

CIND 110 Data Organization for Data Analysts
Assignment 4
Data Mining Concepts

Tamer Abdou, PhD

April 3, 2017

1. On describing discovered knowledge using association rules

One of the major techniques in data mining involves the discovery of association rules. These rules correlate the presence of a set of items with another range of values for another set of variables. The database in this context is regarded as a collection of transactions, each involving a set of items, as shown below.

Trans ID	Items Purchased
101	milk, bread, eggs
102	milk, juice
103	juice, butter
104	milk, bread, eggs
105	coffee, eggs
106	coffee
107	coffee, juice
108	milk, bread, cookies, eggs
109	cookies, butter
110	milk, bread

1.1

Apply the Apriori algorithm on this dataset.

Note that, the set of items is {milk, bread, cookies, eggs, butter, coffee, juice}. You may use 0.2 for the minimum support value.

1.2

Show two rules that have a confidence of 0.7 or greater for an itemset containing three items.

2. On describing discovered knowledge using classification

Classification is the process of learning a model that describes different classes of data and the classes should be pre-determined. Consider the following set of data records:

RID	Age	City	Gender	Education	Repeat Customer
101	20..30	NY	F	College	YES
102	20..30	SF	M	Graduate	YES
103	31..40	NY	F	College	YES
104	51..60	NY	F	College	NO
105	31..40	LA	M	High school	NO
106	41..50	NY	F	College	YES
107	41..50	NY	F	Graduate	YES
108	20..30	LA	M	College	YES
109	20..30	NY	F	High school	NO
110	20..30	NY	F	college	YES

2.1

Assuming that the class attribute is Repeat Customer, apply a classification algorithm to this dataset.

3. On describing discovered knowledge using clustering

Consider the following set of two-dimensional records:

RID	Dimension 1	Dimension 2
1	8	4
2	5	4
3	2	4
4	2	6
5	2	8
6	8	6

3.1

Use the K-means algorithm to cluster this dataset. You can use a value of 3 for K and can assume that the records with RIDs 1, 3, and 5 are used for the initial cluster centroids (means).

3.2

What is the difference between describing discovered knowledge using clustering and describing it using classification.