Pergamon

PII: S0305-0548(96)00060-3

Computers Ops Res. Vol. 24, No. 4, pp. 367-377, 1997

0305-0548/97 $17.00+0.00

TOWARDS ANALYSING STUDENT FAILURES: NEURAL NETWORKS COMPARED WITH REGRESSION ANALYSIS AND MULTIPLE DISCRIMINANT ANALYSIS

A. M. Flitman†‡

Department of Business Systems, Faculty of Computing and Information Technology, Monash University, Clayton, Victoria 3168, Australia

(Received October 1995; in revised form June 1996)

Scope and Purpose—A key aim of any academic department is to maximize the likelihood of success for its students. This study aims to help achieve this objective by identifying which students are most at risk in specific subjects. This enables corrective action to be taken before the problems arise. This study investigates Neural Networks (comparing them with Logistic Regression and Multiple Discriminant Analysis) as tools to identify those students at-risk. We found significant improvement over previous similar studies by reducing the scope of the individual models.

Abstract—Using data from key first year courses, this article considers the development of subject-specific models to identify enrolled students at-risk of failure. The primary technique considered was neural networks, with it's results compared with logistic regression and multiple discriminant analysis. The three different modelling approaches were developed by three different analysts to achieve the benefits accruing from the independent M-Competition. We have found the quality of forecasts achieved to be significantly improved on earlier studies, presumably because of the subject specific nature of the models.

1. INTRODUCTION

Chatfield [1] questioned the validity of neural networks as viable forecasting tools. He pointed out that few comparisons with standard statistical techniques had been made. The research in this article compares results in trying to predict student failures using statistical and network approaches.

We have utilised neural networks and the most common statistical forecasting technique for this style of problem, namely logistic regression and multiple discriminant analysis.

2. THE STUDENT FAILURE PROBLEM

There has been an alarming increase in the percentage of students failing our first year subjects. Given the department's focus on quality of service this trend could not go unchecked. We decided to address the situation by trying to identify the reasons for student failure, thereby facilitating correction actions to help students.

Given that a large database concerning students is available at Monash (in line with most other Universities), we deemed it appropriate to attempt a quantitative analysis approach to the problem. The specific subjects we developed models for are listed in Table 1.

This research aimed to compare neural networks with logistic regression methods in identifying causes of student failure and therefore predicting future failures. This was achieved through the development of models using data from 1992 (and 1993 in the case of the statistical techniques), and then testing the

Table 1. Modeled subjects

Subject code	Description
BUS1021	Business Information Systems
BUS1060	Computer Programming for Business II
BUS1042	Computer Programming for Business I
BUS1100	Quantitative Methods for Business Systems
BUS1110	Computer Models for Business Decisions

† e-mail:AFLITMAN@fcit-m1.monash.edu.au

‡ Dr. Flitman is Director of Research and Deputy Head of the Department of Business Systems, Monash University, Melbourne Australia. He obtained his Ph.D. from Warwick University in England. His research interests include financial forecasting, simulation and strategic planning systems.

models against actual failures in 1993 and 1994.

Previous related studies tend to center on US management graduate student performance (for example Fisher and Resnick [2], Baird [3], Remus and Wong [4] and Hardgrave *et al.* [5]). Hardgrave *et al.* compared regression techniques with categorical techniques (such as neural networks) and concluded that the latter had the greater chance of success, although could only predict <60% of pass/fail. In general all the results were poor; [2] used least squared regression to obtain an R^2 of 8%, later improved upon by [3] (17%) and [4] (16%) by using different explanatory variables. Multiple discriminatory analysis employed by [4] provided the best result correctly classifying 64% of students as pass or fail.

Our research built upon the research by previous authors in a number of ways. Our research was based on prediction of undergraduate rather than postgraduate success or failure. This, of course, lead to a rather different set of explanatory variables, which were not normalized to the extent that the GPA is in the American studies.

Our models also differed in that we developed separate models for each subject (as opposed to the entire course). The American studies have viewed the problem from an admissions perspective (i.e. whether or not to accept an applicant). Our perspective was to consider what factors could adequately be addressed to improve our students chances of success. Since our models predict for a more specific area of skill than an entire course, we expected to obtain better results than the previous researchers.

We also differed in the analysis undertook. In addition to calculating R^2 we tested the null hypothesis that the proportion of students correctly classified by each model is the same, using a non-parametric test (the data is categorical). As a more procedural enhancement we developed the models in the spirit of the M-Competition (see [6]), with different forecasters for each technique, and with no knowledge of the results from the other forecasters.

Utilising previous researchers findings that categorical techniques worked better than standard regression techniques, the techniques we evaluated in this comparative study were:

- Back Propagation Neural Network
- Logistic Regression
- Multiple Discriminant Analysis

This article is organised into four further sections. In Section 3 we present a discussion of neural networks. We then describe the neural networks we have developed using data readily available at the University. In a similar (but abbreviated manner) we also discuss the logistic regression and multiple discriminant analysis techniques involved in this study. Finally we compare the predictive performance of the techniques and then indicate the direction of our research.

3. NEURAL NETWORKS

Neural networks are built from elements that behave somewhat like individual nerve cells (or neurons. Much research has been conducted looking at the application of this technology for classifying information such as the prediction of thrift failures [7]. This article investigates the development of a neural network system to classify students as future passes or fails in given subjects.

To ensure the non-linear aspect of the problem is catered for the networks have a minimum of three levels. The reader is referred to [8] for an overview of neural network theory. This article outlines only the very basics of neural net theory.

Whilst many different architectures and learning algorithms for neural network models exist, most successful applications utilise the 3 layer backpropagation model, as illustrated in Fig. 1.

When the input neurons receive data (in our case the explanatory variables), a calculation is performed at each neuron, with a subsequent signal sent to each connected internal neuron, which in turn passes a signal to each output neuron. The output layer then forms the prediction (in our case of pass or failure).

The calculations performed at each neuron are determined by an activation function, which are usually logistic, linear, linear threshold or hard limiting (on/off). For example, the activation functions below provide a logistic (i.e. sigmodal) function in the ranges (0,1) and [− 1,1] respectively.

$$f(x) = \frac{1}{1+e^{-x}}$$

$$f(x) = \tanh(x)$$

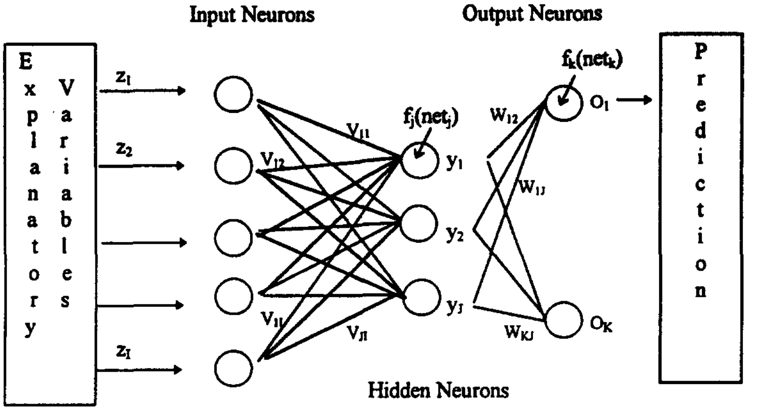

Fig. 1. Basic neural network architecture.

The size of the signal passed between any two neurons depends on both the activation function and the weight of the connection. In the network illustrated in Fig. 1 we have I, inputs; J, hidden neurons and K, outputs, with the inputs being denoted z_i, the outputs from the internal layer y_j, and the final outputs o_k. The weights connecting neurons are denoted v_{ji} and w_{kj} for connections to and from the internal layer respectively.

The network learns by using test data. Explanatory variables are supplied and the resultant output is compared to the desired output. The network then adjusts the interconnection weights between layers. This process is repeated until the network performs well on the training set. The network can then be assessed on data not included in the test set, to evaluate its performance.

The essential difference between neural networks and other forecasting techniques is that the networks use the training data to develop their representation for the modeled entity. This eliminates the situation associated with most models which must pre-determine assumptions about the modeled environment.

This suggests that, in those cases where we are forced to make the most assumptions in order to model them using traditional models, neural networks may provide better results.

However, in designing a neural network, there are a number of parameters that need to be selected. These include Learning Rate, Momentum, Initial Interconnection Weights (all to do with learning) and the number of neurons in the hidden layer(s). The choice of these parameters can greatly influence the network's performance.

4. A NEURAL NETWORK TO PREDICT STUDENT FAILURE

4.1. Selection of variables

Our principal sources of information have been from the student information contained in University and Faculty databases. Each provides a bank of personal student information, such as their country of origin, Victorian Certificate of Education (VCE – high school) results, sex, and family background. In all 61 fields of information were available (Appendix A). An analysis of the data (based on looking at the correlation of intuitively related fields such as VCE Chinese and being of Asian origin) led to a consolidation of the data fields into a more orthogonal set of 39 (Appendix A).

One of the major issues in developing a neural network is the selection of the data to include. The selection of too many variables results in a network 'memorising' the training set, thereby making it useless for predictive purposes. Similarly, too few variables may mean valuable information is missing.

Selection variables were based on simple experimentation with the data set limiting ourselves to six explanatory variables, to avoid the model 'rote' learning the training/test data sets. This compares with 8 variables used by [5] (1994) and 10 by [9]. Using R^2 as the criteria the best results were found with the variables listed in Appendix B.

4.2. The data set

Data for all the variables listed in Appendix A was obtained from University databases for all the students who took the five first year subjects under evaluation. Data were obtained for three years information, 1992, 1993 and 1994.

Unlike most of the MBA student studies (where students would at least have GMAT scores as a basis for comparison), our data was not complete (since not all students would for instance have taken the same VCE subjects). We therefore had to ensure that missing data was estimated from the samples. This was achieved using multiple regression analysis. A series of models were developed to infer a VCE score based on other subject scores, utilising results from the other students.

In the case of the logistic regression analysis and multiple discriminant analysis, a model was fitted selecting the six (to ensure consistency with the restriction we placed on the neural network) most significant variables on both 1992 and 1993 data. All the results for this style of analysis which predicted 1993 success and failure are therefore highly optimistic, since none of the data on which the models are tested is new to the model. In other words, the model was designed to fit the data. The 1994 forecasts, however, were based on out-of-sample testing.

However, in the case of the neural network, used only 1992 data to develop the models. This first year's data was used to develop the neural network. This was further subdivided, with 85% being used as training data and 15% being used as test data. Thus the network was trained using the training data, with the network model being saved each time an improvement was registered on the test data. The subdivision into training and test data was performed using random sampling. The second and third years of data were used as a validation (out-of-sample) set. Thus none of the networks developed were exposed to this data during model development. The validation data were simply used to compare the models with actual results in a practical actual-use scenario.

It is the results against the validation set which are of particular interest, as any statistical or neural network modeling approach can be applied to fit historic data with a good R^2.

4.3. Development of the networks

Several decisions needed to be made regarding network architecture. In particular the activation functions, the interconnection between neurons, and the number of hidden neurons had to be decided. In all analyses we used the unmodified R^2 as our basic criterion of success.

It is not intended here to discuss *activation functions* in any detail. The reader is referred to texts such as Hertz *et al.* (1991) [10]. Our research concentrated on the hidden layer activation functions, though consideration was made regarding the first layer scale functions and final layer activation functions.

The activation and scale functions we used in our comparison are summarised in Table 2 below:

With regards to *neuron connection* we considered the following network architectures with the criteria again being the best R^2 on the validation set:

- Back Propagation (as described above)
- Recurrent Networks with dampened feedback's ('Jordan–Elman')
- Multiple Hidden Slabs with different activation functions
- Jump Connection

Determining the size of the network (the number of neurons) has important consequences for its performance [11]. Too small a network may not reach an acceptable level of accuracy. Too many neurons may result in an inability for the network to generalise (it may rote learn the training patterns).

Selecting the number of hidden neurons is again more of an art than a science. Whilst no definitive view regarding the number of internal neurons has been developed, some basic heuristics have been. Three examples are:

Number of hidden neurons = 1/2 (Inputs + Outputs) + Sqrt(Number of Patterns)

Number of hidden neurons = 2 * square root (number of inputs or defining characteristics + the number of outputs or classifying characteristics) rounded down to the nearest integer.

Number of hidden neurons <= (Number of Patterns * Error Tolerance)/(Inputs + Outputs)

Table 2. Experimented scale and activation functions

Scale functions	Activation functions
Linear [− 1,1]	Gaussian
Linear (− 1,1)	Gaussian comp.
Linear [0,1]	Linear
Logistic	Logistic
Tanh	Sine
	Symmetric Logistic
	Tanh
	Tanh 15

Table 3. Optimal configurations

Subject	Architecture	Activation	Hidden neurons
BUS1021	Backpropagation 4 Layer	Gaussian, Gaussian	12,7
BUS1060	Jump 4 Layer	Gaussian, Gaussian	12,7
BUS1042	Jump 3 Layer	Gaussian	12
BUS1100	Backpropagation 3 Layer	Gaussian	12
BUS1110	Jump 3 Layer	Gaussian	7

The last of these is that provided by Baum and Haussler (see [11]).

As indicated we used R^2 (on the training set) as the principal determinant of network quality. This resulted in the optimal configurations listed in Table 3.

4.4. Training the networks

We trained the networks using basic supervised backpropagation. The training sets consisting of the various inputs and the single output (pass/fail) for 1992 were applied to the network. The data was normalised in the range $[-1,1]$. This required knowledge of the expected maximums and minimums for each data type, not just the maximum and minimum for the training set. The scaling was simply an indication of the relative size of each input with its minimum and maximum. Such scaling is essential as it minimises the effect of input magnitudes, and also aids the backpropagation learning algorithm. This algorithm works by presenting each training vector to the network in turn, computing the forecast error, and thereby determining inter connection weight changes throughout the network.

More specifically, the back propagation algorithm has four steps ([11]), designed to basically ensure the root mean squared error of the network on the training data is minimized. Using the notation introduced in Fig. 1 (and denoting the desired output at node k as d_k), these four steps can be summarised as:

(1) Sum for a single pass of the training set, the following error term, calculated for each training pattern:

$$\varepsilon = \frac{1}{2} \sum_{k=1}^{k} (d_k - o_k)^2$$

(2) Calculate the change in the set of weights connecting to the output level as follows:

$$\Delta_{w_{kj}} = -\eta \frac{\partial \varepsilon}{\partial_{w_{kj}}}$$

$$\Delta_{w_{kj}} = \eta (d_k - o_k) f_k'(net_k) y_j$$

where η is the learning rate.

(3) Now calculate the change of weights connecting to the inner level using:

$$\Delta_{\eta ji} = \eta f_j'(net_j) z_i \sum_{k=1}^{k} \delta_{o_k} w_{kj}$$

where $\delta_{o_k} = (d_k - o_k) f'(net_k)$

(4) Repeat this on each complete pass of the training data until no further improvement in the error term calculated in step 1 is being achieved (after a number of iterations).

We decided to present the training sets to the network in a random order, rather than rotate through the training set, so that bias could be minimised (the network could end up learning based on the position of the training data rather than the data itself).

Each time a number (200) of complete passes (epochs) of the training data through the network, we applied the partially trained network to the test data set, and recorded the average forecast error. If that forecast error was lower than previous forecasts on the test data, the network parameters were saved. Thus we were using the test data to determine the quality of the network. We regard this as preferable to the alternative of saving the network each time an improvement in the forecast on the training set was produced. This latter approach lends itself to the risk of saving a network which has memorised the training set, without the ability of generalising to new data.

This process of training continued until a number of successive tests using the test data yielded no further improvement. We allowed for around 300 tests on the test data without improvement, before concluding our training.

Once we had completed training, our final test was applying the network to the validation test sets. These sets were not used in the training at all, and represented 1993 and 1994 results. In comparing network quality we relied primarily on the R^2 statistic.

It is interesting to consider the relative importance of each input variable in predicting the networks' output. Appendix B includes the contribution factors associated with each of the explanatory variables.

These contribution factors sum the absolute values of the weights leading from each variable. In this regard these factors give an approximate measure relative to the other variables in the same network.

5. LOGISTIC REGRESSION AND MULTIPLE DISCRIMINANT ANALYSIS

5.1. Logistic regression

Multiple linear regression, or ordinary least squares (OLS) regression, is a flexible and well known method for analysing data. It is used for both summarising and understanding the relationships among quantitative data. The basic technique involves fitting a statistical model in which a dependent variable, such as a student's BUS1100 score, is estimated as a combination of several predictors or regressors such as a student's gender and VCE Chemistry score. The parameter estimates or regression coefficients yielded by the model are measures of how much change, on average, a one-unit change in any given regressor makes to the dependent variable.

Logistic regression is similar to OLS regression, differing only in that the statistical premise that it is based on is appropriate to the analysis of binary responses such as Pass/Fail. Accordingly we have used logistic regression to suggest which variables are likely to result in a pass or fail.

Using the data described in 4.2 above, our Logistic Regression specialist identified the variables listed in Appendix C as being significant predictors of student failure. Note that, though the neural networks and regression models were developed by different people, 21/30 (70%) of the explanatory variables selected were the same.

The coefficients and constants listed in Appendix C for each subject were used to calculate a z score for each student. For example, the z-score for BUS1021 was calculated as:

$$\begin{aligned} Z = {} & (1.5901) + (-0.8087) \times \textit{Temporary Visa} + (0.0253) \times \textit{VCE Mark Economics} \\ & + (-0.2240) \times \textit{VCE Mark Applied Engineering} + (-0.0224) \times \textit{VCE Mark Arc Geo. and Arc Graph} \\ & + (1.7204) \times \textit{Whether of Asian Origin} + (0.0137) \times \textit{Student Category} \end{aligned}$$

Once the z-score for a given student/subject combination was calculated the probability of failure was then calculated as:

$$\textit{Prob(Failure)} = \frac{1}{(1 + e^{-z})}$$

For example a z-score of ± 3.346 would result in a probability of failure of 0.0340. Based on this estimate we would predict that the student is unlikely to fail. In general any probability over 0.5 would be classified as a likely failure.

5.2. Multiple discriminant analysis

The other 'traditional' technique we adopted was Multiple Discriminant Analysis (MDA). MDA classifies a population into a number of groups based on a number of explanatory variables, which are combined in a linear equation to derive a fisher z-score for each group. This z-score can then be calculated for any member of the population for which we have the explanatory data, and a group deduced.

Using Fisher linear discriminant functions (and limiting for comparative purposes to 6 explanatory variables) based on the data set described in 4.2 above, our MDA specialist identified the variables listed in Appendix D as being significant predictors of student failure 17/30 (57%) of the explanatory variables selected were the same as for neural networks.

The coefficients and constants listed in Appendix D were used to calculate z-scores for group 1 (fail) and group 2 (pass) for each student/subject combination. If the z-score for group 1 was higher than group 2 the student was classified as a likely fail, otherwise they were regarded as a probable pass.

6. RESULTS

We developed the models using 1992 data only for the network, and 1992 and 1993 data in case of the logistic regression. We then tested the fitted models.

Results can be expressed as the percentage of pass/fail correctly predicted, and more crucially, the percentage of failures correctly predicted (Tables 4 and 5).

Clearly, the difference in overall classification is negligible. However, in terms of correct prediction of failure, the neural network far outperforms the regression technique. Owing to the fact that far more students pass than fail these subjects (around 80%), these greatly improved failure prediction capabilities are hidden in the overall results.

As the data are categorical, the null hypothesis that the proportion of student failures correctly classified by each method is identical was tested using a non-parametric test. In line with [7] we adopted the chi-square test statistic for equality of k proportions:

$$Q = \sum_{j=1}^{k} \frac{(f_j - n_j p)^2}{n_j p(1-p)}$$

where p is the proportion of successful catagorisations, f_j is the observed frequency of success and n_j is the number of observations. The results for each course are summarised in Table 6.

This means that in all cases except the 1993 prediction of BUS1060 results and the 1994 prediction of BUS1100, there is a significant difference in the predictive capability of the methods considered. This strengthens our belief that the neural networks have, in general performed better than the traditional methods for this particular application.

In common with the findings from studies published by other researchers the results indicate that further work needs to be done to develop practical quantitative predictors of student failure. A good summary of the results from previous studies can be found [5]. Utilising this summary, we can compare the results of these studies with our research, as in Table 7.

Our results appear to be an improvement on previous studies. We believe that this is because we developed specific models by subject (as opposed to a whole course), improving the specificity of the individual models. In addition, the inherent bias in most studies (they only include those students who

Table 4. Overall correct classification percentage

Subject	Logistic regression		Multi-discriminant analysis		Neural network	
Forecasted yr>	1993	1994	1993	1994	1993	1994
BUS1021	81%	74%	**82%**	74%	81%	**81%**
BUS1060	**91%**	**80%**	**91%**	76%	79%	79%
BUS1042	85%	83%	85%	83%	**88%**	**88%**
BUS1100	74%	74%	**76%**	69%	75%	**75%**
BUS1110	90%	87%	90%	82%	**91%**	**91%**

The best performances are highlighted in bold text.

Table 5. Correct failure prediction percentage

Subject	Logistic regression		Multi-discriminant analysis		Neural network	
Forecasted yr>	1993	1994	1993	1994	1993	1994
BUS1021	4%	0%	9%	0%	**35%**	**28%**
BUS1060	14%	4%	14%	4%	**22%**	**27%**
BUS1042	0%	0%	3%	0%	**20%**	**8%**
BUS1100	13%	10%	**45%**	**16%**	39%	8%
BUS1110	0%	0%	7%	13%	**50%**	**22%**

The best performances are highlighted in bold text

Table 6. Probability of equality of proportions over the models

	1993	1994
BUS1021	$p<.001$	$p<.001$
BUS1060	$0.2<p<0.3$	$p<.001$
BUS1042	$p<.001$	$p<.001$
BUS1100	$p<.001$	$0.1<p<0.2$
BUS1110	$p<.001$	$p<.001$

Table 7. A comparison with other reported studies

Researchers	Techniques	% Predicted	R^2
Fisher and Resnick	Least squares regression		8%
Deckro and Woundenberg	Least squares regression		15%
Gayle and Jones	Least squares regression		17%
Remus and Wong	Stepwise regression		16%
Poalillo	Stepwise regression		19%
Deckro and Woundenberg	Stepwise regression		15%
Graham	Stepwise regression		17%
Remus and Wong	Discriminant analysis	64%	
Our research (1994 results)	**Techniques**	**% Predicted**	**R^2**
BUS1021	Discriminant analysis	74%	
BUS1060	Discriminant analysis	76%	
BUS1042	Discriminant analysis	83%	
BUS1100	Discriminant analysis	69%	
BUS1110	Discriminant analysis	82%	
BUS1021	Neural network		36%
BUS1060	Neural network		35%
BUS1042	Neural network		20%
BUS1100	Neural network		32%
BUS1110	Neural network		37%

actually apply for admission in a program, and therefore exclude some of the target population who presumably never apply because of their low GMAT scores) is not present in our samples.

7. SUMMARY

We have applied Neural Networks, Logistic Regression and Multiple Discriminant Analysis techniques to predict individual subject performance within our core undergraduate degree. We have found significant improvement over previous studies by reducing the scope of the individual models to cover more specific subject areas. We believe that there is further scope for improving the results through improved handling of missing data. We are currently in the process of undertaking a follow up study, utilising the results of our analysis to provide extra tuition to those students deemed marginal as a result of using our models.

REFERENCES

1. Chatfield, C., Neural networks: forecasting breakthrough or passing fad? *Int. J. Forecasting*, 1993, **9**, 1–3.
2. Fisher, J. B. and Resnick, D. A., Standardized testing and graduate business school admission: a review of issues and an analysis of a baruch college MBA cohort. *College Univ.*, 1990, **55**, 137–148.
3. Baird, L.L., Comparative prediction of first year graduate and professional school grades in six fields. *Educ. Phychol. Measure.*, 1975, **35**, 137–148.
4. Remus, W. and Wong, D. A., An evaluation of five models for the admission decision. *College Student*, 1982, **16**, 53–59.
5. Hardgrave, B. C., Wilson, R. L. and Walstrom. K. A., Predicting graduate student success: a comparison of neural networks and traditional techniques. *Computers Ops Res.*, 1994, **213**, 249–263.
6. Makridakis, S., The accuracy of extrapolation (time series) methods: results of a forecasting competition. *J. Forecasting*, 1982, **1**, 111–153.
7. Salchenberger, L. M., Cinar, E. M. and Lash, N. A., Neural networks: a new tool for predicting thrift failures. *Decision Sciences*, 1992, **23**, 899–916.
8. Lipperman, R., *An Introduction to Computing with Neural Nets*. IEEE, 1987.
9. Graham, L. D., Predicting academic success of students in a master of business administration program. *Educational Psychol. Measure.*, 1991, 721–727.
10. Hertz, J., Krogh, A. and Palmer, R. G., *Introduction to the Theory of Neural Computation*. Addison-Wesley, Redwood City CA, 1991.
11. Wasserman, P. D., *Advanced Methods in Neural Computing*. Van Nostrand, New York, 1993.

APPENDIX A: Available data fields

Raw data fields	Combined data fields
Year	Year
Sex	Sex
Temporary visa	Temporary visa
VCE mark Australian studies	VCE mark Australian studies
VCE mark Chemistry	VCE mark Chemistry
VCE mark Chinese	Whether of Asian origin
VCE mark Economics	VCE mark Economics
VCE mark English (3 and 4)	VCE mark English
VCE mark English (ESL)	VCE mark English (ESL)
VCE mark English	VCE mark Physics/Inf. Proc and Mgt.
VCE mark Physics/Inf. Proc and Mgt.	VCE mark Comm. Language/Legal Studies
VCE mark Comm. Language/Legal Studies	VCE mark App Eng.
VCE mark App Eng.	VCE mark Arc Geo. and Arc Graph
VCE mark Arc Geo.	VCE mark Engineering Drawing
VCE mark Engineering Drawing	VCE mark Physics
VCE mark Arc Graph	VCE mark Accounting
VCE mark Physics	Student Category (Business Systems/Commerce/Other)
VCE mark Accounting	Whether Completed Year 12
VCE mark Hong Kong	First Language English
VCE mark Malaysia	Subject Result
Whether Business Systems Student	Tertiary Entrance Score
Whether Commerce Student	Whether Born Overseas
Tertiary Entrance Score	Sex
Whether Completed Year 12	Not Living at Home with Parents
First Language English	Course was Not First Course
First Language Chinese	Got in Other than via VCE
First Language Malay	Australian Born
Subject Result	English Second Language at Home
Whether Born Overseas	Not Australian Resident
Sex	Working
Not Living at Home with Parents	From Government School
Course was Not First Course	From Catholic School
Got in Other than via VCE	From Independent School
Australian Born	Did Not Attend Australian School
English Second Language at Home	Single Sex School
Not Australian Resident	Social Economic Group 1
Working	Social Economic Group 2
From Government School	Social Economic Group 3
From Catholic School	Delayed Starting Course
From Independent School	
Did Not Attend Australian School	
From Single Sex School	
Father is Manager/Professional	
Father is Clerk/Tradesman	
Mother is Manager/Professional	
Mother is Clerk/Tradesman	
Either Parent is not Employed	
Delayed Starting Course	
Parent's Highest Qualification	

Appendices continued overleaf

APPENDIX B: Combined data fields utilised by the neural networks

Subject	Key variables in model	Contribution factors
BUS1021 Business Information Systems	VCE mark Economics	2.852
	Sex	3.730
	Whether of Asian Origin	3.671
	VCE mark Applied Engineering	2.701
	Temporary Visa	4.113
	VCE mark Arc Geo. and Arc Graph	3.313
BUS1060 Computer Programming for Business II	VCE mark Physics	3.247
	Whether of Asian Origin	3.775
	Temporary Visa	3.031
	Student Category	3.522
	Sex	3.753
	VCE mark Arc Graph	3.948
BUS1042 Computer Programming for Business I	VCE mark Chemistry	3.073
	Whether of Asian Origin	3.365
	VCE mark Arc Geo. and Arc Graph	2.675
	Temporary Visa	2.860
	Sex	2.244
	Student Category	2.551
BUS1100 Quantitative Methods for Business Systems	Whether of Asian Origin	2.068
	VCE mark Applied Engineering	2.732
	Temporary Visa	3.153
	Sex	2.593
	Student Category	3.898
	Whether completed year 12	2.365
BUS1110 Computer Models for Business Decisions	Whether of Australian Origin	1.145
	VCE mark Engineering Drawing	1.178
	Temporary Visa	1.090
	Sex	1.016
	Student Category	1.828
	Whether completed year 12	1.490

APPENDIX C: Combined Data Fields, their Coefficients and Constants Utilised by the Logistic Regression Models

Subject	Key variables in model	Coefficients
BUS1021	*Temporary Visa*	−0.8087
	VCE mark Economics	−0.0253
	VCE mark Applied Engineering	−0.2240
	VCE mark Arc Geo. and Arc Graph	−0.0224
	Whether of Asian Origin	1.7204
	Student Category	0.0137
	(Model Constant:)	1.5091
BUS1060	VCE mark Australian Studies	−0.3669
	VCE mark Arc Geo. and Arc Graph	−0.0020
	Sex	1.6869
	Whether of Asian Origin	7.7300
	VCE mark Comm. Language/Legal Studies	−0.0483
	English Second Language at Home	6.6133
	(Model Constant:)	1.8305
BUS1042	*Sex*	0.9710
	VCE mark Australian Studies	−0.2586
	Temporary Visa	−0.3851
	VCE mark Chemistry	0.0086
	VCE mark Physics mark	0.0254
	VCE mark Arc Geo. or Arc Graph	0.0112
	(Model Constant:)	1.2655
BUS1100	*Whether of Asian Origin*	−0.5860
	VCE mark Physics/Inf. Proc and Mgt.	−0.2991
	Student Category	1.5908
	VCE mark Australian Studies	−0.2434
	Whether Completed Year 12	−0.7238
	VCE mark Comm. Language/Legal Studies	−0.0177
	(Model Constant:)	1.3092
BUS1110	*VCE mark Engineering Drawing*	0.0050
	Sex	0.8146
	Temporary Visa	−0.7380
	VCE Chemistry mark	0.0188
	Student Category	0.1997
	Whether completed year 12	−1.1719
	(Model Constant:)	2.3940

Those key variables in italics match those for the corresponding neural network.

APPENDIX D: Combined data fields utilised by the MDA models

Subject	Key variables in model	Coefficients for Group 1 (Fail)	Coefficients for Group 2 (Pass)
BUS1021	*Temporary Visa*	2.476	1.500
	VCE mark Economics	0.019	0.034
	VCE mark Applied Engineering	0.043	0.015
	VCE mark Arc Geo. and Arc Graph	0.427	1.039
	Whether of Asian Origin	0.252	1.773
	VCE mark Physics	0.021	0.029
	(Model Constant:)	−2.700	−1.210
BUS1060	*Sex*	0.605	2.040
	VCE mark Australian Studies	0.316	0.016
	Whether of Asian Origin	0.361	1.413
	VCE Chemistry mark	−0.010	0.011
	VCE mark Comm. Language/Legal Studies	0.067	0.008
	Tertiary Entrance Score	0.011	0.007
	(Model Constant:)	−3.504	−1.233
BUS1042	*Sex*	0.957	1.857
	VCE mark Australian Studies	0.217	−0.010
	Temporary Visa	2.763	2.268
	VCE mark Chemistry	0.013	0.019
	VCE mark Physics mark	0.057	0.075
	VCE mark Arc Geo. or Arc Graph	0.013	0.023
	(Model Constant:)	−2.980	−1.513
BUS1100	*Whether of Asian Origin*	0.017	−0.041
	VCE mark Physics/Inf. Proc and Mgt.	0.128	−0.468
	Student Category	−0.981	0.828
	VCE mark Australian Studies	0.112	−0.041
	Whether Completed Year 12	5.282	3.819
	VCE mark Comm. Language/Legal Studies	−0.010	0.004
	(Model Constant:)	−3.423	−2.020
BUS1110	*Whether Competed Year 12*	6.562	3.786
	VCE Chemistry mark	−0.024	0.001
	Sex	0.332	1.344
	VCE mark Arc Geo. or Arc Graph	−0.158	0.001
	Whether on Temporary Visa	3.863	3.121
	VCE mark Comm. Language/Legal Studies	−0.421	0.488
	(Model Constant:)	−5.716	−2.264

Those key variables in italics match those for the corresponding neural network.