooi.
CURLER: Finding and visualizing nonlinear correlated clusters.
In Proceedings of the ACM International Conference on Management of Data (SIGMOD), Baltimore, ML, 2005.

[Web01] G. I. Webb.
Discovering associations with numeric variables.
In Proceedings of the 7th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), San Francisco, CA, pages 383–388, 2001.

[WLKL04] K.-G. Woo, J.-H. Lee, M.-H. Kim, and Y.-J. Lee.
FINDIT: a fast and intelligent subspace clustering algorithm using dimension voting.
Information and Software Technology, 46(4):255–271, 2004.

[WWYY02] H. Wang, W. Wang, J. Yang, and P. S. Yu.
Clustering by pattern similarity in large data sets.
In Proceedings of the ACM International Conference on Management of Data (SIGMOD), Madison, WI, 2002.

Literature

[YWWY02] J. Yang, W. Wang, H. Wang, and P. S. Yu.
δ-clusters: Capturing subspace correlation in a large data set.
In Proceedings of the 18th International Conference on Data Engineering (ICDE), San Jose, CA, 2002.