CSE 4502/5717 Big Data Analytics

Homework 3, due on December 3, 2019 at 2 PM

1. Construct a linear regression model for the following data: $(0, 1; 2), (1, 0; 4)$, and $(1, 1; 4)$.

2. Construct a neural network for the following Boolean function: $x_1\bar{x_3} + x_2x_4x_1 + \bar{x_2}x_3 + \bar{x_4}\bar{x_1}$.

3. Let A and B be two successive levels in a neural network. There are m nodes in A and n nodes in level B. There is an edge from every node in level A to every node in level B. In this problem we are focusing on computing the activation values of the nodes in level B, given the activation values of the nodes in level A. Show that this can be done in $O(\log m)$ time using $\frac{mn}{\log m}$ CREW PRAM processors.

4. (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.