ISCS 539 – Data Analytics

Assignment #4

1. Consider the following data set shown in Table 1.

(10 points) Find the frequent itemsets (up to size 4) for the support threshold min\_support = 0.5 by using the Apriori approach.

(15 points) Find the association rules for the support threshold min\_support = 0.5 and confidence threshold min\_confidence = 0.5 by using the Apriori and the rule generation approaches. Calculate the lift of each rule.

Table 1. Market basket transactions

|  |  |
| --- | --- |
| Transaction ID | Items Bought |
| 1 | Chips, Cookies, Regular Soda, Ham |
| 2 | Chips, Ham, Boneless Chicken, Diet Soda |
| 3 | Ham, Bacon, Whole Chicken, Regular Soda |
| 4 | Chips, Ham, Boneless Chicken, Diet Soda |
| 5 | Chips, Bacon, Boneless Chicken |
| 6 | Chips, Ham, Bacon, Whole Chicken, Regular Soda |
| 7 | Chips, Cookies, Boneless Chicken, Diet Soda |

(10 pints) 2. Import the Groceries data set in R. Answer the following questions by using the library arules in R.

How many frequent itemsets will be generated if the support threshold min\_support = 0.02, 0.04, and 0.06, respectively?

Consider the frequent itemsets generated with the support thresholds min\_support = 0.02, 0.04, and 0.6, respectively. How many association rules will be generated from the frequent itemsets if the confidence threshold min\_confidence = 0.1, 0.2 and 0.3, respectively?

(15 points) 3. Consider the following data set shown in Table 2. Place the objects in the data set in **two** clusters using k-means.

Andrew and Carolina are selected as the initial centroids of the two clusters.

Find the cluster to which each object belongs to using k-means with **one** iteration and re-compute the centroids of the two clusters.

Create a **scatter plot for the data set** and the **re-computed centroids**.

The cluster to which each object belongs to is represented by color of the object in the plot.

Table 2. Social network data set

|  |  |  |
| --- | --- | --- |
| Name | Age | Educational level |
| Andrew (A) | 55 | 1 |
| Bernhard (B) | 43 | 2 |
| Carolina (C) | 37 | 5 |
| Dennis (D) | 82 | 3 |
| Eve (E) | 23 | 3.2 |
| Fred (F) | 46 | 5 |