Name: ______________________________

CSE 4502/5717 Big Data Analytics

Fall 2022 Model Exam III

Note: You are supposed to give proofs to the time and processor bounds of your algorithms. Read the questions carefully before attempting to solve them.

1. 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 the number of items in each of the n transactions in DB is $O(1)$. Input also is a threshold $minSupport = 1/4$ for the minimum support. Present an $O((\log n)(\log\log n))$ time Monte Carlo algorithm for finding all the frequent items. Show that the output of your algorithm is correct with a high probability.

2. Input is a database DB with n transactions from a set $I = \{i_1, i_2, \ldots, i_d\}$ of items. Input also is a threshold $minSupport$ for the minimum support. We are required to identify all the frequent k-itemsets, where k is a constant. Present a parallel algorithm for this problem that runs in $O(\log n)$ time. You can use up to $\frac{nd^k}{\log n}$ CREW PRAM proccessors. Assume that each transaction is given as a bit array as discussed in class.

3. Present an $O(n \log^2 n)$ time algorithm to compute all the roots of a given degree-n polynomial $f(x)$. Assume the following: 1) The roots of $f(x)$ are integers in the range $[1, cn]$ where c is a constant; 2) The polynomial is given in coefficients form. (Recall that a is a root of $f(x)$ if $f(a) = 0$.) (*Hint:* Assume that we can evaluate any degree-n polynomial at n arbitrary points in $O(n \log^2 n)$ time).

4. Input is a sequence X of pairs of real numbers $(r_1, a_1), (r_2, a_2), \ldots, (r_n, a_n)$. The problem is to find a polynomial $f(x)$ of minimum degree (d) such that $f(r_i) = a_i$, for $1 \le i \le n$. For example, if the input sequence is $(0, 1), (1, 3), (2, 5), (3, 7), (4, 9)$, the answer is $f(x) = 2x + 1$. Present an algorithm to solve this problem that runs in $O(n \log^3 d)$ time. (Assume that $n > d \log d$). (*Hint:* Assume that we can perform interpolation at n arbitrary points in $O(n \log^3 n)$ time).

5. Construct a linear regression model for the following input examples: $(0, 1; 4)$, $(1, 0; 3)$, $(1, 1; 6)$, $(2, 1; 10)$. The model of interest is $f(x_1, x_2) = w_1 x_1 + w_2 x_2$. Compute the best values for the parameters w_1 and w_2.

6. Present a neural network (specifically, a multilevel perceptron) for realizing the Boolean function $F(x_1, x_2, x_3, x_4) = x_1 \bar{x_3} x_4 + x_2 \bar{x_3} + x_1 x_2 \bar{x_4}$.