Estimation of Student Success with Artificial Neural Networks

Yapay Sinir Ağları ile Öğrenci Başarısı Tahmini

Kemal TURHAN[*]   Burçin KURT[**]   Yasemin Zeynep ENGİN[***]

Karadeniz Teknik Üniversitesi

Abstract

Faculties' counseling systems need to take early action for problematic students by predicting and tracing student success and related problems. In this study, we compared two methods for predicting students' success to take systematic and proactive actions for detecting and solving problems in early phase. In this study, it is aimed to estimate the marks that the students of the 1st semester of School of Medicine of Karadeniz Technical University, who were 111 in total, 56% of whom were males and 44 % were females. By using their final exam (F1) results that were gathered through multiple choice tests, variables that could be considered as affecting student performance were analyzed. Artificial neural networks and regression analyses were carried out with success in various research areas for the settlement of classification, clustering and regression problems   As a result, it is considered that there appears a need of advanced level research and the automatic information system implementations to be developed at the end of those researches in the student performance estimation as also in F1 estimation.

Keywords: Artificial Neural Networks, Prediction, Student Performance, Regression.

Öz

Yapay sinir ağları (YSA), sınıflandırma, kümeleme ve regresyon problemlerinin çözümünde başarıyla uygulanabilmektedir. Bu sebeple öğrenci ders başarısının erken dönemde tahmini ve gerekli önlemlerin alınabilmesi için de kullanılabilir. Bu çalışmada Karadeniz Teknik Üniversitesi Tıp Fakültesi Dönem I öğrencilerinin kurul sonu çoktan seçmeli üç ara sınav ve öğrenci başarısını etkileyebilecek diğer faktörler eklenerek dönem sonu final sınavı YSA ve regresyon analizi ile tahmin edilmeye çalışılmış ve YSA'nın daha yüksek performans gösterdiği gözlenmiştir.

Sonuç olarak daha ileri düzeyde araştırmalar yapılarak erken uyarı sistemleri geliştirilebileceği görülmüştür.

Anahtar Sözcükler: Yapay Sinir Ağları, Tahmin, Öğrenci Performansı, Regresyon.

Introduction

Faculties' counseling systems need to take early action for students who have some problems. It is important to predict and trace student success and related problems. In this study we compare two methods for predicting students' success to take systematic proactive actions for detecting and solving problems in early phase. It is used for the purposes of presupposing student performance; directing the student; triggering student consultancy system; classification of students according to performance and abilities; testing the training program and methods and informing the trainers. Therefore, the estimation of student performance has a strategic importance in the settlement of training problems. There are many different methods for statistical estimation. Regression Analysis and Artificial Neural Network (ANN) are investigated and compared in this study for predicting student success.

* Doç. Dr. Kemal TURHAN, Karadeniz Teknik Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, kturhan.tr@gmail.com

** Burçin KURT, Karadeniz Teknik Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, burcinnkurt@

*** Yasemin Zeynep ENGİN, Karadeniz Teknik Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Yüksek Lisans Öğrencisi, nymph_yze@hotmail.com

To estimate the relationships between variables, regression analysis is a satisfactory and commonly used statistical approach which contains many methods for analyzing and modeling the variables. However, this technique can give misleading results for applications with small effects or questions of causality based on observational data. As another method, ANN was developed using model of the biological nervous system (Figure 1). It becomes popular after the presentation of simplified neurons by (McCulloch and Pitts, 1943). ANN is relatively new technique in educational sciences so that we will give general information about ANN in the following section.

Method

In this study, it is aimed to estimate the marks that the students of the 1st semester of School of Medicine of Karadeniz Technical University, who were 111 in total, 56% of whom were males and 44 % were females. By using their final exam (F1) results that were gathered through multiple choice tests, variables that could be considered as affecting student performance were analyzed.   To estimate the final exam result, gender, type of the high school graduated, physical environment, and family income statue, compatibility level with the friend as well as the family and appreciation as regards the sheltering environment together with the committee exam results were used as independent variables in estimation.

Three committee and final exam results obtained from faculty exams system database as electronic records. The information about the other variables which were thought to effect students' success was obtained from student feedback system. The student feedback system is repeated every semester to collect data with the help of valid and reliable scales. After obtaining the electronic data from two different sources, the data were transformed into statistical software data sheet. In the fallowing paragraphs, basic information about ANN will be included. Because, ANN is relatively less known and less used for educational estimation problems when compared with regression analysis.

Artificial neurons or *nodes* are the primary processing elements of neural networks. In mathematical model of ANN, synapses show connection weights and are related to input signals. Furthermore, a transfer function defines the nonlinear characteristic of neurons. The weighted sum of the input signals which represents the neuron impulse is computed and then transformed using the transfer function. Setting the weights in accordance to the chosen learning algorithm, the learning capability of neuron is obtained (Abraham, 2005).
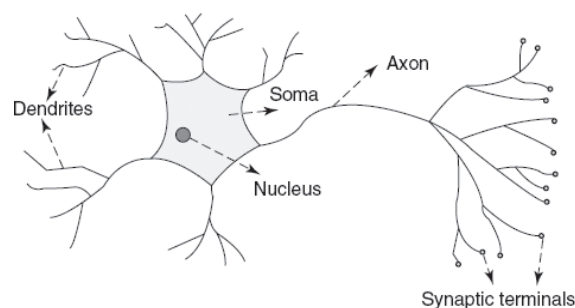

Figure 1: Biological neuron

By using the biological model, the artificial neuron and a multilayered neural network can be described as in Figure 2.
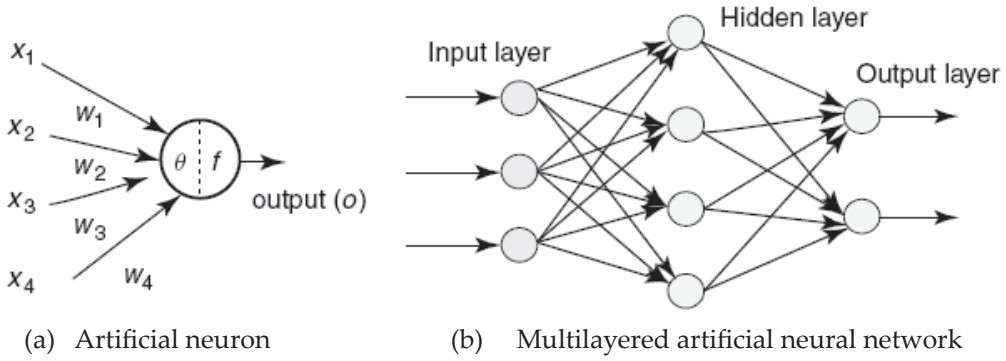
(a) Artificial neuron   (b)   Multilayered artificial neural network

Figure 2: Architecture of an artificial neuron and a multilayered neural network

As seen in Figure 2, *x1, . . . , xn* show the inputs and the neuron output *O* can be computed using the following formula:

$$O = f(net) = f\left(\sum_{j=1}^{n} w_j x_j\right) \qquad (1)$$

where w_j is the weight vector, and the function *f(net)* represents the activation (transfer) function. The variable *net* is defined as a scalar product of the weight and input vectors,

$$net = w^T x = w_1 x_1 + \ldots + w_n x_n \qquad (2)$$

where T is the transpose of a matrix, and, in the simplest case, the output value *O* is computed as

$$O = f(net) = \begin{cases} 1 & if \quad w^T x \geq \emptyset \\ 0 & otherwise \end{cases} \qquad (3)$$

where θ is called the threshold level; and this type of node is called a *linear threshold unit*.

1. Neural Network Architectures

Neural Networks (NNs) include three neuron layers which are input, hidden, and output layers. When the signal flow direction is from input to output layers, these networks are called the feed forward networks. The recurrent network is a commonly used neural network which contains feedback connections. Furthermore, according to the properties and requirement of the applications, there are many neural network architectures such as Elman network, adaptive resonance theory maps, competitive networks, etc. More information about the different neural network architectures and learning algorithms has been given in the study of Bishop (Bishop, 1995).

ANN is implemented as the application which produces the desired set of outputs for the set of inputs. There are two common methods for strengthen the connections, one of which is to set the weights explicitly, using a priori knowledge and the other is to train the neural network by feeding. In the training method, a learning rule is used which teaches patterns by changing the weights. There are three learning approaches which are supervised learning, unsupervised learning, and reinforcement learning (Abraham, 2005). In the *supervised* learning, the desired

outputs are defined by an external teacher and the input vector includes these desired outputs for training. The errors or discrepancies between the desired and actual response for each node in the output layer are found by forward pass. Then, according to the learning rule, the weight changes in the net are computed using these errors and discrepancies. This process is called back propagation and commonly used techniques for this process are delta and perception rules.

In supervised learning, an output is trained to respond to clusters of patterns in the input. In this approach, statistically salient features of the input population can be discovered. Unlike the supervised learning, the system must determine the categories of the patterns; rather the priori set of categories. Reinforcement learning is mapping situations to actions where the learner is not told which actions to take and instead of that the learner must discover the actions to take by trying.

2. Backpropagation Algorithm

We can learn the moving direction of the error in the network by taking the partial derivative of the error with respect to each weight. When the derivative is negative, the rate change of the error increases and it is added to the weight so the error will decrease until it reaches local minima. Positive derivative means that the error is increasing. In this case, a negative value should be added to the weight and vice versa if the derivative is negative (Abraham, 2005). This process starts from the output layer to hidden layer weights, then the hidden layer to input layer weights thus this algorithm is called the *backpropagation algorithm*.

For training, training samples are set and given as input vectors to the network, then the error of the output layer is calculated and the weights of the network are arranged for minimizing the error. For the outputs, the average of all the squared errors (*E*) is computed and the weights are updated separately.

$$\Delta w_{ij}(n) = -\eta * \frac{\delta E}{\delta w_{ij}} + \alpha * \Delta w_{ij}(n-1) \qquad (4)$$

where η and α are the learning rate and momentum respectively. The effect of past weight changes on the current direction of movement defines the momentum. The η and α values affect the training success and the speed of the neural network learning. Therefore, the choice of these parameters is important. The backpropagation learning proves that the network with enough hidden layers can approximate any nonlinear function to arbitrary accuracy. Therefore, backpropagation learning neural network can be used for signal prediction and system modeling.

3. Conjugate Gradient Descent

Backpropagation cannot reach a global minimum of weight space for some initial weight settings; on the other hand it can reach for different initial weight settings. To optimize the weights, four types of optimization algorithms are used. Three of them are general optimization methods which are quasi-Newton, gradient descent, and conjugate gradients and try to minimize a quadratic error function (Abraham, 2005).

For networks with a large number of weights (more than a few hundred) and/or multiple output units, conjugate gradient descent (Bishop, 1995; Shepherd, 1997) is a recommended method of training multilayer networks. It usually performs significantly better than back propagation and can be used where back propagation is used. Furthermore, Levenberg-Marquardt or quasi-Newton may be better for smaller networks and can be used for low-residual regression problems.

While the back propagation sets the network weights after each case, conjugate gradient descent computes the average gradient of the error surface across all cases before updating the weights once at the end of the epoch. Furthermore, conjugate gradient descent does not use learning or momentum parameters. In this work, we used Conjugate Gradient Descent method for our network.

4. Training

For the best training, using a wide range of examples which show all the different characteristics of the problem is recommended especially for more complex problems. Poor data causes an unreliable and unpredictable network. Usually, an epoch number or a particular error threshold is defined for training the network as a condition of stop. The network should not be over trained because in this situation, the network may become too adapted in learning the samples from the training set. As a result, the network may not have the capability to classify samples outside of the training set correctly (Abraham, 2005).

Results

Faculty provides training in an integrated system. Each committee is composed of courses integrated to keep the medical information unity. There are three committees in the half term and a committee exam, composed of 100 multiple-choice questions specified according to the courses take place in each committee, is applied (Q1,Q2,Q3). A final exam similarly composed of 100 questions at the end of the half term is made (F1). F1 is taken as a dependent variable in the study. Data apart from Q1, Q2, Q3 were obtained through surveys.

ANN achieved for the problem:

It is composed of Multilayer Perception Network, 9 input variable, 1 output variable, 27 input neurons, 7 hidden layer neurons, 1 output neurons. Network is trained with conjugate gradient descent algorithm and 65th epoch was chosen as the most successful network as a result of 48 hours-network development since it gives the least testing error. 50% for training, 25% for choosing and the rest, which was 111 in total samples, was used for test purpose (Figure 3).
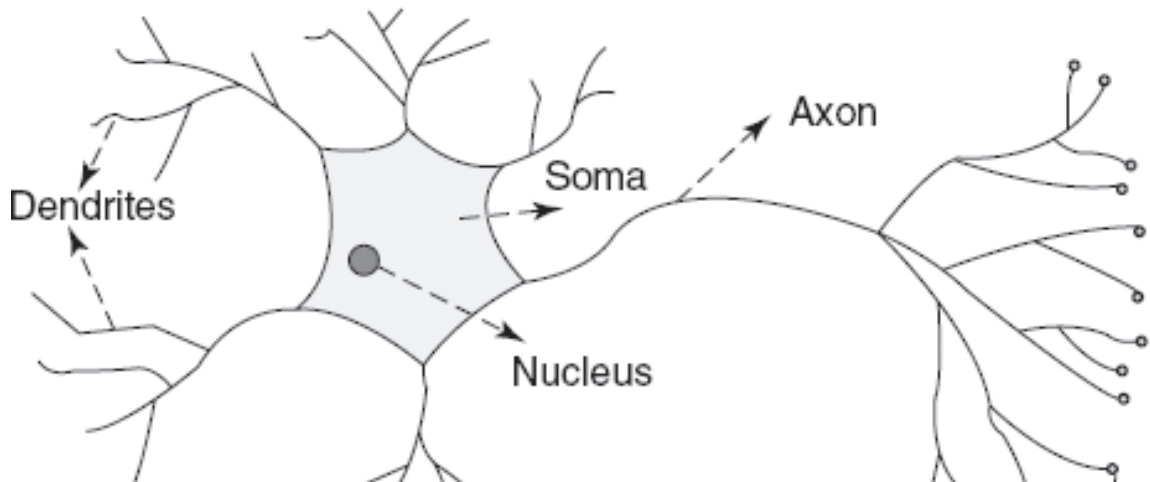

Figure 3: Developed ANN For F1 Estimation

While the three committee marks achieved in the first half term (Q1, Q2, Q3) is provided as the continuous predictor variables; gender (M, F), type of the high school graduated (5 different high school type), medical faculty, appreciation of the physical environment, appreciation of the sheltering environment, degree of matching with the family, degree of matching with friends are provided to the network as categorical predictor variables in terms of displaying the increasing appreciation from 1 to 5 or compatibility. Output variable that means the variable (F1), which is expected to be estimated by the network, is the final exam grade carried out at the end of the half term. When the network performance is examined, results estimated by the network shows similarities with the real exam results Pearson-R=.93. Absolute error ratio is 2.9%. This implies that since the exam results are

evaluated over 100, there is ± 2.9 point deviation in estimation (Table 1, Graph 1).

When the sensitivity analysis is examined for the network, it is considered that the most important estimation variable is the last committee point (Q3, ratio=1.80 rank=1).

Table 1.

ANN Sensitivity Analysis

Input Variable	Ratio	Rank
Q3	1,797782	1
Physical Environment	1,427646	2
Gender	1,355480	3
Sheltering	1,346690	4
Friend Compatibility	1,294957	5
High School Type	1,239995	6
Family Compatibility	1,179669	7
Q2	1,134809	8
Q1	1,088786	9

When multiple regression analysis is carried out with the same variables, Pearson-R 0,85, R-square .73 are calculated. Regression parameters, beta coefficients and significance levels are given in Table 2. The absolute error ratio of the difference with the real values of the regression analysis calculated values is 4.59 %. Similarly, it could be considered that the result to be obtained at the end of the regression analysis has ± 4.59 deviation in average.
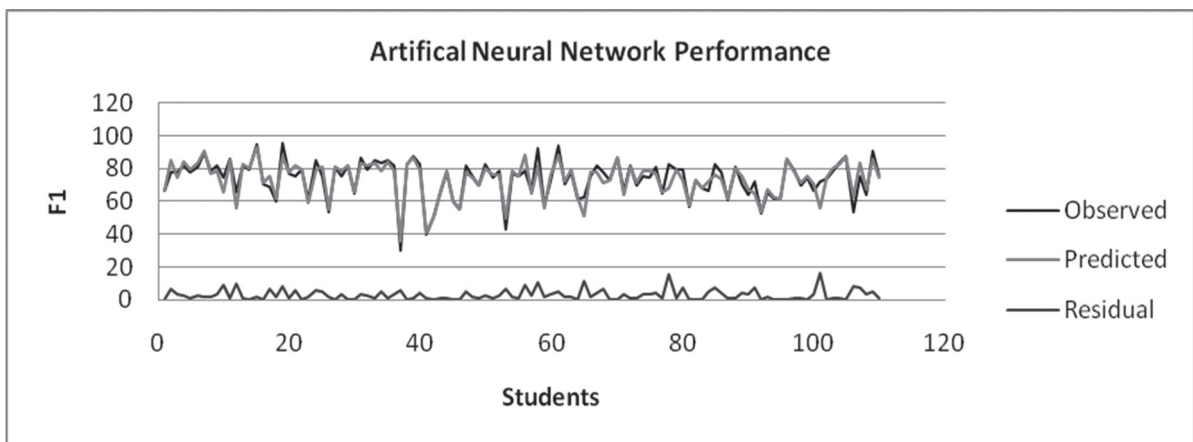
Graph 1: ANN Prediction Results

When the determined variables are used, it could be considered that ANN is able to estimate the student grades in high levels (R=.93, absolute error = ±2.9). According to the multiple regression (R=.85 absolute error = ± 4.59) results with the same variables, ANN displays a much greater performance. If the independent variables provided as input values to ANN are organized by a more detailed study, then it could be possible to have a greater success (Pulito, 2007). Those results go parallel with the other studies carried out in terms of ANN performance (Ashby, 1996; Gorr, 1994; Lykourentzou, 2009; Zaidah, 2007).

KEMAL TURHAN, BURÇİN KURT,
YASEMİN ZEYNEP ENGİN

Although the same variables are used in both methods, the estimation power of the variables varies. It is considered that only Q3, that means the third committee mid-term exam grade, is the variable having the most efficient estimate power in both methods.

Table 2.

Regression Parameters

Regression Variable	Beta Coefficient	P
Q3	0,3973	0,000005
High School Type	-1,3030	0,004346
Q2	0,3074	0,005672
Q1	0,2153	0,009330
Friend Compatibility	-0,9727	0,121647
Family Compatibility	-0,6565	0,514299
Physical Environment	-0,4275	0,627716
Sheltering	-0,3859	0,672113
Gender	0,2228	0,856065

Discussion and Conclusion

When the studies carried on this issue is examined, it is seen that ANN is also commonly used in estimating student performance in addition to the classical statistical methods (Thobega, 2008). ANN was used in canalizing the students newly commenced the Department of Management to one of the departments of marketing, finance, management information systems and general (Badri, 1999) and it is reported that this model worked better when compared with the classical statistical methods. In another study, it was predicted in which level should the students attend mathematics course with a ANN-based model and it was indicated that it displayed a better performance when compared with ANN discriminant analyses (Sheel, 2002). In a study pre-supposing the success of the final exam by the help of the mid-term grades, it was again indicated that ANN made a more successful prediction when compared with the regression analysis (Lykourentzou, 2009). With a view to estimate the academic performance of the students before accepting them to the Department of Management, it was demonstrated that ANN is more successful especially in estimating the continuous data when compared with discriminant analysis and linear regression. In a similar study, it was indicated that ANN created a better result in estimating student success (Zaidah, 2007). A successful sample of web-based ANN application is given in accepting students in Nigeria (Adewale, 2007). In another study, linear regression and stepwise polynomial regression were compared with ANN and no statistically significant difference was found (Wilpen, 1994).

ANN is not only used for estimating student success, but also it is used for different problems of medical education as in some other studies (Greenes, 1990; Monique, 2000; Vendlinski, 2002).

It is obvious that the adverse outcomes that will appear because of the inability in settling the problems as regards the medical education in the early period is more acute when compared with the other disciplines. Since it is developed as based on human learning model, ANN is probably more effective in solving the non-linear problems when compared with the classical statistical methods. The potential of ANN to be used in other fields of the medical education should be considered such as classification of the students according to their ability; providing the appropriate training programs in terms of knowledge, ability and attitude; confirming the

deficiencies in the training programs and ensuring the student consultancy system to be active during the training period.

This pilot study showed that student achievement can be predicted using regression analysis or ANN. ANN showed higher prediction success on the same data set. To generalize these results further studies should be performed. Our results have shown that advanced level research and the automatic information system implementations should be developed at the end of the studies that aim to gather student performance estimation as well as in F1 exam estimation.

There is an information system, which carries out question preparation, exam management, evaluation of the results and material analysis in the faculty. With this system, additions, which are able to produce dynamic as well as smart warnings, will be carried out by integrating ANN implementations that could be introduced by a detailed analysis as to be F1 point estimation also.

References

Abraham, A. (2005). Artificial neural networks. *Handbook of Measuring System Design.* London: John Wiley and Sons Ltd.

Adewale, O. S., Adebiyi, A. B., Solanke, O.O. (2007). Web-based neural network model for university undergraduate admission selection and placement. *The Pacific Journal of Science and Technology*, 2, 367-384.

Ashby, D. & Kumar, N. (1996). A comparison of neural networks and classical discriminant analysis in anticipating default among high-yield bonds. *Association for Information System 1996 Americas Conference*, Phoenix, Arizona.

Badri, M. A. (1999). A neural network based DSS for advising students in the school of business. *Neural Networks, 99. International Joint Conference*, Washington, DC , USA, 5, 3533-3538.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford University Press London, UK.

Gorr, W. L., Nagin, D. & Szczypula, A. (1994). Comparative study of artificial neural network and statistical models for predicting student grade point averages. *International Journal of Forecasting, 10*(2), 17-34.

Greenes, B. P., Bergeron, A. N. & Morse, R. A. (1990). A Generic Neural Network-Based Tutorial Supervisor For Computer Aided Instruction. *SCAMC Inc.*, 435-439.

Lykourentzou, I., Giannoukos, I., Mpardis, G., Nikolopoulos, V. & Loumos, V. (2009). Early and dynamic student achievement prediction in e-learning courses using neural networks. *Journal of The American Society for Information Science and Technology, 60*(2), 372–380.

McCulloch, W. S. & Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–133.

Monique, F. & Claude, F. (2000). Decision-support and intelligent tutoring systems in medical education. *Clin Invest Med., 23*(4), 266-269.

Paliwal, M. & Kumar, U. A. (2009). A study of academic performance of business school graduates using neural network and statistical techniques. *Expert Systems with Applications, 36*, 7865–7872.

Pulito, A. R., Donnelly, M. B. & Plymale, M. (2007). Factors in faculty evaluation of medical students' performance. *Med Educ., 41*(7), 667-75.

Sheel, S. J., Renner, R. S., Dawsey, S. K. (2002). Alternatives to Math Placement Exams: A Look at Discriminant Analysis, Neural Networks and Ensembles of Networks. Department of Computer Science: Coastal Carolina University, Chico Conway, SC.

Shepherd, A. J. (1997). Second-order methods for neural networks: Fast and reliable training methods for multi-layer perceptions. *Springer-Verlag.*, London.

Thobega, M. & Masole, T.M. (2008). Predicting students' performance on agricultural science examination from forecast grades. *US-China Education Review*, *5*(10), 45-51.

Vendlinski, T. & Stevens, R. (2002). Assessing Student Problem-Solving Skills With Complex Computer-Based Tasks. *The Journal of Technology, Learning and Assessment*, *1*(3), 1-24.

Wilpen, L., Nagin, D., Szczypula, J. (1994). Comparative study of artificial neural network: A statistical models for predicting student grade point averages. *International Journal of Forecasting*, *10*(1), 17-34.

Zaidah, I. & Daliela, R. (2007). Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression. *21ˢᵗ Annual SAS Malaysia Forum*, Kuala Lumpur, 1-16.