Name: ____

CSE 4502/5717 Big Data Analytics

Exam III; December 8, 2022

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. (17 points) 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.)

2. (17 points) 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 at most c items, c being a constant. Input also is a threshold *minSupport* for the minimum support. Present an algorithm to find **all** the frequent itemsets. The expected run time of your algorithm should be $O(n)$.

3. (17 points) Present an algorithm to compute $f(x) = (x + a_1)(x + a_2) \cdots (x + a_n)$ where $a_1, a_2, \ldots, a_n$ are scalars (given as input). The output should be the coefficients of $f(x)$. Your algorithm should run in time $O(n \log^2 n)$.

4. (17 points) 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.

5. (16 points) Construct a linear regression model for the following input examples: $(0,1;3)$, $(1,0;4)$, $(1,1;6)$, $(2,1;10)$. The model of interest is $f(x_1,x_2) = w_1x_1 + w_2x_2$. Compute the best values for the parameters w_1 and w_2.

6. (16 points) Present a neural network (specifically, a multilevel perceptron) for realizing the Boolean function $F(x_1, x_2, x_3, x_4) = x_2x_3 + \bar{x_1}\bar{x_4} + x_2\bar{x_3}\bar{x_4}$.