Section 0 - Grading

Exam questions are a mix of multiple-choice and are written formulated questions in accordance with the course instruction handbook. They are only based on the topics that were covered in lectures as previously indicated. The exam comprises of:

- a single explanatory ungraded introductory section (0 points),
- a section on data and preparation (10 points),
- a section on associations (30 points),
- a section on clustering (30 points),
- a section on classification (30 points).

There is digital calculator available to you with Inspera.

Grading:
There are no constraints on how much one scores from each section however in total per course instruction documentation:
5 [100-90) Pass with distinction: Outstanding performance with only minor errors.
4 [90-70) Pass with credit: Generally sound work with a number of notable errors.
3 [70-50) Pass: Fair but with significant shortcomings.
U [50-0) Fail: Considerable further work is required.

Good luck!

Lycka till!

1 Section 1 - Attributes

Classify the following attributes as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Most cases may have more than one interpretation. [2 points].

Brightness as measured by people's judgments.

- [ ] Binary
- [ ] Discrete
- [ ] Continuous
- [ ] Qualitative
- [ ] Nominal
- [ ] Ordinal
- [ ] Quantitative
- [ ] Interval
- [ ] Ratio

Brightness as measured by a light meter.

- [ ] Binary
- [ ] Discrete
- [ ] Continuous
- [ ] Qualitative
- [ ] Nominal
- [ ] Ordinal
- [ ] Quantitative
- [ ] Interval
- [ ] Ratio

Ability to pass light in terms of the following values: opaque, translucent, transparent.

- [ ] Binary
- [ ] Discrete
- [ ] Continuous
- [ ] Qualitative
- [ ] Nominal
- [ ] Ordinal
- [ ] Quantitative
- [ ] Interval
- [ ] Ratio

Is interval +/- operations also valid for ratio operations?

- ( ) True
- ( ) False

Totalpoäng: 2

2 Section 1 - Similarity

For the following vectors, x = (0, 1, 0, 1), y = (1, 0, 1, 0), calculate the indicated similarity or distance measures given the equations: equationssection1similarity and fill the values in the boxes [2 points].

1. cos(x,y) =
2. corr(x,y) =
3. Euclidean(x,y) =
4. Jacard(x,y) =

Totalpoäng: 2

3 Section 1 - Standardization

Standardization of datasets is a common requirement for data mining.

A) One standardization method is scaling features to lie between a given minimum and maximum value, often between zero and one, or so that the maximum absolute value of each feature is scaled to unit size where X is a data vector and X.min defines minimal value and X.max defines maximal value in the data set X.

X_std = (X - X.min) / (X.max - X.min)

B) Another standardization method is normalization where the process is of scaling individual samples to have unit norm. Maximal normalization is normalize all the values by the maximal value of the vector.

Given, $Data = \begin{bmatrix} 0 & 0 & 2 \\ 1 & 2 & 0 \\ 2 & 2 & 1 \end{bmatrix}$**, please answer the following questions. [2 points]**

Column max normalization will produce same results as min max standardization.

- True
- False

Row max normalization will produce same results as min max standardization.

- True
- False

Totalpoäng: 2

4 Section 1 - Principal Component Analysis

No calculator is needed for this question. Some of the most common approaches for dimensionality reduction, particularly for continuous data, use techniques from linear algebra to project the data from a high-dimensional space into a lower-dimensional space. Principal Components Analysis (PCA) is a linear algebra technique for continuous attributes that finds new attributes (principal components) that (1) are linear combinations of the original attributes, (2) are orthogonal (perpendicular) to each other, and (3) capture the maximum amount of variation in the data. Please answer the following questions to demonstrate your understanding of PCA [2 points]

Please mark the first ranked principal component (eigenvector) and its eigenvalue(variance) when we assume that we have the following covariance matrix for the data X. No calculator is needed.

$$XX^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

- Eigenvector= [0 0 1]^T and eigenvalue=3
- Eigenvector= [1 0 0]^T and eigenvalue=3
- Eigenvector= [0 1 0]^T and eigenvalue=3
- Eigenvector= [1 2 3]^T and eigenvalue=1

Are these statements true or false?
(A) Learning lower dimensional representation can save memory usage.
(B) Learning lower dimensional representationcan remove redundancies and noises in data.

- A)True, B)False
- A)False, B)True
- A)True, B)True
- A)False, B)False

Are these statements true or false?

(A) When we use PCA, we need data to be labelled.

(B) PCA extracts the variance structure from high dimensional data such that the variance of projected data is minimized.

- A) False, B) False
- A) True, B) False
- A) False, B) True
- A) True, B) True

Are these statements true or false?

(A) Ignoring the components of small eigenvalues will not lose information.

(B) With better parameter tuning, we can get a better first principal component which maximizes the variability more precisely.

- A) False B) False
- A) False B) True
- A) True B) False
- A} True B) True

Totalpoäng: 2

5 Section 1 - Noise Outliers

Distinguish between noise and outliers by answering following questions [2 points].

Is noise ever interesting or desirable for data (disregard data mining privacy) ?

- Yes
- No

Is outliers ever interesting or desirable?

- No
- Yes

Can noise objects be outliers?

- Yes
- No

Are noise objects always outliers?

- No
- Yes

Are outliers always noise objects?

- Yes
- No

Totalpoäng: 2

6 Section 2 - Association rules basics

$$\text{Support, } s(X \longrightarrow Y) = \frac{\sigma(X \cup Y)}{N};$$
$$\text{Confidence, } c(X \longrightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}.$$

$\sigma() = the\ number\ of\ transactions\ that\ contain\ a\ particular\ itemset.$
$N = number\ of`transactions$

Given the association rules confidence and support definitions, please answer the following questions for the given table:

Example of market basket transactions.

Customer ID	Transaction ID	Items Bought
1	0001	$\{a, d, e\}$
1	0024	$\{a, b, c, e\}$
2	0012	$\{a, b, d, e\}$
2	0031	$\{a, c, d, e\}$
3	0015	$\{b, c, e\}$
3	0022	$\{b, d, e\}$
4	0029	$\{c, d\}$
4	0040	$\{a, b, c\}$
5	0033	$\{a, d, e\}$
5	0038	$\{a, b, e\}$

a) Compute the support for itemsets s({e}) [ ] , s({b, d}) [ ] , and s({b, d, e}) [ ] by treating each transaction ID as a market basket.

b) Is confidence a symmetric measure? Please feel free to validate by using any rules that you like such as $\{b, d\} \rightarrow \{e\}\ and\ \{e\} \rightarrow \{b, d\}$ if you are not sure of the answer.

- ○ True
- ○ False

c) Repeat parts of (a) by treating each customer ID as a market basket. Each item should be treated as a binary variable (1 if an item appears in at least one transaction bought by the customer, and 0 otherwise.). Then please mark one of the below.

- ☐ s({b,d}) and s({b,d,e}) are the same in both cases
- ☐ s{{e}) are equal in both cases

d) Suppose s1 and c1 are the support and confidence values of an association rule r when treating each transaction ID as a market basket. Also, let s2 and c2 be the support and confidence values of r when treating each customer ID as a market basket. Are there any relationships between s1 and s2 or c1 and c2?

- Yes
- No

e) Mark all the correct answers for the confidence rules for $a \rightarrow \emptyset$ where $\emptyset$ denotes empty set.

- c(∅ → A) = s(∅ → A)
- c(∅ → A) ≠ s(∅ → A)
- c(A → ∅) = 100%.
- c(A → ∅) ≠100%.

Totalpoäng: 8

7 Section 2 - Association candidates and rules general

Consider the market basket transactions shown in Table:

Market basket transactions.

Transaction ID	Items Bought
1	{Milk, Beer, Diapers}
2	{Bread, Butter, Milk}
3	{Milk, Diapers, Cookies}
4	{Bread, Butter, Cookies}
5	{Beer, Cookies, Diapers}
6	{Milk, Diapers, Bread, Butter}
7	{Bread, Butter, Diapers}
8	{Beer, Diapers}
9	{Milk, Diapers, Bread, Butter}
10	{Beer, Cookies}

a) What is the maximum number of association rules that can be extracted from this data (including rules that have zero support) that is given by $R = 3^d - 2^{d+1} + 1$ that was derived/discussed during the lectures using binomial theorem where d is number of different items. R=

b) What is the maximum size of frequent itemsets that can be extracted (assuming minsup > 0)?

- 6
- 2
- 4
- 1

c) Find an itemset (of size 2 or larger) that has the largest support.

- Bread, Milk
- Diapers, Bread
- Butter, Milk
- Bread, Butter

d) Given the option of two equations

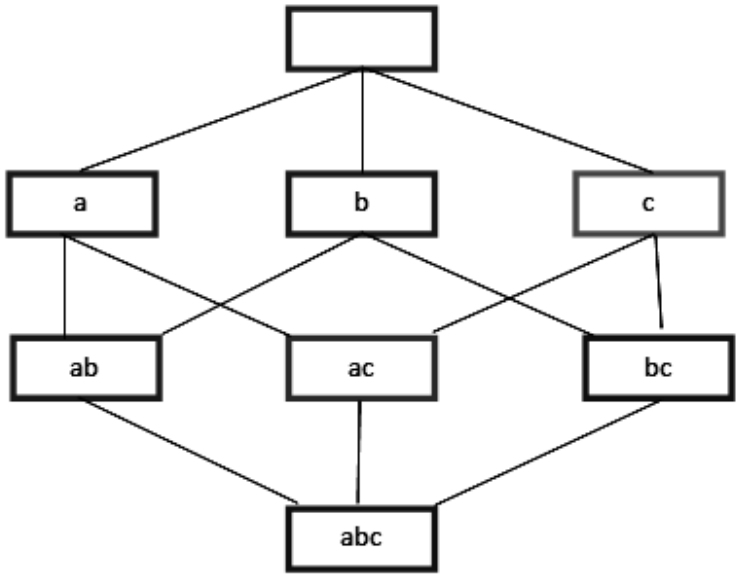
$$Combinations (Binomial\ Coefficient)\ C(n,r) = \binom{n}{k} = \frac{n!}{k!(n-k)!} Permutations\ P(n,r) = \frac{n!}{(n-k)!}$$

please choose the correct one between permutation or combination as discussed in the lectures and calculate maximum number of size-3 candidate itemsets that can be derived from this data set. Herein n=size of the itemset and k=size of the candidates and "!" is factorial for example 3!=3 x 2 x 1. Mark the correct answer

.

- 720
- 15
- 120
- 20

e) Given the below lattice structure for three items {a,b,c} candidates, answer the below questions and choose all the correct answers:



Please use the lattice structure to define correct answers.

- [ ] All the possible candidates are given by 2^n and this is derived from the binomial theorem. The total number of candidates are 8.
- [ ] Apriori algorithms aims to reduce the number of this candidates.
- [ ] The total rules R=12 is the highest candidate scenario for frequency counting.
- [ ] The total number of rules that can be generated from this item set is R as defined at part a) and it is 12.

Totalpoäng: 8

8 Section 2 - Frequent item set generation

$\sigma() = support\ count$
$i = items$
$k = k\ itemsets$
$minsup = minimum\ support$

Frequent itemset generation of the *Apriori* algorithm.

1: $k = 1$.
2: $F_k = \{ i \mid i \in I \wedge \sigma(\{i\}) \geq N \times minsup\}$. {Find all frequent 1-itemsets}
3: **repeat**
4: $k = k + 1$.
5: $C_k = \text{candidate-gen}(F_{k-1})$. {Generate candidate itemsets.}
6: $C_k = \text{candidate-prune}(C_k, F_{k-1})$. {Prune candidate itemsets.}
7: **for** each transaction $t \in T$ **do**
8: $C_t = \text{subset}(C_k, t)$. {Identify all candidates that belong to t.}
9: **for** each candidate itemset $c \in C_t$ **do**
10: $\sigma(c) = \sigma(c) + 1$. {Increment support count.}
11: **end for**
12: **end for**
13: $F_k = \{ c \mid c \in C_k \wedge \sigma(c) \geq N \times minsup\}$. {Extract the frequent k-itemsets.}
14: **until** $F_k = \emptyset$
15: Result $= \bigcup F_k$.

Which of the following 3-itemsets is in the list of candidate 3-itemsets generated by the APRIORI algorithm, if the list of frequent 2-itemsets is {i1, i2}, {i1, i3}, {i2, i4}, {i4, i3}, {i3, i2}?

- {i1, i2, i3}
- {i2, i3, i4}
- There are no candidate 3-itemsets
- All the subsets of {i1, i2, i3, i4} with three elements
- {i1, i2, i3}, {i2, i3, i4}
- None of the previous answers
- I cannot answer using only the information provided in the question

The hash function, h(p) = (p − 1) mod (x), where mode refers to the modulo (remainder) operator determines which branch of the current node should be followed and it defines the hash tree. In an APRIORI algorithm that employs hash function for support count, consider a hashing function with two branches for mod(2) and with maximum itemset size 6: items 1, 3, 5 are associated to the left branch, and items 2, 4, 6 are associated to the right branch. A hash tree with maximum node capacity 2 is generated to store the following candidate 2-itemsets: (1,2), (1,4), (2,3), (2,5), (2,6). How many itemsets are stored in the right-most non-empty leaf of the hash tree?

- 4
- 5
- 2
- 1
- 3
- 0
- None of the previous answers.

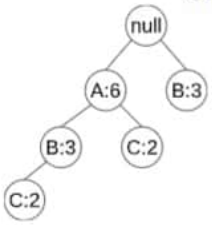
Totalpoäng: 2

9 Section 2 - FP- tree

Consider the following transactions:
• A B C
• A B
• A C
• A C D
How many nodes labeled A are present in the corresponding FP-tree?

- 3
- 2
- 1
- 4

Consider the following FP-tree:



What number is associated to node A in the conditional FP-tree on C assuming minimum support count 2?

- None of the above
- 4
- 3
- 2

Totalpoäng: 6

10 Section 2 - Evaluation of association patterns

$$Lift = \frac{P(Y \mid X)}{P(Y)}$$

$$Interest = \frac{P(X,Y)}{P(X)P(Y)}$$

lift is used for rules while interest is used for itemsets

$s(X,Y) = P(X,Y)$ *where s denotes support*

Suppose we have market basket data consisting of 100 transactions and 20 items. If the support for item a is 25%, the support for item b is 90% and the support for itemset {a, b} is 20%. Let the support and confidence thresholds be 10% and 60%, respectively.

a) Compute the confidence of the association rule {a} → {b}. Is the rule interesting according to the confidence measure?

- None of the above
- Confidence is 30% and The rule is interesting because it exceeds the confidence threshold.
- Confidence is 60% and The rule is interesting because it exceeds the confidence threshold.
- Confidence is 80% and The rule is interesting because it exceeds the confidence threshold.
- Confidence is 40% and The rule is interesting because it exceeds the confidence threshold.

b) Compute the interest measure (lift) for the association pattern {a, b}. Describe the nature of the relationship between item a and item b in terms of the interest measure.

- The interest measure is 0.889. It is interesting.
- The interest measure is 0.92. It is interesting.
- None of the above
- The interest measure is 0.640. It is not interesting.
- The interest measure is 0.12. It is not interesting.

Totalpoäng: 6

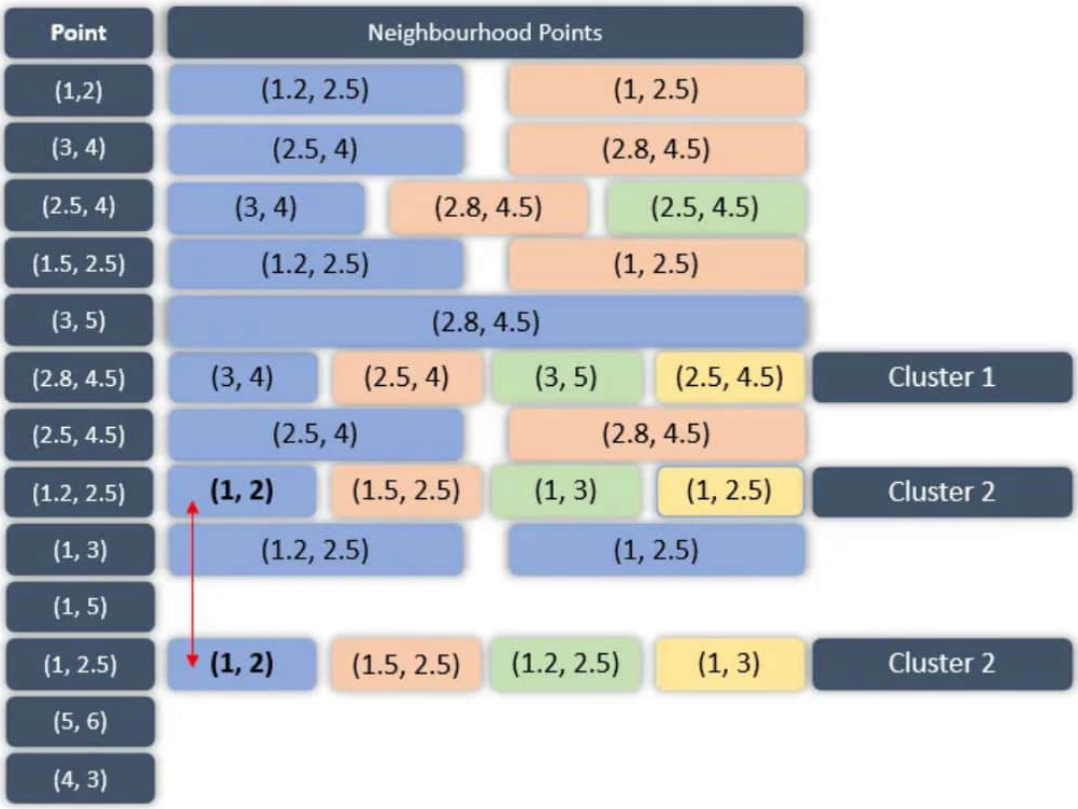
11 Section 3 - DBScan

The DBSCAN algorithm takes two input parameters. Radius around each point (*eps*) and the minimum number of data points that should be around that point within that radius (*MinPts*). In DBSCAN each point is checked for these two parameters and the decision about the clustering is made as described through the below steps:

1. Choose a value for *eps* and *MinPts*
2. For a particular data point (**x**) calculate its distance from every other datapoint.
3. Find all the neighbourhood points of **x** which fall inside the circle of radius (*eps*) or simply whose distance from **x** is smaller than or equal to *eps*.
4. Treat **x** as **visited** and if the number of neighbourhood points around **x** are greater or equal to *MinPts* then treat **x** as a **core point** and if it is not assigned to any cluster, create a new cluster and assign it to that.
5. If the number of neighbourhood points around **x** are less than MinPts and it has a core point in its neighbourhood, treat it as a border point.
6. Include all the **density connected points** as a single cluster. (What density connected points mean is described later)
7. Repeat the above steps for every unvisited point in the data set and find out all core, border and outlier points.

If the number of neighbourhood points around **x** is greater or equal to MinPts then **x** is treated as a core point, if the neighbourhood points around **x** are less than MinPts but is close to a core point then **x** is treated as a border point. If **x** is neither core nor border point then **x** is treated as an outlier.

We choose eps = 0.6 and MinPts =4, the point tagged as core point has 4 other points (>= MinPts) in its neighbourhood & the one tagged as border point is in the neighbourhood of a core point but has only one point in its neighbourhood (< MinPts). The outlier point is one which is neither border point nor core point.



Based on the above information, please mark all that are accurate.

- [ ] (3,5) is border point.
- [ ] (2.8, 4,5) is core point.
- [ ] (3, 4) is in cluster 2.
- [ ] (3, 4) is in cluster 1.
- [ ] (3, 4) is in cluster 3.
- [ ] (1,2) is an outlier.
- [ ] (1,5) is an outlier.
- [ ] (3, 4) is border point.

Totalpoäng: 10

12 Section 3 - K-means

Given K equally sized clusters, the probability that a randomly chosen initial centroid will come from any given cluster is 1/K, but the probability that each cluster will have exactly one initial centroid is much lower. (It should be clear that having one initial centroid in each cluster is a good starting situation for K-means.) In general, if there are K clusters and each cluster has n points, then the probability, p, of selecting in a sample of size K one initial centroid from each cluster is given by Equation:

$$p = \frac{\text{number of ways to select one centroid from each cluster}}{\text{number of ways to select } K \text{ centroids}} = \frac{K!n^K}{(Kn)^K} = \frac{K!}{K^K}$$

Given the above formula, choose the correct answers.

- [ ] This equation shows the exact number of samples that is required to be chosen to make the clustering accurate in first iteration.
- [ ] At two clusters, the probability is already at 50%.
- [ ] The probability is not sample size dependent for this setting.
- [ ] The probability is sample size dependent for this setting.

Algorithm Basic K-means algorithm.

1: Select K points as initial centroids.
2: **repeat**
3: Form K clusters by assigning each point to its closest centroid.
4: Recompute the centroid of each cluster.
5: **until** Centroids do not change.

Centroid= the point defined by the arithmetic mean all the selected points.

Data points	X1	X2
A	2	3
B	6	1
C	1	2
D	3	0

Given the K-Means algorithm above, consider 4 data points A,B,C,D as above and initialize the K-means algorithm by two centroids c1 and c2, calculated as c1= (A+B)/2 and c2= (C+D)/2, and given the distance measure Euclidean. The first iteration creates the following Euclidean distance table to initial centroids and their clusters:

Cluster label	Cluster 2	Cluster 1	Cluster 2	Cluster 2
Data point	A	B	C	D
c1	5	5	9	5
c2	4	16	2	2

For a cluster of size two, which point will merge to with which clusters in the final iteration?

Select one alternative

- A,C,D to c1 and B to c2
- A,B to c1 and C,D to c2
- A to c1 and B,C,D to c2
- None of the above.

Totalpoäng: 10

13 Section 4 - Hierarchical clustering - Graphs - Max link and Graph (Girwan Newman algorithm)

Question on Hierarchical clustering by employing complete link or MAX method.

Algorithm Basic agglomerative hierarchical clustering algorithm.

1: Compute the proximity matrix, if necessary.
2: **repeat**
3: Merge the closest two clusters.
4: Update the proximity matrix to reflect the proximity between the new cluster and the original clusters.
5: **until** Only one cluster remains.

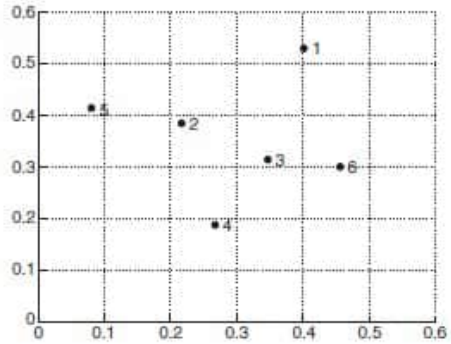


Figure Set of six two-dimensional points.

Point	x Coordinate	y Coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

Table xy-coordinates of six points.

Given the hierarchical clustering algorithm above and the data points and tables, please choose Euclidean distance metric for your proximity distance matrix equationssection1similarity and employ complete link or Max approach. For the complete link or MAX version of hierarchical clustering, the proximity of two clusters is defined as the maximum of the distance (minimum of the similarity) between any two points in the two different clusters. Using graph terminology, if you start with all points as singleton clusters and add links between points one at a time, shortest links first, then a group of points is not a cluster until all the points in it are completely linked, i.e., form a clique.

Please calculate the second iteration of the proximity matrix. The order follows 1 to 6 from left to right in the first distance matrix iteration. In the table entry below, and do follow the update order and enter the correct values below in that order. Please note that all values, including zeros needs to be entered. Please use dot such as 0.22. No more then two digits are required after dot.

Based on the above analysis, which clusters are available to us at the second iteration?

- None of the above
- p1,{p2,p3},{p4,p5},p6
- {p1}, {p2,p5},{p3,p6},p4
- {p1,p2},p3,{p5,p6},p4

Question on Hierarchical clustering by employing graph clustering (Girwan Newman algorithm) .

The Girvan–Newman algorithm detects communities by progressively removing edges from the original network. The connected components of the remaining network are the communities. Instead of trying to construct a measure that tells us which edges are the most central to communities, the Girvan–Newman algorithm focuses on edges that are most likely "between" communities. The measure used for hierarchical clustering that betweenness is defined by the following formula:

The betweenness centrality of a node v is given by the expression:

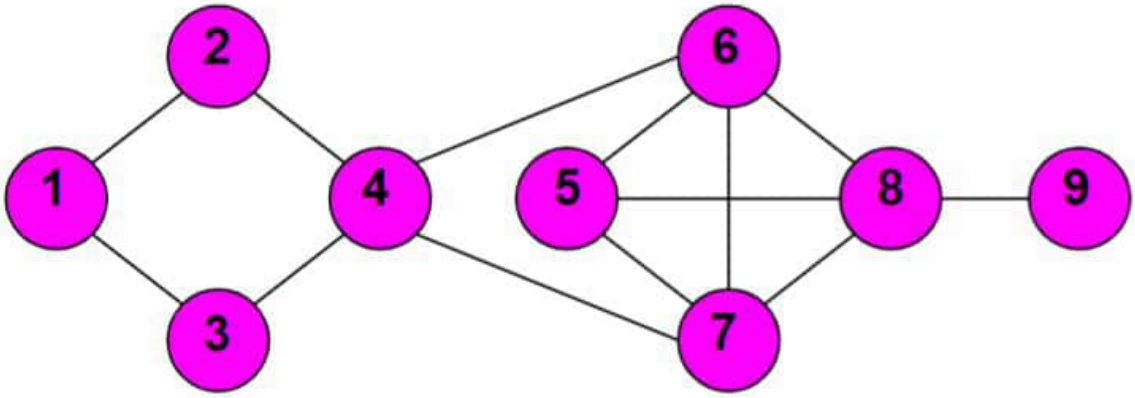
$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v (not where v is an end point).

The algorithm's steps for community detection are summarized below

1. The betweenness of all existing edges in the network is calculated first.
2. The edge(s) with the highest betweenness are removed.
3. The betweenness of all edges affected by the removal is recalculated.
4. Steps 2 and 3 are repeated until no edges remain.

Given the social network below:



Please mark all the answers that apply:

- [ ] The clusters in the second iteration will be {1,2,3,4} and {5,6,7,8,9}.
- [ ] After first iteration of the algorithm. There will be still one cluster that is all the nodes.
- [ ] After first iteration of the algorithm. There will be two clusters.
- [ ] The clusters in the first iteration will be {1,2,3,4} and {5,6,7,8,9}.
- [ ] The clusters in the first iteration will be {1,2,3,4,5,6,7,8} and {9}.

Totalpoäng: 10

14 Section 4 - Decision tree classifier and validation

To evaluate the impurity of a node t in a decision tree, below is three measures:

$$\begin{aligned} \text{Entropy} &= -\sum_{i=0}^{c-1} p_i(t) \log_2 p_i(t), \\ \text{Gini index} &= 1 - \sum_{i=0}^{c-1} p_i(t)^2, \\ \text{Classification error} &= 1 - \max_i [p_i(t)], \end{aligned}$$

where pi(t) is the relative frequency of training instances that belong to class i at node t, c is the total number of classes. All three measures give a zero impurity value if a node contains instances from a single class and maximum impurity if the node has equal proportion of instances from multiple classes.

Data set

Customer ID	Gender	Car Type	Shirt Size	Class
1	M	Family	Small	C0
2	M	Sports	Medium	C0
3	M	Sports	Medium	C0
4	M	Sports	Large	C0
5	M	Sports	Extra Large	C0
6	M	Sports	Extra Large	C0
7	F	Sports	Small	C0
8	F	Sports	Small	C0
9	F	Sports	Medium	C0
10	F	Luxury	Large	C0
11	M	Family	Large	C1
12	M	Family	Extra Large	C1
13	M	Family	Medium	C1
14	M	Luxury	Extra Large	C1
15	F	Luxury	Small	C1
16	F	Luxury	Small	C1
17	F	Luxury	Medium	C1
18	F	Luxury	Medium	C1
19	F	Luxury	Medium	C1
20	F	Luxury	Large	C1

a) Compute the overall Gini index for the overall collection of training examples. Choose the correct value.

- [ ] 0.48
- [ ] 0.5
- [ ] 0.3
- [ ] None of the above.

(b) Compute the overall Gini index for the Customer ID attribute.

- 0
- 0.5
- 0.2
- 0.1

(c) Compute the overall Gini index for the Gender attribute.

- 0.5
- 0.48
- 0.2
- None of the above

(d) Compute the overall Gini index for the Car Type attribute.

- 0.1625
- 0.10
- 0.22
- 0.3

e) Compute the overall Gini index for the Shirt Size attribute.

- 0.56
- 0.49
- 0.33
- 0.22

f) Which attribute is better, Gender, Car Type, or Shirt Size?

- Shirt size
- Gender
- Car type

Data set

Instance	a_1	a_2	a_3	Target Class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	−
4	F	F	4.0	+
5	F	T	7.0	−
6	F	T	3.0	−
7	F	F	8.0	−
8	T	F	7.0	+
9	F	T	5.0	−

Given the above data set, consider the training example for a binary classification problem.

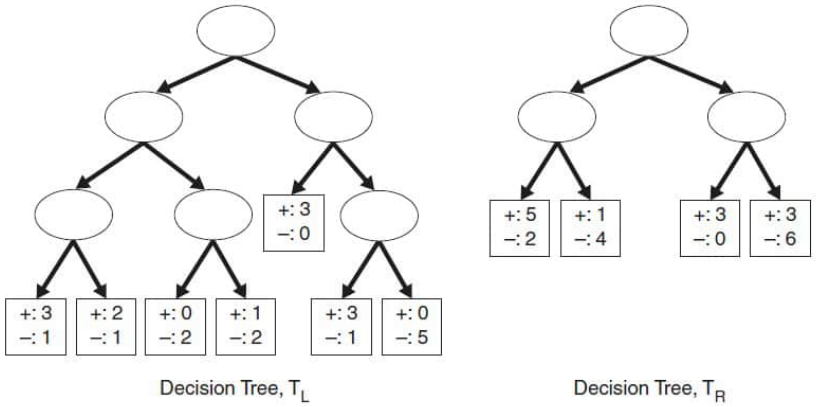
What are the information gains of a1 and a2 relative to these training examples? Please enter values to four decimals after decimal point as in the example 0.1301

Entropy $a_1 =$ [ ] and Entropy $a_2 =$ [ ] and maximum information gain [ ]

Totalpoäng: 22.5

15 Section 4 - Model selection

Consider the two binary decision trees,T_L and T_R, shown in Figure. Both trees are generated from the same training data and T_L is generated by expanding three leaf nodes of T_R. The buckets show the class predictions while (-) represents one class and (+) represents another class for each trial of 24 test data sets.



Example of two decision trees generated from the same training data.

Let k be the number of leaf nodes and N_{train} be the training instances. The generalization error rate of a decision tree T can then be computed using

$$err_{gen}(T) = err(T) + \Omega \times \frac{k}{N_{train}},$$

where Ω is a training tune parameter which is taken as one for this example. The training error rate for the tree is given by err(T).

Based on generalized errors $err_{gen}(T)$, which decision tree model is the best model for classification?

- Right tree with generalization error of 0.233
- Right tree with generalization error of 0.417
- Left tree with generalization error of 0.458
- Right tree with generalization error of 0.323

Totalpoäng: 7.5