Comparing discriminant function, logistic regression and neural network for classification of stocks using ROC curve: Evidence from Egyptian Stock Exchange

Dr. MEDHAT MOHAMED AHMED ABDELAAL,
Mathematics and Statistics department,
Faculty of Commerce, Ein Shams University,
CAIRO, EGYPT.

Abstract: Classification rates on out-of-sample predictions can often be improved through the use of model selection when fitting a model on the training data. Trained neural networks, with ESE financial ratios used as input, improve on ESE variables LDF and LR for discriminating between active and inactive stocks. The performance of Multilayer Perceptrons, Delta-Bar-Delta neural networks, LDF and LR can be improved with optimization of the features in the input. Neural network analyses show promise for increasing diagnostic accuracy of classifying the stocks. The areas under the ROC curves for MLP, and DBD were 0.929, and 0.927 respectively. For the full models of LDF and LR were 0.887 and 0.917 respectively. With the use of forward selection and backward elimination optimization techniques, the areas under the ROC curves for MLP and the LR were increased to approximately 0.93.

Keywords: Multilayer Perceptrons, back-propagation, Delta-Bar-Delta, linear discriminant function, logistic regression, financial indicators, receiver operating characteristic curve.

1. Introduction

With the financial indicators (independent variables) of the stock markets, 132 companies shared in ESE have been investigated to classify their stocks. With all the ESE financial ratios used as input, receiver operating characteristic (ROC) curves were generated for the classification of stocks, by two neural network techniques: a multilayer perceptron (MLP) with back propagation technique, and Delat-Bar-Delat (DBD) and, as well as linear discriminant function (LDF) and logistic regression (LR) analysis.

LDF analysis assumes that data representing different groups are linearly separable; also LR analysis assumes that data representing two groups are non-linearly separable. If these assumptions are not well met, the classifier's performance is degraded. Other investigators have used artificial neural networks (specifically, multilayer perceptrons [MLP] with back-propagated learning) trained on ESE financial indicators (independent variables) to classify stocks as active or inactive. Using this method, the neural network classifier is trained to detect a relationship between input (independent variables) and a predefined investor's decision by comparing neural network prediction with the dependent variable and by learning from its mistakes. In general, neural network techniques differ from basic statistical techniques such as LDF and LR, because they can adapt to the distribution of the data rather than assume a predefined distribution. The active of statistical or neural network classification methods is most often measured by reporting areas under the receiver operating characteristic (ROC) curve or by reporting sensitivity at different specificities.

The purpose of the current study was to compare the performance of ESE financial ratios LDF and LR with two artificial neural network methods in a single sample. Comparing different classification methods in a single sample reduces the effects of confounding variables. Because of their adaptability, we hypothesized that neural network techniques would perform as well as or better than LDF or LR classifiers in discriminating between active and inactive stocks.

2. METHODOLOGY

This part of the study includes shedding light on the case study used, the collected data description and the applied statistical techniques.

2.1. The aim of the study

The purpose of this study is to determine whether neural network techniques can improve differentiation between active (an investor can buy a stock) and inactive stocks (an investor can sell a stock) in Egyptian Stock Exchange (ESE). In an attempt to classify the stocks effectively as active or inactive, analysis strategies have been developed by using statistical methods such as linear discriminant function (LDF), logistic regression (LR) analysis and neural network technique. Also; this research aims at identifying a group of variables which deeply effect the decision of the investor whether to buy or sell a stock and also to suggest a model which can be used to classify any stock as they are parts of the financial instruments society. This research is extremely crucial for investors to identify the variables which will help them in taking the investment decision. Furthermore, the importance of this research according to companies which depend on issuing financial instruments is to disclose financial ratios which are of importance to the investors in the appropriate way and in the appropriate time. So that it could provide the investors with the needed information to interact with the stocks belonging to a company. The importance of this research according to the responsible authorities in the government is to supply to it a way to monitor the type of information needed by the investor to take the investment decision to achieve the transparency concept.

2.2. The research limits

This research deals with stocks only and doesn't deal with bonds in the ESE also it is based on the financial indicators and does not use political and economical variables which can be added to suggested models in future researches.

2.3. The data

One hundred thirty two companies randomly selected from 658 companies shared in ESE in year 2003 were included in the study. All ESE financial ratios have been calculated for each company; also the stocks of these companies have been observed as active or inactive stocks to construct the binary variable (dependent variable) which take two values; one for an active stock and zero for an inactive stock. The variables of this study are:

I. The dependent variable: It's a binary variable; it has 2 values zero and one. The value zero denotes that the stock is inactive, and the value one denotes that the stock is active. Active stock means an investor can buy a stock, while an inactive stock means an investor can sell a stock.

II. The Independent variables: This paper uses the financial indicators as independent variables because the investor's decision is based on these indicators. So, the independent variables are:

Current Ratio (X_1), Quick Acid Ratio (X_2), the Financial Leverage (X_3), Accounts receivable turnover ratio (X_4), Asset return ratio (X_5), Rate of return on equity (X_6), Profitability multiplier (X_7), Dividends (X_8), Market to Book Value ratio (X_9), Liquidity of Stock in the Market (X_{10}), Cash Flow from Operations (X_{11}), and the ratio between Cash in Flow and Current Liabilities (X_{12}).

3. Study Techniques

This part will focus on the deployed analytical techniques: linear discriminant function, logistic regression, the multilayer perceptron (MLP) neural network model with the back-propagation (BP) algorithm and Delta-Bar-Delta neural network technique.

3.1. Discriminant Analysis

Discriminant analysis approach suggests that the stocks represent two distinct populations: "active" population and "inactive" population. Before the stock history is accumulated, the prior probability of a stock to belong to the "active" population is $\pi_1 = P(S_i = 1)$, while the probability of a stock to belong to the "inactive" population is $\pi_0 = 1 - \pi_1$. Stock history represented by $X = (X^1, \ldots X^k)$ supplies us with additional information regarding the posterior:

$$P(S = s | X = 1) = \frac{\pi_1 p_1(x)}{\pi_1 p_1(x) + \pi_0 p_0(x)},$$

where $p_r(x) = P(X = x | S = r), r = 0,1.$

Maximum likelihood estimation in this case requires the maximization of the unconditional likelihood

$$p(s|x)p(x) = \pi_1^{N_1} \prod_{i=1}^{N_1} p_1(x_i) \pi_0^{N_0} \prod_{i=N_1+1}^{N_1+N_0} p_0(x_i) \qquad (1)$$

where N_1 and $N_0 = N - N_1$ are the numbers of active and inactive stocks respectively. This maximization is usually equivalent to separate maximum likelihood estimation of distribution parameters for active and inactive populations. Prior probabilities can be empirically estimated by the relative frequency of activation and inactivation in the sample (Cooley et al., 1985). Evaluation the performance of linear discriminant analysis formula was carried out for classifying the stocks as active or inactive. Also, developed and evaluated LDF that used all 12 financial ratios as input. This LDF was developed and tested, with 10-fold cross-validation used to reduce bias in developing and testing on the same samples.

3.2. Logistic Regression

Logistic regression suggests that there exists a stable statistical relationship, which allows one to determine the probability of active $P(S = 1)$ given the values of $X = (X^1, \ldots X^k)$:

$P(S = 1 | X = x) = 1 \div [1 + exp(\alpha - \beta x)].$

No assumption is made concerning the distribution of X^j. Coefficients α and β can be found directly by maximum likelihood estimation. Essentially, we may look for the values of α and β, maximizing the conditional likelihood

$$\begin{aligned} p(s|x) &= \prod_{i=1}^{N} P(S_i = s_i | X_i = x_i) \\ &= 1 \div \prod_{i=1}^{N} (1 + exp(\alpha - \beta x_i)) \end{aligned} \qquad (2)$$

Two models (discriminant analysis and logistic regression) may be considered equivalent if we assume

$$\beta x - \alpha = log \frac{\pi_1}{\pi_0} + log \frac{p_1(x)}{p_0(x)}. \qquad (3)$$

Nevertheless, corresponding estimation procedures may lead to different results due to different maximization procedures in (1) and (2). However, if we do not have a good model for the distribution of X to start with, the distribution-free logistic regression, optimizing conditionally on the observed values of X_i, seems to be a logical choice. Logistic regression via formula (3) can even be used for diagnostics and validation of parametric models for X. Evaluation the performance of logistic regression analysis formula was carried out for classifying the stocks as active or inactive. Also, developed and evaluated LR that used all 12 financial ratios as input. This LR was developed and tested, with 10-fold cross-validation used to reduce bias in developing and testing on the same samples.

3.3. Neural network techniques:

Evaluation the performance of two artificial neural network techniques (MLP and DBD) for classifying stocks as active or inactive was carried out. For both neural network techniques, all 12 financial ratios described earlier were included initially in the training set.

3.3.1. Multilayer perceptron neural network:

The MLP, a feed-forward back-propagation (BP) network, is the most frequently used neural network technique in recent researches. Two important characteristics of the MLP are: it's nonlinear PEs which have a nonlinearity that must be smooth (such as the logistic function and the hyperbolic tangent); and their massive interconnectivity, such as any element of a given layer feeds all the elements of the next layer (Haykin S. 1994).

The MLP is trained with error correction learning, which means that the desired response for the system must be known. In pattern recognition this is normally the case, since we know which data belongs to which experiment or the input data are labeled. Error correction learning works in the following way: from the system response at ith PE at nth iteration, $Y_i(n)$ and the desired response $d_i(n)$ for a given input pattern an instantaneous error $e_i(n)$ can be expressed as follows:

$$e_i(n) = d_i(n) - Y_i(n) \tag{4}$$

Using the theory of gradient descent learning, each weight in the network can be adapted by correcting the present value of the weight with a term that is proportional to the present input and error at the weight, which can be expressed as follows:

$$w_{ij}(n+1) = w_{ij}(n) + \eta\delta_i(n)x_j(n) \tag{5}$$

The local error $\delta_i(n)$ can be directly computed from $e_i(n)$ at the output PE or can be computed as a weighted sum of errors at the internal PEs. The constant η is called the step size. This procedure is called the BP algorithm. BP computes the sensitivity of a cost functional with respect to each weight in the network, and updates each weight proportional to the sensitivity. Momentum learning is an improvement to the straight gradient descent in the sense that a memory term is used to speed up and stabilize convergence. In momentum learning the equation to update the weights becomes

$$\begin{aligned} w_{ij}(n+1) &= w_{ij}(n) + \eta\delta_i(n)x_j(n) \\ &+ \alpha\left(w_{ij}(n) - w_{ij}(n-1)\right) \end{aligned} \tag{6}$$

where α is the momentum factor. Normally α should be set between 0.1 and 0.9. Training can be implemented in two ways: either we present a pattern and adapt the weights (on-line training), or we present all the patterns in the input file (an epoch), accumulate the weight updates, and then update the weights with the average weight update. The input and output of the i^{th} node in a MLP mode, according to the BP algorithm computations can be expressed as follows:

$$Input: X_i = \sum W_{ij}O_j + B_i \tag{7}$$

$$Output: O_i = F(X_i) \tag{8}$$

Where

Wij : the weight of the connection from node i to node j

Bi : the numerical value called bias

F : the activation function

The sum in equation (7) is over all nodes j in the previous layer. The output function is a nonlinear function which allows a network to solve problems that a linear network cannot solve. In this study the Sigmoid function given in equation (9) is used to determine the output state.

$$F(X_i) = 1 \div (1 + exp(-X_i)) \tag{9}$$

BP learning algorithm is designed to reduce an error between the actual output and the desired output of the network in a gradient descent manner. The summed squared error (SSE) is defined as:

$$SSE = 0.5\left[\sum_p \sum_i O_{pi} - t_{pi}\right]^2 \tag{10}$$

Where p index the all training patterns and i indexes the output nodes of the network. $\boldsymbol{O_{pi}}$ and $\boldsymbol{T_{pi}}$ denote the actual output and the desired output of node, respectively when the input vector p is applied to the network (Fahlman S.E. 1988). A set of representative input and output patterns is selected to train the network. The connection weight Wij is adjusted when each input pattern is presented. All the patterns are repeatedly presented to the network until the SSE function is minimized and the network learns the input patterns.

3.3.2. Delta-bar-Delta technique:

Delta rule is also well known as Widrow/Hoff's rule, or the rule of least mean squares, because it aims to minimize the objective function by determining the weights values. The aim is to minimize the sum of error squares. Delta rule equation is

$$\Delta w_{ij} = \eta \cdot y_{ci} \cdot \varepsilon_j \tag{11}$$

where Δw_{ji} is the adjustment of the connection weight from neuron ***j*** to neuron ***i*** computed by:

$$\Delta w_{ij} = \Delta w_{ij}^{new} - \Delta w_{ij}^{old} \tag{12}$$

$\boldsymbol{y_{ci}}$ is the output value computed in the neuron ***i*** , ε_j is the raw error computed by:

$$\varepsilon_j = y_{cj} - y_{dj} \tag{13}$$

η is the learning coefficient and $\boldsymbol{y_{dj}}$ is the desired output that is used to compute the error. In a classical BP neural network, the error is backpropagated through the network using the gradient descent algorithm. Since Delta rule is commonly used in supervised networks, it is necessary to mention the main problem that can occur in BP the error, i.e. the local minima. The local minima problem occurs when the minimum error of the function is found only for the local area and learning is stopped without reaching a global minimum. Generalized delta rule is obtained by adding a derivation of input neurons into a Delta rule equation such that weight adjustment is computed according to the formula:

$$\Delta w_{ij} = \eta \cdot y_{ci} \cdot \varepsilon_j \cdot f'(I_j) \tag{14}$$

where $\boldsymbol{I_j}$ is the input into neuron ***j.*** This rule is appropriate to be used with non-linear transfer functions.

Learning coefficient η is an important parameter for the speed and efficiency of neural network learning, and is typically determined as a single learning rate for all connections in the network. Delta-Bar-Delta learning rule was developed in 1988 by Jacobs in order to improve the convergence speed of the classical Delta rule. It is a heuristic approach of localizing the learning coefficient η in a way that each connection in the network has its own learning rate and changes those rates continuously as the learning progresses. Dynamic weight adjustment in the DBD rule is done according to the Saridis heuristic approach. On the other hand, when the sign of the weight is changed for a certain number of time steps, the rate for that connection is decreased. Thus, Delta rule equation (11) is modified so that the learning rate is different for each connection.

In order to overcome these shortcomings, Extended-Delta-Bar-Delta rule (EDBD), introduces a momentum term α_k which also varies with time. The momentum term is used to prevent the network weights from saturation, and the EDBD rule enables local dynamic adjustment of this parameter, such that the learning equation becomes:

$$\Delta w^{t}_{ij(k)} = \eta_k \cdot y_{cj} \cdot \varepsilon_j + \alpha_k \Delta w^{t-1}_{ij(k)} \qquad (15)$$

where α_k is the momentum of the connection ***k*** in the network and ***t*** is the time point in which the weights of the connection ***k*** are adjusted. DBDs are newly developed techniques used for solving classification and regression problems. DBD architecture resembles the architecture of MLPs (input layer, hidden layer, output layer). During training, the DBD nonlinearly maps the training data to a high dimensional space where a hyper plane is fit that maximizes the margin of separation between classes while minimizing the generalization error (ability to generalize results from finite training set to data set), with the use of statistical learning theory. The DBD attempts to split the positive and negative vectors to optimize the distance between the hyper plane and the nearest of the positive and negative examples. Each weight has its own learning rate that increases linearly as long as the weight' s direction of change does not rapidly alternate, in which case learning rate decreases exponentially (Jacobs, 1988).

4. Study results

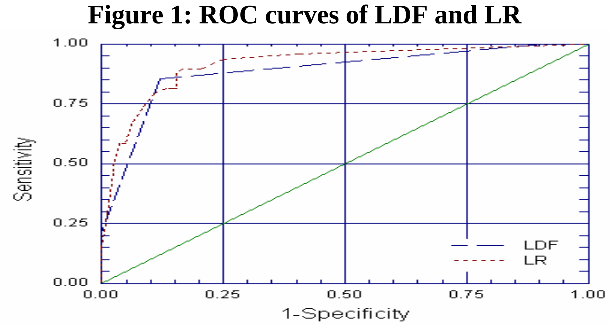
ROC curves for classifying stocks as active or inactive were determined for all techniques. These curves describe the continuous relationship between sensitivity and specificity at specificities ranging from 0% to 100%.

A Receiver Operating Characteristic curve (ROC) summarizes the performance of a two-class classifier across the range of possible thresholds. It plots the sensitivity (class two true positives) versus one minus the specificity (class one false negatives). An ideal classifier hugs the left and top sides of the graph, and the area under the curve is 1.0. A random classifier should achieve approximately 0.5 while a classifier with an area less than 0.5 can be improved simply by flipping the class assignment (Zweig H.M. 1993). The ROC curve is recommended for comparing classifiers, as it does not merely summarize performance at a single arbitrarily selected decision threshold, but across all possible decision thresholds. The ROC curve can be used to select an optimum decision threshold.

For LDF and LR, 10-fold cross-validation was used to evaluate the classifiers. The active and inactive stocks were each divided randomly into 10 approximately equal subsets. Ten mutually exclusive partitions were formed for cross validation (to measure the true rather than the estimated error rate) by combining one of the 10 inactive subsets with one of the 10 active subsets. One partition was used as the test set and the remaining nine partitions were combined to form the training set. The process was iterated, with each partition serving once as the test set. The results obtained for the 10 test sets were combined to generate a single ROC curve for each classification method. For MLP, cross-validation was similar, except eight partitions were used for training, one was used as a test set, and one was used as a stopping set to avoid overtraining. We provided sensitivities at specificities of 75% (representing moderate specificity) and 90% (representing high specificity), although this information is available in the graphic representations of ROC curves also presented. Finally, we reported the area under the ROC curve when specificity was 90% or more for the different techniques. These areas are bound by the ROC curve, the point at 100% specificity, and the line that passes through the point at 90% specificity and is perpendicular to the diagonal that represents chance discrimination. This information was provided to examine differences between techniques when specificity was high. The 90% specificity level was chosen because it theoretically forces the cases presumed to be the most difficult into the active group by allowing only 10% of these cases into the inactive group. Areas under the ROC curves (with sensitivities at 75% and 90% specificities) for all classification techniques evaluated are shown in Table 1.

The area under the ROC curve (±SE) for the LDF was (0.887±0.033). Area under the ROC curve of the LR was (0.917±0.029) which was significantly greater than the LDF. No other statistically significant differences between areas under the ROC curves of LDF and LR were observed. ROC curves for LDF and LR are shown in Figure 1.

Figure 1: ROC curves of LDF and LR



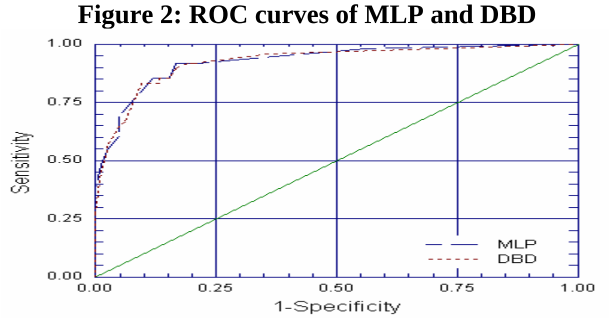
Areas under the curves when specificity was constrained from 90% to 100% were 0.180, and 0.138 respectively. Sensitivities at 75% specificity for LDF and LR were 86% for both methods. Sensitivities at 90% specificity were 79% and 69% respectively. Areas under the curves when specificity was constrained from 90% to 100% were 0.200 and 0.180, for MLP and DBD respectively. Sensitivities at 75% specificity for MLP and DBD were 90% for both techniques; sensitivities at 90% specificity were 82% and

Table 1: Area under ROC curves and Sensitivities.

	Technique	Area under ROC curve			Sensitivity %	
		Total	SE	Specificity 0.9-1.0	Specificity 75%	Specificity 90%
Statistical methods	LDF	0.887	0.033	0.180	86	79
	LG	0.917	0.029	0.138	86	69
Neural network techniques	MLP	0.929	0.027	0.200	90	82
	DBD	0.927	0.027	0.180	90	77
Optimizing techniques	MLP forward selection	0.935	0.026	0.177	94	77
	MLP backward elimination	0.935	0.026	0.128	89	71
	LR forward selection	0.933	0.026	0.228	94	88
	LR backward elimination	0.928	0.027	0.205	94	82

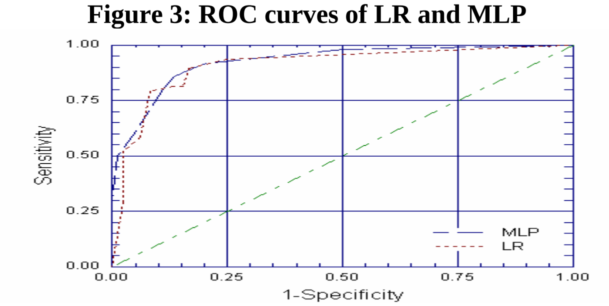
77%, respectively. ROC curves for MLP and DBD are shown in figure 2.

Figure 2: ROC curves of MLP and DBD



Areas under the ROC curves were significantly higher for MLP than DBD. ROC curves for the best neural network (MLP) and the best statistical method LR are shown in Figure 3.

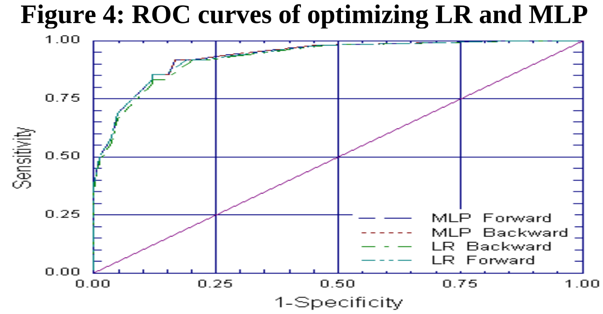
Figure 3: ROC curves of LR and MLP



4.1. Optimizing Neural Network and LR Results

When all financial ratios (independent variables) were included as input to the training set; the largest area under the ROC curve was the MLP for the neural network technique and LR for the statistical technique. Feature selection both by sequential forward selection and sequential backward elimination of features were carried out to determine whether relying on more effective features and removing less effective features would improve the performance of a classifier as measured by area under the ROC curve. During forward selection, an optimum training (input) set was determined by starting with an empty subset and adding one input parameter at a time (e.g., the one that most increased the area under the curve in combination with the previously selected variables) to the previously selected features until the area reached a maximum. During backward elimination, an optimal training set was found by starting with the full dimensional set from which the least effective input parameter was removed, one input parameter at a time (e.g., the one that resulted in the smallest increase in area under the ROC curve) until the maximum area was reached. This technique has been applied to LR and MLP. ROC curves for the optimizing MLP and LR are shown in Figure 4.

Figure 4: ROC curves of optimizing LR and MLP



Figures 5 and 6 shows that the optimal areas under the ROC curve with either forward selection or backward elimination when we were using approximately 40% of the input variables. These figures show areas under the ROC curve (y-axis) as a function of the number of ESE financial ratios in the training set (x-axis). The areas were maximized with a reduced dimension data set (subset of available input parameters) that contained an optimal combination of features determined by each optimization method, compared with using the full-dimensional feature set (all available input parameters). Using forward selection, the area under the ROC curve (± SE) increased from 0.935 (± 0.026) with all input variables, to a maximum of 0.935 (± 0.026) with 7 input variables. When the optimal feature set was analyzed at specificities constrained from 90% to 100%, the area under the ROC curve decreased from 0.200 to 0.177. Sensitivity at 75% specificity increased from 90% to 94%, and sensitivity at 90% decreased from 82% to 77%. When backward elimination was used, the area under the ROC curve increased to 0.928±0.027 and reached its maximum with 5 input variables. When specificity was constrained from 90% to 100%, the area was 0.205. Sensitivity at 75% specificity was 94% and sensitivity at 90% specificity was 82%.

Figure 5: Forward selection and backward elimination of LR

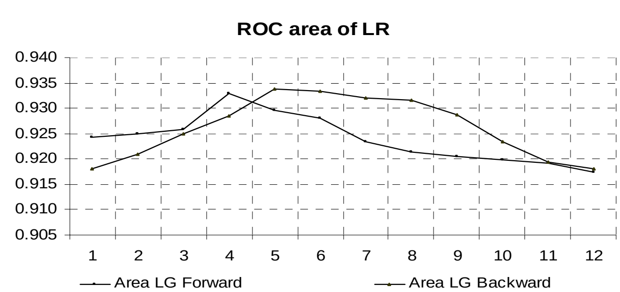
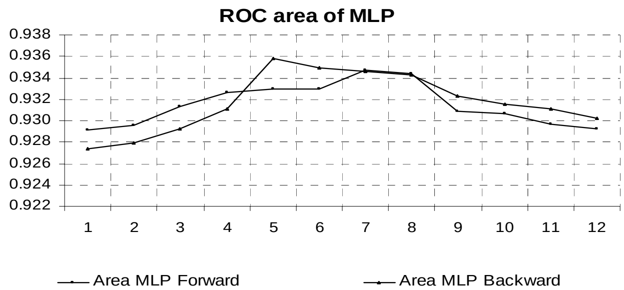


Figure 6: Forward selection and backward elimination of MLP



ESE variables included in the optimized MLP and LR training set with both methods is listed below:

LR Forward selection: X_5, X_7, X_9 and X_{12}

LR Backward selection: X_3, X_4, X_7, X_9 and X_{12}

MLP Forward selection: X_3, X_4, X_5, X_6, X_7 and X_9

MLP Backward selection: X_3, X_6, X_7, X_9 and X_{12}

When forward selection was used, the area under the ROC curve (±SE) increased from 0.917 ±0.029, with all input variables, to 0.933 ± 0.026, with 4 input variables. The areas when specificity was constrained from 90% to 100% increased from 0.138, with all input variables, to 0.177, with 4 input variables. Sensitivity at 75% specificity increased from 86% to 94%, and sensitivity at 90% increased from 69% to 77%. When backward elimination was used, the area under the ROC curve increased to 0.928 ± 0.027, with 5 input variables. When specificity was constrained from 90% to 100%, the area was 0.205. Sensitivity at 75% specificity was 95% and sensitivity at 90% specificity was 82%. The investor's decision classification performance of the optimized LR was similar to that of the optimized MLP, indicating that, with an optimal feature set, the data are linearly separable, and adaptive classifiers may not be necessary. Areas under the ROC curves for the optimized and full-dimensional LR and MLP are shown in Figure 4.

5. Conclusion

In our sample, all investigated financial ratios neural network techniques performed as well as or better than the LR functions. ROC curves for non-optimized neural network techniques ranged from 0.927 to 0.935, compared with 0.887 to 0.917 for LDF and LR methods. Further, optimization of the feature set significantly increased discrimination ability, probably because of the removal of variables that add information that has less value than the cost of including them in the training process. These results suggest that neural network classification techniques trained on ESE financial ratios are promising for discriminating between stocks. In the present study, the non-optimized technique that resulted in the largest area under the ROC curve for discriminating between active and inactive stocks (area under ROC curve = 0.929 for MLP) yielded a sensitivity of 82% at 77% specificity.

Although neural networks successfully discriminated between active and inactive stocks in this study, these techniques incur one general analysis. Due to the complexity of the classifiers, they do not allow the interaction of important variables to be identified and measured. Other classification techniques, such as Bayesian networks, allow better assessment of the relative contribution of features. In summary, neural network techniques were more successful at discriminating between active and inactive stocks than previously proposed LR. This improvement suggests that neural network techniques show at least as much potential for use in diagnosis of active and inactive stocks as nonlinear discriminant techniques. In addition, support vector machines demonstrated better generalization performance, and therefore better classification performance than MLPs. This result, coupled with the fact that MLPs are faster to train than DBD, suggests that DBD show superior potential for use in diagnosis of active and inactive stocks when compared with MLPs.

References

[1] Bishop C. *Neural Networks for Pattern Recognition*. University Press Oxford. 1995.

[2] Cooley, W.W. and Lohnes, P.R. *Multivariate Data Analysis*. Robert F. Krieger Publishing Co. Malabar, Florida. 1985.

[3] DeLong ER, DeLong DM, Clarke-Pearson DL. *Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach*. *Biometrics.* 1988; 44: pp 837–845.

[4] Fahlman S.E. *Faster-learning variations on back-propagation: an empirical study*. In Touretzky D., Hinton G.E. and T.J. 1988.

[5] Haykin S. *Neural Networks: A Comprehensive Foundation*. Macmillan Publishing New York. 1994.

[6] Jacobs R., "Increased Rates of Convergence through Learning Rate Adaptation", J. Neural Networks. 1988, Vol. 1, pp 295-307.

[7] Minsky M.L.and Papert S.A. *Perceptrons*. MA: MIT Press. Cambridge. 1996.

[8] Zweig M.H.and Campbell G. *Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine*. Chem. 1993. 4 39, pp 561-577.

About the author

Dr. Medhat Mohamed Abdelaal is lecturer in the Department of Statistics and Mathematics, Faculty of Commerce, Ein Shams University. He received his Ph.D. degree in applied statistics from School of mathematics and Statistics, University of Plymouth, UK. His research interests include stochastic modelling, neural network modelling, model selection, bootstrapping application, generalized additive modelling and computer intensive statistical methods.