Name: _____

CSE 4502/5717 Big Data Analytics

Exam III; December 5, 2019

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. (20 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 $f(x_1, x_2) = w_1 x_1 + w_2 x_2$. Compute the best values for the parameters w_1 and w_2.

2. (20 points) Present a neural network (specifically, a multilevel perceptron) for realizing the Boolean function $F(x_1, x_2, x_3, x_4) = x_1\bar{x}_3x_4 + x_2\bar{x}_3 + x_1x_2\bar{x}_4$.

3. (a) (18 points) Consider a neural network with L layers. There are n neurons at each layer. Assume that the activation function at each node is the linear function (i.e., $\sigma(x) = x$). In this case, show that one forward propagation can be completed in $O(\log L \log n)$ time using $\frac{Ln^3}{\log L \log n}$ CREW PRAM processors.

   (b) (2 points) When $\sigma(x) = x$ for each node in a neural network does it make sense to have multiple layers?

4. (20 points) 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.

5. (20 points) Input are a set X of n real numbers and an integer k $(1 \le k \le n)$. Show how to cluster X into k clusters using the (single linkage) hierarchical clustering algorithm in $O(n \log n)$ time. If a and b are two real numbers, then the distance between a and b is $|a - b|$.