A STATISTICAL STEPWISE ALGORITHM FOR FITTING THREE-LAYER FEED-FORWARD NEURAL NETWORKS WITH DICHOTOMOUS OUTCOMES: A NONPARAMETRIC APPROACH BASED ON COMPARING AREAS UNDER CORRELATED RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES.

S.M. Sadat-Hashemi[†1], R. Ghorbani[1], B. Kavehie[2], M.R. Meshkani[3]

[1]*Semnan University of Medical Sciences, Semnan, Iran;*
[2]*Ministry of Science, Researches and Technology, Tehran, Iran;*
[3]*Shahid Beheshti University, Tehran, Iran*

[†] E-mail: *Sadat-Hashemi@sem-ums.ac.ir*

In traditional statistical models such as regressions, entering all predictors (independent variables or inputs) into the model is not necessary due to problems such as collinearity, unworthiness for some predictors to predict the response (outputs, dependent variable or outcome) that also make the model more dependent on data and more complex. Hence various techniques (e.g., stepwise or all possible regressions) are used to reduce the dimension of models′ input space and the predictor selection. Similarly, we need some proper techniques for fitting artificial neural networks on our data sets; However, due to the complexity of their structure and lack of suitable probability criteria as those, for modeling in Statistics (likelihood functions and etc), determination of the models′ structure is more difficult than the statistical regressions. There are three major common problems regarding the structure of any type of neural networks:
1. How many hidden layers are necessary?
2. How many hidden units in each hidden layer are necessary?
3. Which predictors are worth to enter the models input space?
Feed-forward neural networks have been widely used to predict outcomes, especially dichotomous outcomes. However, appropriate statistical procedures are not available for choosing input and hidden neurons. We have designed a stepwise algorithm to solve problems 2 and 3 for three layers networks based on powerful statistical inferences for correlated receiver operating characteristic (ROC) curves. We have also compared the results from our own algorithm with the genetic algorithm using numerical simulation.