Name: ______________________________

CSE 4502/5717 Big Data Analytics

Exam III (model); Fall 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. (25 points) The goal in this problem is to construct a linear regression model for the following input examples: $(0,1;2)$, $(1,0;1)$, $(1,1;5)$, $(2,1;4)$. 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.

2. Let $f : \Re^n \rightarrow \Re$ be any function on n variables. Given a series of examples to learn f, we can fit them using a linear model: $f(x_1, x_2, \ldots, x_n) = w_1x_1 + w_2x_2 + \cdots + w_nx_n$. As we showed in class, linear regression computes the optimal values for the parameters by equating the gradient to zero. Let the examples be $(x_i^1, x_i^2, \ldots, x_i^n; y_i)$ for $1 \leq i \leq m$. Let $\vec{w} = (w_1\ w_2\ \cdots\ w_n)^T$ be the parameter vector.

   Show that we can compute an optimal $\vec{w}$ in $O(\log^2 n + \log m)$ time using $n^4 + \frac{n^2 m}{\log m}$ CREW PRAM processors. Assume that we can invert an $n \times n$ matrix in $O(\log^2 n)$ time using n^4 CREW PRAM processors. Assume that $m \geq n$.

3. (25 points) Present a neural network (specifically, a multilevel perceptron) for realizing the Boolean function $F(x_1, x_2, x_3, x_4) = x_2x_3 + \bar{x_1}\bar{x_4} + x_2\bar{x_3}\bar{x_4}$.

4. (25 points) 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 2-itemsets. Present a parallel algorithm for this problem that runs in $O(\log n)$ time. You can use up to $\frac{nd^2}{\log n}$ CREW PRAM proccessors. Assume that each transaction is given as a bit array as discussed in class.