CSCI 5080 Assignment 5

Use Microsoft **WORD** to write your answer and submit it to the Dropbox in D2L.
Please indicate **how much time** you spend on each problem.

1. (40 points) Quantitative association rules may disclose exceptional behaviors within a data set, where "exceptional" can be defined based on statistical theory. For example, Section 7.2.3 shows the association rule:

$$sex = female \Rightarrow meanwage = \$7.90/hr \ (overallmeanwage = \$9.02/hr),$$

which suggests an exceptional pattern. The rule states that the average wage for females is only $7.90 per hour, which is a significantly lower wage than the overall average of $9.02 per hour.

Read the following papers posted in D2L. Discuss how such quantitative rules can be discovered systematically and efficiently in large data sets with quantitative attributes.

Y. Aumann and Y. Lindell. A statistical theory for quantitative association rules. *Proc. of the 5th Intl. Conference on KDD*, 1999.

R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. *Proc. of the 20th Intl. Conference on VLDB*, 1994.

2. (60 points) The price of each item in a store is nonnegative. The store manager is only interested in rules of certain forms, using the constraints given below. For each of the following cases, identify the kinds of constraint they represent and briefly discuss how to mine such association rules using constraint based pattern mining.

(1) Containing items whose sum of the prices is less than $150
(2) Containing one free item and other items whose sum of the prices is at least $200
(3) Where the average price of all the items is between $100 and $500