SSIE 637 Advanced Topics in Healthcare
Assignment 01
Fall 2016

Instructor: *Daehan Won*

Due 09/28/2016

Question 1: Literature review (30 points)
Read the following articles (posted on Blackboard) and provide your essay (at least 500 words each).
1. PCAST - Systems Engineering in Healthcare
2. Operations Research in Healthcare a survey
3. Knowledge discovery in medicine: Current issue and future trend

Question 2: Association rule mining (30 poibts=15+15 points)
(Part 1, *Apriori algorithm*)
The transcation data below has 6 transactions. What association rules can be found in this set, if the minimum support (minsup) is 33.33% and minimum confidence is 60%?

Transaction ID	Items
1	Hotdogs, Sandwich, Jam
2	Hotdogs, Sandwisch
3	Hotdogs, Milk, Chips
4	Chips, Milk
5	Chips, Jam
6	Hotdogs, Milk, Chips

Hint:
-Step 1: Count support for all single item set and form a table indicating item set and their corresponding supports
-Step 2: Form a table for 2 items. We only take items sets from the previous step whose support is 33.33% or more
-Step 3: Form a table for 3 items. We only take items sets from the previous step whose support is 33.33% or more
-Step 4: Enumerate all possible association rules based on the previous steps and calculate confidence for all rules. Indicate the rules whose confidences are greater than 60%.

(Part 2, *FP growth*)
Use the transaction data from the Part 1, buid a FP-tree with same minsup. Extract a frequent itemsets for item 'Milk'. You must show for each transaction how the tree evolves.

Hint:
-Step 1: Generate the list of frequent 1-itemsest and make the table (or list) where the items are sorted by frequency (support)
-Step 2: Proceed the tree evolution for every transaction. Each transcation should be sorted by item support
-Step 3: From the tree made by Step 2, take branches from item Milk and contruct a conditional tree for Milk. Extract freqent patterns from the tree.

Question 3: k-mean clustering (40 points) Given data ("random_data.txt"), let try k-mean clustering.
- Provide your clustering results which contain visualization effect (e.g., use scatter plot)

- Use different colors for each cluster.
- Mark a centroid (=mean) for each cluster
- Apply various range of k and show your conclusion from the clustering results breifly.