CSE 4502/5717 Big Data Analytics

Homework 3, due on November 21, 2024 at 11:00 AM

1. (a) Input is a database DB with n transactions from a set $I = \{i_1, i_2, \ldots, i_d\}$ of items. It is known that each transaction in DB has $O(1)$ items. Input is also a threshold *minSupport* for the minimum support. Present an algorithm to find all the frequent 2-itemsets. The expected run time of your algorithm should be $O(n)$.

   (b) Let I be a set of items with $|I| = d$. Show that we can construct $3^d - 2^{d+1} + 1$ association rules from I.

2. Input is a database DB with q transactions from a set $I = \{i_1, i_2, \ldots, i_d\}$ of items. The total number of items in all of these q transactions is n. Assume that $d = O(n^c)$ for some constance c. Input also is a threshold *minSupport* for the minimum support. We are required to identify all the frequent items. Present an $O(n)$ time algorithm for this problem. Assume that each transaction is given as a list of items in it.

3. Input is a database DB with n transactions from a set $I = \{i_1, i_2, \ldots, i_d\}$ of items. Assume that $d = O(n^c)$ for some constant c. It is known that each transaction in DB has $O(1)$ items. Input is also a threshold *minSupport* for the minimum support. Present an algorithm to find all the frequent 2-itemsets. The **worst case** run time of your algorithm should be $O(n)$. (*Hint:* We can sort N integers in the range $[1, N^i]$ in $O(N)$ time, where i is any constant.)

4. Present an $O(n)$ time algorithm to compute the coefficients of the polynomial $(1 + x)^n$. How much time is needed if you use the FFT algorithm to solve this problem?

5. An $n \times n$ *Toeplitz* matrix is a matrix A with the property that $A[i, j] = A[i-1, j-1],\ 2 \leq i, j \leq n$. Give an $O(n \log n)$ algorithm to multiply a Toeplitz matrix with an arbitrary $(n \times 1)$ column vector.

6. Input are two polynomials $f(x)$ and $g(x)$ of degree n and m, respectively, in coefficients form. Present an $O(n \log m)$ time algorithm to multiply these two polynomials. The product should be output in coefficients form as well.