KNOWLEDGE DISCOVERY IN ARTIFICIAL NEURAL NETWORKS AND REGRESSION MODELS

Ojo, Adebola K. & Adeyemo, Barnabas A.
Department of Computer Science
University of Ibadan
Ibadan, Nigeria
adebola_ojo@yahoo.co.uk, esanadeyemo@gmail.com
*+2347032736013, +2348034747690

ABSTRACT
In this paper, Artificial Neural Networks (ANN) and Regression Analysis models were considered to determine which of them performs better. Prediction was done using one hidden layer and three processing elements in the ANN model. Furthermore, prediction was done using regression analysis. The parameters of regression model were estimated using Least Square method. To determine the better prediction, mean square errors (MSE) attached to ANN and regression models were used. Seven real series were fitted and predicted with in both models. It was found out that the mean square error attached to ANN model was smaller than regression model which made ANN a better model in prediction.

Keywords: Artificial Neural Networks, Regression, Least Square, Processing Element, Hidden Layer, Mean Square Error.

1. INTRODUCTION
Neural Networks (NN), also commonly referred to as Artificial Neural Networks, are information-processing models inspired by the way the densely interconnected, parallel structure of the brain processes information. In other words, neural networks are simplified mathematical models of biological neural networks. The key element of the NN is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements that are analogous to neurons, and tied together with weighted connections that are analogous to synapses. (Kosta Metaxiotis, 2010).

NN are capable of finding internal representations of interrelations within raw data. NN are considered to be intuitive because they learn by example rather than by following programmed rules. The ability to learn is one of the key aspects of NN. This typical characteristic, together with the simplicity of building and training NN, has encouraged their application to the task of prediction. Because of their inherent non-linearity, NN are able to identify the complex interactions between independent variables without the need for complex functional models to describe the relationships between dependent and independent variables.

Recently, the NN approach has been proposed as a substitute for statistical approaches for classification and prediction problems. The advantages of NN over statistical methods include the ability to classify in the presence of nonlinear relationships and the ability to perform reasonably well using incomplete databases. The comparison of the results from NN and statistical approaches indicated that neural networks offer an accurate alternative to classical methods such as multiple regression or autoregressive models (Feuston & Thurtell, 1994; AlFuhaid et al., 1997).

Although the NN concept was first introduced in 1943 (McCulloh & Pitts, 1943), it was not used extensively until the mid-1980’s owing to the lack of sophisticated algorithms for general applications, and its need for fast computing resources with large storage capacity. Since the 1980’s, various NN architectures and algorithms were developed (e.g. the multi-layer perceptron (MLP) which is generally trained with the error backpropagation algorithm, Hopfield Network, Kohonen Network, etc.). Neural networks techniques are becoming useful as alternate approaches to conventional techniques or as components of integrated systems. They are used to solve complicated practical problems in various areas which include telecommunications. They are widely accepted as a technology offering an alternative way to tackle complex and ill-defined problems. They are being used in diverse applications in telecommunications.

ANNs are members of a family of statistical techniques, and flexible nonlinear regression models, discriminate models, data reduction models, and nonlinear dynamic systems (Sarle, 1994; Cheng and Tetterington, 1994). They are trainable analytic tools that attempt to mimic information processing patterns in the brain. Because they do not necessarily require assumptions about population distribution, economists, mathematicians and statisticians are increasingly using ANNs for data analysis. Not only do they not require assumptions about the underlying population but are also powerful forecasting tools that draw on the most recent developments in artificial intelligence research. Neural networks are used in a widening range of applications, including airline security control, investment management and risk control (Brockett, Cooper, Golden, and Pitaktong, 1994), bank insolvency Prediction (Al-Shayea, El-Refae, El-Iter and Al-Zaytoonah, 2010), industrial management and production (Satake, Morikawa, and Nakamura, 1994; Eberts and Habibi, 1995), as well as in forecasting stock price indexes and derivative securities (Hutchinson, Poggio, and Lo, 1994; Li, 1994; Fish, Barnes, and Milam, 1995; Shachmurove and Witkowska, 2001), and predicting exchange rates (Kuan and Liu, 1995) and thrift failures.
Consequently, NN models have been used extensively as a tool for modeling, control, forecasting, and optimization in many fields of engineering and sciences such as process control, manufacturing, nuclear engineering, and pattern recognition. So also, regression analysis is used when two or more variables are thought to be systematically connected by a linear relationship. Regression analysis involves finding the best straight line relationship to explain how the variation in an outcome (or dependent) variable, \(Y\), depends on the variation in a predictor (or independent or explanatory) variable, \(X\). (Vinsnes et al 2001)

### 2. MATERIALS AND METHODS

#### 2.1 Artificial Neural Networks

These are predictive models loosely based on the action of biological neurons. They provide models of data relationships through highly interconnected, simulated “neurons” that accept inputs, apply weighting coefficients and feed their output to other “neurons” which continue the process through the network to the eventual output. Some neurons may send feedback to earlier neurons in the network. Neural networks are “trained” to deliver the desired result by an iterative (and often lengthy) process where the weights applied to each input at each neuron are adjusted to optimize the desired output.

**Input Layer** — A vector of predictor variable values \((x_1, ..., x_d)\) is presented to the input layer. The input layer (or processing before the input layer) standardizes these values so that the range of each variable is -1 to 1. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the bias that is fed to each of the hidden layers; the bias is multiplied by a weight and added to the sum going into the neuron.

**Hidden Layer** — Arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight \((w_{ij})\), and the resulting weighted values are added together producing a combined value \(u_i\). The weighted sum \((u_i)\) is fed into a transfer function, \(\sigma\), which outputs a value \(h_i\). The outputs from the hidden layer are distributed to the output layer.

**Output Layer** — Arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight \((w_{kj})\), and the resulting weighted values are added together producing a combined value \(v_j\). The weighted sum \((v_j)\) is fed into a transfer function, \(\sigma\), which outputs a value \(y_j\). The \(y\) values are the outputs of the network.

If a regression analysis is being performed with a continuous target variable, then there is a single neuron in the output layer, and it generates a single \(y\) value. For classification problems with categorical target variables, there are \(N\) neurons in the output layer producing \(N\) values, one for each of the \(N\) categories of the target variable.

#### 2.2 The Neural Networks Topology

This neural network is called Time-Lagged Feedforward Network (TLFN). It is a Multi-Layer Perceptron (MLP) with memory components to store past values of the data in the network. The memory components allow the network to learn relationships over time. It is the most common temporal supervised neural network. It consists of multiple layers of processing elements (PEs) connected in a feedforward fashion. The PEs in NeuroSolutions are the orange circular icons and are called axons. The connections between the PEs are the icons with horizontal and diagonal lines between the axons and are called synapses. NeuroSolutions uses the backpropagation of errors to train the MLP. The smaller icons on top of the axons and synapses are called backpropagation components and pass the error backwards from the end of the network to the beginning. The green axons on top of the backpropagation components are called gradient search components and adjust the weights contained in the synapses and axons – this is how the network is trained.

Networks constructed using the “low complexity” setting has one hidden layer. Those with medium or high complexity have two hidden layers. The number of PEs in the first hidden layer is contained in the properties of the 2nd AXON from the left. The number of PEs in the second hidden layer is contained in the properties of the 3rd AXON from the left. The rightmost axon is called the criterion and reads the desired file from the attached file component and determines the error in the network. The axon 2nd from the right is the output axon and generates the actual network outputs. The axon on the far left is...
called the input axon and doesn’t do anything but accept the input from the file component.

2.3 Linear Regression Model
Here we wish to determine the relationship between a single independent variable X and a dependent variable Y. The independent variable X is assumed to be a continuous mathematical variable controllable by the experiment suppose that the true relationship between Y and X is linear and that the observation Y at each level of X is a random variable. The expected value of Y for each value of X is E(Y/X) = a+bx where a and b are unknown constants. We also assume that each observation Y can be described by model:

\[ Y = \beta_0 + \beta_1 X + \epsilon_i \]

Where \( \epsilon_i \) is random variable with

\[ E(\epsilon_i) = 0 \]
\[ Var(\epsilon_i) = \sigma^2 \]
\[ E(\epsilon_i, \epsilon_j) = 0 \]

and \( E(X_i, \epsilon_i) = 0 \)

Where Y is a dependent variable
X is an independent variable
E is an error or disturbance term

2.4 Least Square Estimators
To estimate \( \beta_0 \) and \( \beta_1 \) we consider n observations \((X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n)\). If we write the line of best fit \( \hat{Y} = \beta_0 + \beta_1 X \) where \( \hat{Y} \) is used to distinguish between the observed value Y and the corresponding value \( \hat{Y} \) on the line, then the least square criterion required that we minimized the sum of the square of the deviation between Y and \( \hat{Y} \) i.e. \( Y = \beta_0 + \beta_1 X + \epsilon_i \)

Then \( \epsilon_i = Y - \beta_0 - \beta_1 X \)

Squaring both sides,
\[ \epsilon_i^2 = (Y - \beta_0 - \beta_1 X)^2 \]

And summing over the n-observation we have,

\[ \sum \epsilon_i^2 = \sum (Y - \beta_0 - \beta_1 X)^2 \]

Let \( L = \sum \epsilon_i^2 \) and minimizing L w.r.t. \( \beta_0 \) and \( \beta_1 \) and set each to zero.

\[ \frac{\partial L}{\partial \beta_0} = 0 \]
\[ \frac{\partial L}{\partial \beta_1} = 0 \]
\[ \frac{\partial L}{\partial \beta_0} = -2 \sum (Y - \beta_0 - \beta_1 X) \]
\[ \frac{\partial L}{\partial \beta_1} = -2 \sum (Y - \beta_0 - \beta_1 X) \]
\[ \beta_0 = \frac{\sum Y - \beta_0 \sum X}{n} = \bar{Y} - \bar{X} \bar{Y} \]

Substituting \( \beta_0 \) in equation (2) we have,

\[ \beta_1 = \frac{n \sum Y - \beta_0 \sum X}{n \sum X^2 - (\sum X)^2} \]

The regression equation or the line of best fit is now
\[ \hat{Y} = \beta_0 + \beta_1 X \]

Note that \( \beta_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} \)
2.5 Prediction
Prediction can be said to be expectation about a given aspect of social behaviour that may be verified by subsequent observations. Within this general conception, the term is used in two principal senses: for deductions from known to unknown events within a conceptually static system and for statement about future outcomes based on recurring sequences of events. Given the sequence $X_t$ which is zero mean stationary stochastic process, we want to predict the value of $X_{t+m}$, that is, $m$ step ahead predictor and denote by $X_t(m)$ (Shangodoyin and Ojo 2002).

3. RESULTS AND DISCUSSION
The real series used is the payment and consumption lists of a telecommunication industry in Nigeria. Regression and ANN models are fitted using the series.

3.1 Artificial Neural Networks (ANN) Learning Curves
Payment list of telecommunication industry in Nigeria was considered as our input and using the model - one hidden layer and three PEs - the learning curve was generated which showed the plot of the MSE across iterations.

3.2 Prediction Equations for Regression Models:
Series 1 to 7

\[
\hat{Y} = 2.6E + 08 - 1052975X_1 - 0.395X_2 \quad (1)
\]
\[
\hat{Y} = 1.7E + 07 - 252186X_1 - 0.109X_2 \quad (2)
\]
\[
\hat{Y} = 3.6E + 08 - 442018X_1 - 0.205X_2 \quad (3)
\]
\[
\hat{Y} = 3.8E + 07 - 67128X_1 + 3.489E - 03X_2 \quad (4)
\]
\[
\hat{Y} = 2.7E + 07 - 14135X_1 - 2.76E - 02X_2 \quad (5)
\]
\[
\hat{Y} = 2.9E + 07 - 60156.1X_1 - 2.14E.02X_2 \quad (6)
\]
\[
\hat{Y} = 2.4E + 08 + 831100.0X_1 - 0.222X_2 \quad (7)
\]
### 3.3 Prediction Performances of Regression and Artificial Neural Networks (ANN) Models

Table 1: Prediction Performances of Regression Models

<table>
<thead>
<tr>
<th>S/N</th>
<th>Series 1</th>
<th>Series 2</th>
<th>Series 3</th>
<th>Series 4</th>
<th>Series 5</th>
<th>Series 6</th>
<th>Series 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64000266.71</td>
<td>1282013.347</td>
<td>139338558.40</td>
<td>38241603.75</td>
<td>22806333.45</td>
<td>24070328.35</td>
<td>96216537.92</td>
</tr>
<tr>
<td>2</td>
<td>68261000.42</td>
<td>10762905.03</td>
<td>128425945.70</td>
<td>37639085.40</td>
<td>22619523.54</td>
<td>23458775.82</td>
<td>98415342.43</td>
</tr>
<tr>
<td>3</td>
<td>63803448.58</td>
<td>10618854.97</td>
<td>122052070.50</td>
<td>37003535.31</td>
<td>22025949.33</td>
<td>23768579.09</td>
<td>95358822.46</td>
</tr>
<tr>
<td>4</td>
<td>59345896.74</td>
<td>10474804.92</td>
<td>115678195.60</td>
<td>36367985.22</td>
<td>21432375.12</td>
<td>24078382.35</td>
<td>92302302.48</td>
</tr>
<tr>
<td>5</td>
<td>65493560.77</td>
<td>10773409.65</td>
<td>121591915.20</td>
<td>35682160.66</td>
<td>21163213.93</td>
<td>24029649.94</td>
<td>109945299.3</td>
</tr>
<tr>
<td>6</td>
<td>69623620.90</td>
<td>10945982.81</td>
<td>128041342.10</td>
<td>35046282.02</td>
<td>20916574.05</td>
<td>23971478.07</td>
<td>114699529.3</td>
</tr>
<tr>
<td>7</td>
<td>75410835.14</td>
<td>10866862.61</td>
<td>122364909.70</td>
<td>41894257.77</td>
<td>20885262.08</td>
<td>28261762.63</td>
<td>108909561.8</td>
</tr>
<tr>
<td>8</td>
<td>86343089.26</td>
<td>10396441.69</td>
<td>119169344.60</td>
<td>33717992.94</td>
<td>20707232.59</td>
<td>23830030.90</td>
<td>126465100.0</td>
</tr>
<tr>
<td>9</td>
<td>84924026.37</td>
<td>11393559.95</td>
<td>118288468.20</td>
<td>33058289.30</td>
<td>20536604.60</td>
<td>23727377.09</td>
<td>121009114.8</td>
</tr>
<tr>
<td>10</td>
<td>86412407.96</td>
<td>11092297.01</td>
<td>115360871.00</td>
<td>32404320.22</td>
<td>20664910.88</td>
<td>23776715.81</td>
<td>109877400.2</td>
</tr>
<tr>
<td>11</td>
<td>85624057.09</td>
<td>11334428.50</td>
<td>115269559.90</td>
<td>31717208.20</td>
<td>20427126.98</td>
<td>23596778.36</td>
<td>111907942.3</td>
</tr>
<tr>
<td>12</td>
<td>91683956.83</td>
<td>10433678.67</td>
<td>97146106.11</td>
<td>31170611.76</td>
<td>19597070.93</td>
<td>23070980.77</td>
<td>97797811.56</td>
</tr>
<tr>
<td>13</td>
<td>75190700.70</td>
<td>10221370.22</td>
<td>87536144.33</td>
<td>30549464.18</td>
<td>19248198.33</td>
<td>22781637.35</td>
<td>78395338.36</td>
</tr>
<tr>
<td>14</td>
<td>58830904.72</td>
<td>9479918.44</td>
<td>90916857.32</td>
<td>29911393.28</td>
<td>19273187.32</td>
<td>22687018.43</td>
<td>94558013.27</td>
</tr>
<tr>
<td>15</td>
<td>43446115.47</td>
<td>9673292.31</td>
<td>80845251.29</td>
<td>29265155.13</td>
<td>19120200.36</td>
<td>22665100.40</td>
<td>66193016.25</td>
</tr>
<tr>
<td>16</td>
<td>38232021.35</td>
<td>9758512.74</td>
<td>80713633.38</td>
<td>28573841.45</td>
<td>19182321.39</td>
<td>23342287.77</td>
<td>70731556.27</td>
</tr>
<tr>
<td>17</td>
<td>33071927.23</td>
<td>9843732.28</td>
<td>80582015.46</td>
<td>2789527.67</td>
<td>19244442.41</td>
<td>24019475.13</td>
<td>75270906.28</td>
</tr>
<tr>
<td>18</td>
<td>17777349.35</td>
<td>9783130.00</td>
<td>73396476.05</td>
<td>27260735.29</td>
<td>19160099.70</td>
<td>23930827.36</td>
<td>66564204.30</td>
</tr>
<tr>
<td>19</td>
<td>2922868.84</td>
<td>9136223.16</td>
<td>6575262.65</td>
<td>26630157.51</td>
<td>18698439.52</td>
<td>23616871.69</td>
<td>58052031.29</td>
</tr>
<tr>
<td>20</td>
<td>94082566.00</td>
<td>8869369.06</td>
<td>10038953.60</td>
<td>25780617.28</td>
<td>19205200.58</td>
<td>23942157.13</td>
<td>78845503.48</td>
</tr>
<tr>
<td>21</td>
<td>9377009.18</td>
<td>9336668.38</td>
<td>110691975.40</td>
<td>25073373.47</td>
<td>19254382.47</td>
<td>23921196.95</td>
<td>99402471.63</td>
</tr>
<tr>
<td>22</td>
<td>98706338.00</td>
<td>9141748.04</td>
<td>115485525.90</td>
<td>24434527.17</td>
<td>19131731.98</td>
<td>24043421.79</td>
<td>109899083.6</td>
</tr>
<tr>
<td>23</td>
<td>93288887.13</td>
<td>8990668.54</td>
<td>122004444.00</td>
<td>23778337.95</td>
<td>18918002.18</td>
<td>23935050.92</td>
<td>141690878.1</td>
</tr>
<tr>
<td>24</td>
<td>69495486.08</td>
<td>6361472.10</td>
<td>112097057.60</td>
<td>22977463.65</td>
<td>18445505.68</td>
<td>22867533.55</td>
<td>139762700.1</td>
</tr>
</tbody>
</table>

MSE 5130795 249447.1 4066876 1048972 256039.8 215749.4 4597734
Table 2: Prediction Performances of Neural Networks Models

<table>
<thead>
<tr>
<th>S/N</th>
<th>Series 1</th>
<th>Series 2</th>
<th>Series 3</th>
<th>Series 4</th>
<th>Series 5</th>
<th>Series 6</th>
<th>Series 7</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>92627512.00</td>
<td>116920840.00</td>
<td>18413458.00</td>
<td>114416600.00</td>
<td>31271328.00</td>
<td>6625410.5</td>
<td>25266454.0</td>
<td>0.016421</td>
</tr>
<tr>
<td>2</td>
<td>73450688.00</td>
<td>96658064.00</td>
<td>16162011.00</td>
<td>64530476.00</td>
<td>35329680.00</td>
<td>12975855.0</td>
<td>24398382.0</td>
<td>0.031317</td>
</tr>
<tr>
<td>3</td>
<td>70684096.00</td>
<td>140757264.00</td>
<td>96658064.00</td>
<td>16162011.00</td>
<td>64530476.00</td>
<td>12975855.0</td>
<td>24398382.0</td>
<td>0.028438</td>
</tr>
<tr>
<td>4</td>
<td>72926232.00</td>
<td>112658800.00</td>
<td>22023948.00</td>
<td>12352656.00</td>
<td>33289278.00</td>
<td>9729835.0</td>
<td>24404934.0</td>
<td>0.014321</td>
</tr>
<tr>
<td>5</td>
<td>77139792.00</td>
<td>105567464.00</td>
<td>23440324.00</td>
<td>121069432.00</td>
<td>33289278.00</td>
<td>9729835.0</td>
<td>24404934.0</td>
<td>0.025876</td>
</tr>
<tr>
<td>6</td>
<td>77474184.00</td>
<td>125376424.00</td>
<td>23440324.00</td>
<td>121069432.00</td>
<td>33289278.00</td>
<td>9729835.0</td>
<td>24404934.0</td>
<td>0.052776</td>
</tr>
<tr>
<td>7</td>
<td>87508632.00</td>
<td>122989428.00</td>
<td>12352656.00</td>
<td>33289278.00</td>
<td>9729835.0</td>
<td>24404934.0</td>
<td>0.034173</td>
<td></td>
</tr>
</tbody>
</table>

From Tables 1 and 2 above, we could see the prediction of Artificial Neural Networks and regression models for series 1 to 7 resulting from regression equation model and training of runs as a function of epoch in artificial neural networks. Also, from the two tables, we could see the mean square error (MSE) attached to each of the series. The MSE attached to artificial neural networks was smaller when compared with the regression models, his implies that artificial neural networks model perform better than regression model when it comes to prediction.

4. CONCLUSION
The study shows the efficacy of artificial neural network in prediction. Having examined statistical and artificial neural networks models for prediction, the future values of payment were predicted using the appropriate models and the performance of the artificial neural networks compared with the regression. Artificial Neural Networks had the ability to analyze complex patterns quickly and with the higher degree of accuracy. Artificial Neural Networks had the lowest mean square error compared to statistical models.
REFERENCES


BIBLIOGRAPHY


