

## Data Mining Homework 1

Submit the PDF file to Dropbox on D2L.

1. The Apriori algorithm makes use of prior knowledge of subset support property. Given frequent itemset  $l$ ,  $s$  is a subset of  $l$  and  $s'$  is a subset of  $s$ .
  - a. Show that the confidence of the rule " $s' \rightarrow (l-s')$ " cannot be larger than the confidence of the rule " $s \rightarrow (l-s)$ ".
  - b. Show that the confidence of the rule " $(l-s) \rightarrow s$ " must be smaller than the confidence of the rule " $(l-s') \rightarrow s'$ ".
2. Given the following transaction database,  $\text{min\_support\_count} = 2$ , and the  $\text{min\_confidence} = 70\%$

TID	Items bought
T100	{K, A, D, B}
T200	{D, A, C, E, B}
T300	{C, A, B, E}
T400	{B, A, D}

- 2.A. apply Apriori algorithm to derive from the database:
  - (a) all frequent itemsets, and
  - (b) all strong association rulesShow all intermediate steps ( $C_k$ ,  $L_k$ , Join steps, ...) as done in class.

- 2.B. Given the following transaction database, apply FP-Growth algorithm to derive from the database all frequent itemsets. Show all the tree building and frequent item derivation steps.