Nova Southeastern University
College of Engineering and Computing
MMIS 643 Data Mining
Summer 2016
(May 9 – August 14, 2016)
Written Assignment #3
Due Date: July 24, 2016
Instructor: Dr. Junping Sun

A. A neural network typically starts out with random coefficients (weights); hence, it produce essentially random predications when presented with its first case. What is the key ingredients by which the net (neural network) evolves to produce a more accurate predication? (Please answer your question as clearly and concisely as possible.) (10 points)

B. Consider the Boston Housing Data file (The schema of the data file is given on page 33 in Table 2.2 of the textbook, the 3rd edition, page 27 of the 2nd edition.)(40 points)

a. Study the Neural Networks Prediction example from the URL: http://www.solver.com/xlminer/help/neural-networks-classification-intro, and following the example step by step.

b. Using XLMINER's neural network routine to fit a model using XLMINER default values for neural network parameters by using the predictors such as CRIM, ZN, INDUS, CHAS, NOX, RM, AGE, DIS, RAD to classify the value of CAT.MEDV.

   i. Record the RMS errors for the training data and the validation data, and observe the lift charts for repeating the process, changing the number of epochs to 300, 3000, 10,000, 20,000.

   ii. What happens to RMS error for the training data set as the number of epochs increases?

   iii. What happens to RMS error for the validation data set as the number of epochs increases?

   iv. Comments on the appropriate number of epochs for the model.

   Note: (Please use the Prediction Option of the Neural Network in order to get RMS)

C. For Association Rule Mining, please define the following terms: (10 points)

a. Support

b. Confidence

c. Lift

D. Study the Association Mining example from the URL: http://www.solver.com/xlminer/help/association-rules.

E. Problem 14. 3 on page 333-334 of the textbook, *Data Mining for Business Intelligence: Concepts, Techniques, and Applications with XLMiner, 3rd edition, 2016*, by Galit Shmueli, Nitin R. Patel, and Peter C. Bruce, ISBN: 978-1-118-72927-4. The data file is attached. (40 points)

Note:

1. The data files are posted along Written Assignment #3.