Diffrent models:

Partitioning Methods

- Means
  - K-means
    - Användning:
      - Given data and k
      - Select start point
      - Ge varje punt till ett kluster
      - Updatera varje k
      - Upprepa 3 och 4 till konvergera
    - Styrka:
      - O(tkn) (fast)
    - Kommentar:
      - Hittar ofta local optimum isället för global optimum
    - Svaghet:
      - Kan inte hantera categorisk data så bra
      - Specificera k
      - Känslig mot uteliggare
    - Lösning:
      - Använd K-medoids istället

- K-medoids
  - PAM
    - Användning:
      - Given data and k
      - Select k start medoids
      - Ge varje punt till ett kluster
      - Beräkna kostnaden att bytta ut en var medoidsen (behålla en och bytta ut den andra vid k=2). Se exempel i anteckningarna
      - Ge varje punkt till den närmaste medoidsen
      - Upprepa till den minsta kostnaden är nåd
    - Styrka
      - Kan hantera uteliggare bra
      - Funkar bra på stora dataset
    - Svaghet:
      - Långsam O(k(n-k)^2)
    - Lösning:
      - Använd CLARA
  - CLARA
    - Användning:
      - Given data, k and n
      - Sample S
      - Perform pam on the sample S
      - Ge de som ej tillhör S en klass och beräkna avg dissimilarity
      - Om detta värde är mindre än förra, använd dessa medioids
      - Upprepa detta n gånger
  - CLARANS
    - Användning:
      - Given data,
        Numlocal= number of local minima to be found
        Maxneighbor= maximum number of neighbors to compare
      - A Clustering Algorithm based on Randomized Search
    - Styrka:
      - Mer effektiv än både PAM och CLARA

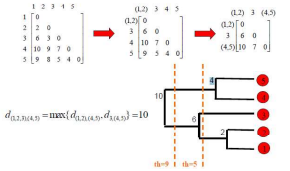
Hierarchical Methods

Typical Alternatives to Calculate the Distance between Clusters

- Single link: smallest distance between an element in one cluster and an element in the other, i.e., $dis(K_i, K_j) = min(t_{ip}, t_{jq})$
- Complete link: largest distance between an element in one cluster and an element in the other, i.e., $dis(K_i, K_j) = max(t_{ip}, t_{jq})$
- Average: avg distance between an element in one cluster and an element in the other, i.e., $dis(K_i, K_j) = avg(t_{ip}, t_{jq})$
- Centroid: distance between the centroids of two clusters, i.e., $dis(K_i, K_j) = dis(C_i, C_j)$
- Medoid: distance between the medoids of two clusters, i.e., $dis(K_i, K_j) = dis(M_i, M_j)$
  - Medoid: one chosen, centrally located object in the cluster

- Distance matrix
  - Användning:
    - Beräkna en anvståndsmatris
    - Alla är ett eget kluster från början
    - Slå ihop de två kluster som när närmast (givet hur man beräknar avståndet mellan två kluster)
    - Gör detta fram till att det endast finns ett kluster.
    - Kolla i trädet efter stort hopp och bryt där

Complete-link Clustering Example



- AGNES
  - Användning:
    - Som Distance matrix men att den använder singel link

- DIANA
  - Användning:
    - Inverse order av AGNES
    - Alla tillhör ett kluster och i slutet är alla observationer ett eget kluster

- Svaghet hos Agglomerative clustring: This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
  - Långsam O(n^2)
  - Kan aldrig ogöra det den redan har gjort

- Integration of hierarchical with distance-based clustering
  - BRICH: uses CF-tree and incrementally adjusts the quality of sub-clusters
    - Styrka
      - O(linjer)
    - Svaghet:
      - Bara numeric data
      - Känslig mot ordningen I data
    - Användning:
      - CF-tree
      - Read more....
  - ROCK: clustering categorical data by neighbor and link analysis
    - Användning:
      - Random sample from data
      - Hierarchical clustering with links using goodness measure of merging
      - Label data in disk: a point is assigned to the cluster for which it has the most neighbors after normalization
  - CHAMELEON: hierarchical clustering using dynamic modeling
    - Användning:
      - Two clusters are merged only if the interconnectivity and closeness (proximity) between two clusters are high relative to the internal interconnectivity of the clusters and closeness of items within the clusters
        - Use a graph partitioning algorithm: cluster objects into a large number of relatively small sub-clusters
        - Use an agglomerative hierarchical clustering algorithm: find the genuine clusters by repeatedly combining these sub-clusters

Density-Based Methods

- DBSCAN
  - Styrka:
    - Inte känslig mot uteliggare
    - Kan hitta alla möjliga former av kluster
  - Svaghet:
    - Känslig mot parametrarna Eps och MinPts. Ändras dessa lite kan hela modelen ändras.
  - Användning:
    - Givet data, Eps och MinPts
    - Gå en observation i taget
    - Kolla om observationen uppfyller antalet grannar. Gör den det bilda ett kluster.
    - Fortsätt till att punkter har blivit kollade
- OPTICS

Typ av variabler

Type of data in clustering analysis

- Interval-scaled variables
  - Continuous measurements (weight, temperature, ...)
- Binary variables
  - Variables with 2 states (on/off, yes/no)
- Nominal variables
  - A generalization of the binary variable in that it can take more than 2 states (color/red,yellow,blue,green)
- Ordinal
  - ranking is important (e.g. medals(gold,silver,bronze))
- Ratio variables
  - a positive measurement on a nonlinear scale (growth)
- Variables of mixed types

Euclidean och manhattan

Distances between objects

- Euclidean distance:

$$d(i,j)=\sqrt{(|x_{i1}-x_{j1}|^2+|x_{i2}-x_{j2}|^2+...+|x_{ip}-x_{jp}|^2)}$$

- Manhattan distance:

$$d(i,j)=|x_{i1}-x_{j1}|+|x_{i2}-x_{j2}|+...+|x_{ip}-x_{jp}|$$

where $i = (x_{i1}, x_{i2}, ..., x_{ip})$ and $j = (x_{j1}, x_{j2}, ..., x_{jp})$ are two p-dimensional data objects,

Binary variables

Binary Variables

- symmetric binary variables: both states are equally important; 0/1
- asymmetric binary variables: one state is more important than the other (e.g. outcome of disease test); 1 is the important state, 0 the other

Symmetric

Distance measure for symmetric binary variables

Object i \ Object j	1	0	sum
1	a	b	$a+b$
0	c	d	$c+d$
sum	$a+c$	$b+d$	p

$$d(i,j)=\frac{b+c}{a+b+c+d}$$

Asymmetric

Distance measure for asymmetric binary variables

Object i \ Object j	1	0	sum
1	a	b	$a+b$
0	c	d	$c+d$
sum	$a+c$	$b+d$	p

$$d(i,j)=\frac{b+c}{a+b+c}$$

Jaccard coefficient = 1- d(i,j) = $sim_{Jaccard}(i,j)=\frac{a}{a+b+c}$

Dissimilarity binary

Dissimilarity between Binary Variables

- Example

Name	Gender	Fever	Cough	Test-1	Test-2	Test-3	Test-4
Jack	M	Y	N	P	N	N	N
Mary	F	Y	N	P	N	P	N
Jim	M	Y	P	N	N	N	N

  - gender is a symmetric attribute
  - the remaining attributes are asymmetric binary
    → distance based on these
  - let the values Y and P be set to 1, and the value N be set to 0

$$d(jack,mary)=\frac{0+1}{2+0+1}=0.33$$
$$d(jack,jim)=\frac{1+1}{1+1+1}=0.67$$
$$d(jim,mary)=\frac{1+2}{1+1+2}=0.75$$

Nominal distance and categorical

Nominal or Categorical Variables

- Method 1: Simple matching
  - m: # of matches, p: total # of variables

$$d(i,j)=\frac{p-m}{p}$$

- Method 2: use a large number of binary variables
  - creating a new asymmetric binary variable for each of the M nominal states

Ordinal distance

Ordinal Variables

- An ordinal variable can be discrete or continuous
- Order is important, e.g., rank
- Can be treated like interval-scaled
  - replace x_{if} by their rank $r_{if} \in \{1,...,M_f\}$
  - map the range of each variable onto [0, 1] by replacing i-th object in the f-th variable by

$$z_{if}=\frac{r_{if}-1}{M_f-1}$$

  - compute the dissimilarity using methods for interval-scaled variables

Partitioning clustering:

1. Clustering by partitioning (2p+2p=4p)

a. Describe the principles and ideas regarding PAM.
- Give a sketch of the algorithm.
- Define swapping cost. = TCih = Σj summan av alla COSTS

b. Describe the principles and ideas regarding CLARA. TCih = ΣjCjih
- Give a sketch of the algorithm.
- What are the strengths and weaknesses of CLARA?

1. Clustering by partitioning (2p+2p=4p)

a. Describe the principles and ideas regarding PAM.
- Give a sketch of the algorithm.
- Define swapping cost.

b. Given the graph representation of the clustering problem where n is the number of data points and k is the number of clusters.
(i) What does a node represent?
(ii) How can this graph be used for finding a solution for the clustering problem?
(iii) When are two nodes neighbors and how many neighbors does a node have?
(iv) Considering PAM, CLARA and CLARANS, the graph for which algorithm/algorithms contains/contain the most nodes? Explain.

1. Clustering by partitioning (2p+3p=5p)

a. Given the graph representation of the clustering problem where n is the number of data objects and k is the number of clusters.
(i) What does a node represent?
(ii) When are two nodes neighbors and how many neighbors does a node have?
(iii) Considering PAM, CLARA and CLARANS, the graph for which algorithm/algorithms contains/contain the most nodes? Explain.
(iv) Which of PAM, CLARA and CLARANS guarantees to find a global optimum?

b. Given the data set {0, 2, 3, 8}. Assume we use Euclidean distance and k = 2. Draw the graph representation of the clustering problem. Then start at an arbitrary node and show one iteration of the PAM algorithm on the graph. Give all steps in the computation and show at what node that iteration ends.

1. Clustering by partitioning (3p+1p=4p)

a. Given the data set {0, 3, 4, 10}. Assume we use Euclidean distance and k = 2. Draw the graph representation of the clustering problem. Then start at an arbitrary node and show one iteration of the PAM algorithm on the graph. Give all steps in the computation and show at what node that iteration ends.

b. Why is PAM more robust than K-means in the presence of outliers?

1. Clustering by partitioning (2p+3p=5p)

a. Given the graph representation of the clustering problem where n is the number of data objects and k is the number of clusters.
   i. What does a node represent?
   ii. When are two nodes neighbors and how many neighbors does a node have?
   iii. Considering PAM, CLARA and CLARANS, the graph for which algorithm/algorithms contains/contain the most nodes? Explain.
   iv. Which of PAM, CLARA and CLARANS guarantees to find a global optimum?

b. Describe the principles and ideas regarding PAM.
   i. Describe the algorithm.
   ii. Define swapping cost.
   iii. Draw an example of a data set in two dimensions where the swapping cost TCih is 0 and one where the swapping cost TCih is strictly negative.

Hierarchical

2. Hierarchical clustering (4p)

Describe the principles and ideas regarding BIRCH.
- Give a sketch of the algorithm.
- Explain Clustering Feature Vector. Given a cluster with the data points (0,0), (1,1) and (2,2), what is its clustering feature vector?
- Explain what a CF-tree is and how it is used in BIRCH. How is a CF-tree traversed?
- What parameters are used as input?

2. Hierarchical clustering (3+2=5p)

a. Describe the principles and ideas regarding Agglomorative Hierarchical Clustering. Show the different steps of the algorithm using the dissimilarity matrix below and complete link clustering. Give partial results after each step.

	1	2	3	4	5
1	0				
2	5	0			
3	9	10	0		
4	3	2	6	0	
5	7	1	4	8	0

b. Describe the principles and ideas regarding the CHAMELEON algorithm. Explain the major steps.

2. Hierarchical clustering (3p)

Describe the principles and ideas regarding Agglomorative Hierarchical Clustering. Show the different steps of the algorithm using the dissimilarity matrix below and *single* link clustering. Give partial results after each step.

	1	2	3	4	5
1	0				
2	8	0			
3	2	10	0		
4	1	5	7	0	
5	9	3	6	4	0

2. Hierarchical clustering (3+2=5p)

a. Describe the principles and ideas regarding BIRCH by answering the following:
- Give a sketch of the algorithm.
- Explain Cluster Feature Vector. Given a cluster with the data points (1,2), (1,3) and (2,2), what is its cluster feature vector?
- Explain what a CF-tree is and how it is used in BIRCH.
- What parameters are used as input?

b. For the ROCK algorithm:
- What is Link defined between two clusters?
- Give and explain the goodness measure in ROCK. Also explain how it is used.

2. Hierarchical clustering (3p+1p=4p)

a. Describe the principles and ideas regarding BIRCH by answering the following:
   i. Give a sketch of the algorithm.
   ii. Explain Cluster Feature Vector. Given a cluster with the data points (1,2), (1,3) and (2,2), what is its cluster feature vector?
   iii. Explain what a CF-tree is and how it is used in BIRCH.
   iv. What parameters are used as input?

b. For the ROCK algorithm:

Given the *similarity matrix* below. What is link(A,B) if the threshold is 0.6?

	A	B	C	D	E
A	1				
B	0.9	1			
C	0.8	0.7	1		
D	0.1	0.3	0.6	1	
E	0	0.2	0.4	0.5	1

Density

3. Clustering categorical data (4p)

Describe the principles and ideas regarding the ROCK algorithm. Within your description, make sure to give a sketch of the algorithm and to define **and** give examples for neighbor, common neighbor, link for objects, link for clusters, and G (goodness measure).

3. Clustering categorical data (2+1=3p)

a. Describe the principles and ideas regarding the DBSCAN algorithm. In your description, make sure to give a sketch of the algorithm and to define core point, direct density-reachable, density-reachable, and density-connected.
b. What is the main idea behind OPTICS?

3. Density-based clustering (2p+1p+1p=4p)

a. Describe the principles and ideas regarding the DBSCAN algorithm. In your description, make sure to give a sketch of the algorithm and to define core point, direct density-reachable, density-reachable, and density-connected.
b. DBSCAN: Consider the following statement: if p is density-connected to q wrt Eps and Minpts then p is density-reachable from q wrt Eps and Minpts. Is this statement true? If yes, then prove. If no, then give a counterexample.
c. What is the main idea behind OPTICS?

3. Density-based clustering (2+1=3p)

a. Describe the principles and ideas regarding the DBSCAN algorithm. In your description, make sure to give a sketch of the algorithm and to define core point, direct density-reachable, density-reachable, and density-connected.
b. What is the relationship between DBSCAN and OPTICS?

3. Density-based clustering (2p)

Describe the principles and ideas regarding the DBSCAN algorithm. In your description, make sure to describe the algorithm and to define core point, direct density-reachable, density-reachable, and density-connected.

Distance or extra

4. Different types of data and their distance measures (2p+1p+1p=4p)

a. What is the distance between Item K and Item L?

	A	B	C	D	E	F	G
Item K	(20000,1)	(1,2)	Y	N	Y	N	55
Item L	(20000,100)	(3,5)	Y	N	Y	N	no-value-available

Attribute A is interval-based and Euclidean distance is used.
Attribute B is interval-based and Manhattan distance is used.
Attributes C and D are binary symmetric variables.
Attributes E and F are binary asymmetric variables.
Attribute G is interval-based.

b. In the vector model for information retrieval documents are represented by vectors with positive real numbers. How is the similarity between two vectors defined?

c. Show how an interval-based measure can be defined for ordinal variables.

4. Data mining concepts (1p+1p+1p+1p=4p)

a. The purpose of data mining is to extract interesting patterns from a huge amount of data. When is a pattern 'interesting' in this case?
b. Data in the real world can be dirty. Give 3 reasons and an example for each.
c. Show how an interval-based distance measure can be defined for ordinal variables.
d. Show how a distance measure can be defined for categorical (or nominal) variables.

4. Different types of data and their distance measures (2p+2p=4p)

a. What is the distance between Item K and Item L?

	A	B	C	D	E	F	G
Item K	(20,1)	(2,2)	Y	N	Y	N	8
Item L	(20,100)	(3,6)	N	N	N	N	no-value-available

Attribute A is interval-based and Euclidean distance is used.
Attribute B is interval-based and Manhattan distance is used.
Attributes C and D are binary symmetric variables.
Attributes E and F are binary asymmetric variables.
Attribute G is interval-based.

b. Assume we have categorical data. One method to define a distance between two data objects is (p-m)/p where p is the total number of categorical variables and m is the number of categorical variables for which there is a match between the objects. A second method is to introduce a new asymmetric binary variable for each of the possible values for each of the categorical variables. Give a formula for the distance between two objects in the second method in terms of p and m (where p and m have the same meaning as above; i.e. p is the number of categorical variables - *not* the number of introduced binary variables, and m is the number of matches in the categorical variables).

4. Different types of data and their distance measures (2p+2p=4p)

a. What is the distance between Item K and Item L?

	A	B	C	D	E	F	G
Item K	(40,50)	(2,1)	Y	N	Y	N	8
Item L	(40,55)	(1,5)	Y	N	Y	N	no-value-available

Attribute A is interval-based and Euclidean distance is used.
Attribute B is interval-based and Manhattan distance is used.
Attributes C and D are binary symmetric variables.
Attributes E and F are binary asymmetric variables.
Attribute G is interval-based.

b. Assume we have categorical data. One method to define a distance between two data objects is (p-m)/p where p is the total number of categorical variables and m is the number of categorical variables for which there is a match between the objects. A second method is to introduce a new asymmetric binary variable for each of the possible values for each of the categorical variables. Give a formula for the distance between two objects in the second method in terms of p and m (where p and m have the same meaning as above; i.e. p is the number of categorical variables - *not* the number of introduced binary variables, and m is the number of matches in the categorical variables).

4. Different types of data and their distance measures (3p+1p+1p=5p)

a. Asymmetric binary variables.
   i. Give and explain the distance measure for objects with asymmetric binary variables using contingency tables.
   ii. Give and explain the distance measure for objects with variables of mixed types.
   iii. Can the formula in question b also be used for objects with only asymmetric variables? If no, explain why. If yes, state whether you would get the same results as with the method in question a and explain why or why not.

b. In the vector model for information retrieval documents are represented by vectors with positive real numbers. How is the similarity between two vectors defined?

c. Show how an interval-based measure can be defined for ordinal variables.

6. Apriori algorithm (2p+1p+1p+1p=5p)

a. Describe the Apriori algorithm, i.e. describe how it works in general not in a particular example.

b. Describe how to incorporate a monotone constraint into the Apriori algorithm.

c. Describe how to incorporate an antimonotone constraint into the Apriori algorithm.

d. Give an example where the Apriori algorithm fails, i.e. there is a frequent itemset in your example that is not discovered by the Apriori algorithm.

6. Apriori algorithm (2p+1p+1p+1p=5p)

d. Describe the Apriori algorithm, i.e. describe how it works in general not in a particular example.

e. Describe how to incorporate a monotone constraint into the Apriori algorithm.

f. Describe how to incorporate an antimonotone constraint into the Apriori algorithm.

a. Prove formally the correctness of the Apriori algorithm.

5. Apriori algoritm (2p+1p+1p+2p=6p)

a. Run the Apriori algorithm on the following transactional database with minimum support equal to one transaction. Explain step by step the execution.

Transaction id	Items
1	A, B, C
2	X, Y, Z
3	A, Y, C
4	X, B, Z

b. Repeat the exercise 5a with the following additional constraint: Find the frequent itemsets that do not contain the itemset AB. Make clear when and how the constraint is used. Incorporate the constraint into the algorithm, i.e. do not simply run the algorithm and afterwards consider the constraint.

c. Let the items A, B, C, X, Y and Z have a price of respectively -3, -2, -1, 1, 2 and 3 units. Repeat the exercise 5a with the following additional constraint: Find the frequent itemsets whose most expensive item has positive price. Make clear when and how the constraint is used. Incorporate the constraint into the algorithm, i.e. do not simply run the algorithm and afterwards consider the constraint.

d. Sketch a proof of the correctness of the Apriori algorithm.

5. Apriori algoritm (2p+1p+1p+1p=5p)

a. Run the Apriori algorithm on the following transactional database with minimum support equal to one transaction. Explain step by step the execution.

Transaction id	Items
1	A, B, C
2	X, Y, Z
3	A, Y, C
4	X, B, Z

b. Repeat the exercise 5a with the following additional constraint: Find the frequent itemsets that contain the item A. Make clear when and how the constraint is used. Incorporate the constraint into the algorithm, i.e. do not simply run the algorithm and afterwards consider the constraint.

c. Let the items A, B, C, X, Y and Z have a price of respectively -3, -2, -1, 1, 2 and 3 units. Repeat the exercise 5a with the following additional constraint: Find the frequent itemsets whose range is smaller than 3 (recall that the range is the price of the most expensive item minus the price of the cheapest item). Make clear when and how the constraint is used. Incorporate the constraint into the algorithm, i.e. do not simply run the algorithm and afterwards consider the constraint.

d. Sketch a proof of the correctness of the Apriori algorithm.

5. Apriori algoritm (2p+1p+1p+2p=6p)

a. Run the Apriori algorithm on the following transactional database with minimum support equal to one transaction. Explain step by step the execution.

Transaction id	Items
1	A, B, C
2	X, Y, Z
3	A, Y, C
4	X, B, Z

b. Repeat the exercise 5a with the following additional constraint: Find the frequent itemsets that contain the item A. Make clear when and how the constraint is used. Incorporate the constraint into the algorithm, i.e. do not simply run the algorithm and afterwards consider the constraint.

c. Let the items A, B, C, X, Y and Z have a price of respectively -3, -2, -1, 1, 2 and 3 units. Repeat the exercise 5a with the following additional constraint: Find the frequent itemsets whose range is smaller than 3 (recall that the range is the price of the most expensive item minus the price of the cheapest item). Make clear when and how the constraint is used. Incorporate the constraint into the algorithm, i.e. do not simply run the algorithm and afterwards consider the constraint.

d. Sketch a proof of the correctness of the Apriori algorithm.

5. FP grow algoritm (2p+2p+2p=6p)

a. Run the FP grow algorithm on the following transactional database with minimum support equal to one transaction. Explain step by step the execution.

Transaction id	Items
1	A, B, C
2	X, Y, Z
3	A, Y, C
4	X, B, Z

b. Repeat the exercise 5a with the following additional constraint: Find the frequent itemsets that do not contain the item Z. Make clear when and how the constraint is used. Incorporate the constraint into the algorithm, i.e. do not simply run the algorithm and afterwards consider the constraint.

c. Let the items A, B, C, X, Y and Z have a price of respectively -3, -2, -1, 1, 2 and 3 units. Repeat the exercise 5a with the following additional constraint: Find the frequent itemsets whose most expensive item has positive price. Make clear when and how the constraint is used. Incorporate the constraint into the algorithm, i.e. do not simply run the algorithm and afterwards consider the constraint.

5. FP grow algoritm (2p+2p+2p=6p)

a. Describe the FP grow algorithm, i.e. describe how it works in general not in a particular example.

b. Describe how to incorporate a monotone constraint into the FP grow algorithm.

c. Describe how to incorporate an antimonotone constraint into the FP grow algorithm.

6. FP grow algorithm (2p+1p+1p+1p=5p)

Stands for ... It is a scalable technique for m frequent pattern in a databa

a. Describe the FP grow algorithm.

b. Explain how you incorporate a monotone constraint into the FP grow algorithm.

c. Explain how you incorporate an antimonotone constraint into the FP grow algorithm.

e. What is the main advantage that the FP grow algorithm has over the Apriori algorithm ?

6. FP growth algorithm (2p+1p+1p+1p=5p)

a. Describe the FP growth algorithm. Do not use examples.

b. Explain how you incorporate a monotone constraint into the FP growth algorithm.

f. Explan how you incorporate an antimonotone constraint into the FP growth algorithm.

d. What is the main advantage of the FP growth algorithm over the Apriori algorithm ?

6. FP grow algorithm (2p+1p=3p)

a. Explain how you incorporate monotone and antimonotone constraints in the FP grow algorithm.

b. What is the main advantage of the FP grow algorithm over the Apriori algorithm ?

7. Constraints and lift (1p+1p+1p=3p)

Are the following statements true or false ?

a. If the itemsets AB and BC are frequent, then the itemset AC is frequent too.
b. Every constraint is monotone, antimonotone, convertible monotone or convertible antimonotone.
c. Adding items to the antecedent of an association rule increases the lift of the rule.

7. Constraints and lift (1p+1p+1p=3p)

a. Describe what a monotone constraint is. Do not give a particular example. Describe the constraint in general terms.

b. Describe what a convertible antimonotone constraint is.

c. Describe what the lift of an association rule is.

7. Constraints and lift (1p+1p+1p=3p)

a. Prove that a constraint C cannot be both monotone and antimonotone, unless C(A)=C(B) for all itemsets A and B. Note that you have to prove the statement and thus it does not suffice with giving an example.

b. Apply the Simple algorithm to the frequent itemset XBZ on the database in exercise 5 in order to find association rules with confidence greater or equal than 50 %.

c. Give an example of an association rule with lift greater than one and another example of a rule with lift smaller than one.

7. Constraints and lift (2p+1p+1p=4p)

a. Give the definitions of monotone, antimonotone, and convertible monotone and antimonotone constraints.

b. Apply the Simple algorithm to the frequent itemset XBZ on the database in exercise 5 in order to find association rules with confidence greater than 50 %.

c. Give an example of an association rule with lift greater than one and another example of a rule with lift smaller than one.

7. Constraints (2p+2p=4p)

a. Give an example of convertible monotone and convertible antimonotone constraints that are not monotone and antimonotone, respectively.

b. Show that your constraints are really convertible monotone and antimonotone. Do not give examples to show it, i.e. provide an abstract and formal argument.

8. Rule generation (2p)

Apply the Simple algorithm to the frequent itemset XBZ on the database in exercise 5 in order to find association rules with confidence greater than 50 %.