CS6720: Data Mining
Assignment 2

March 26, 2014

Instructions:

1. The assignment consists of 5 questions.
2. You are required to a submit a report in PDF (typeset in Latex)
3. Some of the problem numbers referred to in this assignment are based on version 1.2 of the book "Mining of Massive Datasets"
4. For the questions of type implementation, the report should contain implementation and documentation. There will be a viva on these questions. The date and the time for the same will be announced later.
5. The submission deadline for this assignment is 6^{th} April, 11.59 pm.

Questions

1. **Implementation:** Implement CURE algorithm using the given data.
2. **Implementation:** Perform clustering using DB-scan algorithm using the given data. You can use related libraries.
3. **Theory question:** 4.4.1.
4. **Theory question:** 4.4.2.

5. **Theory question:** In frequent itemset mining, closed and maximal itemsets are defined in the following manner:
   An itemset X is closed if X is frequent and there exists no frequent super-pattern $Y \supset X$, with the same support as X.
   An itemset X is a maximal if X is frequent and there exists no frequent super-pattern $Y \supset X$.
   These definitions can also be applied to Multiple Support frequent itemset mining scenario, where each item has their own minimum support threshold, and the minSup of an itemset S is the minimum of all item support thresholds in S.

   - We know that in traditional frequent itemset mining, the set of closed itemsets is a superset of the maximal itemsets. Is this also true in the multiple support scenario? Is yes, prove theoretically. If not, give a contradictory example.
   - In traditional frequent itemset mining, the set of closed itemsets is a lossless compression. In other words, given just the set of closed itemsets and their frequencies, we can generate the entire set of frequent itemsets and with their frequencies without going through the database. Is this also true in Multiple support frequent itemset mining? If yes, prove that theoretically. If not, explain with an example.