MODELING GHANAIAN TOURIST ARRIVALS USING ARTIFICIAL NEURAL NETWORKS

BY

Eben Allen Tetteh
(BSc Mathematics)

A Thesis Submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Technology in partial fulfillment of the requirement for the degree of

MASTER OF PHILOSOPHY APPLIED MATHEMATICS

DEPARTMENT OF MATHEMATICS
College of Sciences

June, 2013

DECLARATION.

I hereby declare that this submission is my own work and that, to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the University, except where due acknowledgement has been made in the text.

EBEN ALLEN TETTEH
PG 5072810

....	....	14/05/2013
Student Name & ID	Signature	Date

Certified by:

E. OWUSU-ANSAH

....	....	14/05/2013
Supervisor's Name	Signature	Date

Certified by:

MR. KWAKU DARKWAH

....	....	31/5/2013
Head of Dept. Name	Signature	Date

DEDICATION

This thesis is dedicated to the Almighty God for His abundant love and grace shown me all these years. To my good and new friend Myrtle Oforiwaa Agyapong, my aunty Miss Rose Tetteh, my mum Miss Mary Ama and my dad Mr. Emmanuel Offei Tetteh. Thank you all for your love and support.

ACKNOWLEGDEMENT

I thank the Almighty God for this successful thesis. I am grateful to my supervisor Mr. E. O. Ansah for guiding me through this study. Without him this work would have been a fiasco. My sincere thanks also go to Mr. Kwaku Darkwa, Head of Department of Mathematics, Mr. Francis Boateng and Dr. Peter Amoako-Yirenkyi of KNUST. To all staff of School of Graduate Studies especially Mr. Joseph Owusu-Sekyere and Mr. Kumi Gyan for their love. To Theresah Amankwah and Priscilla Anti I say a big thank you and God bless you all.

Contents

1 Introduction 1

1.1 Background Study 1

1.2 Problem Statement 2

1.3 Objective 2

1.4 Justification 3

1.5 Methodology 3

1.6 Organization of Thesis 4

2 Literature Review 5

2.1 Introduction 5

2.2 Artificial Neural Networks 5

2.3 Tourism 11

3 Methodology 18

3.1 Introduction 18

3.2 The Human Brain 18

3.3 Mathematical Representation 20

3.4 Activation Functions 23

3.5 Learning Mechanisms 24

3.6 Components of Artificial Neural Networks (ANN) 27

3.7 Artificial Neural Network Architectures 28

3.8 Feedforward Networks 29

3.9 Recurrent Networks 30

3.10 Linear and Nonlinear Temporal Neuron Models 31

3.11 Nonlinear Spatio-temporal Neuron Model 35

3.12 Recurrent Network with Hidden Neuron Feedback 36

3.13 The Gradient Descent Algorithm 38

3.14 The Backpropagation Algorithm 40

3.15 Steps in using Backpropagation Algorithm 43

3.16 Example Training using Backpropagation Algorithm 44

4 Results and Discussion **49**

4.1 Introduction 49

4.2 Data Acquisition and Preparation 49

4.3 Creating the Network Architectures 53

4.4 Training of the Networks 55

4.5 Results and Discussion 56

5 Conclusion **68**

5.1 Introduction 68

5.2 Summary of Findings 68

5.3 Recommendation 69

5.4 Conclusion 69

5.5 Appendix 71

5.6 Reference 73

Abstract

In this study we have used two types of ANN architectures to forecast the yearly tourist demands of Ghana. We did this analysis with the purpose of applying two types of ANN networks that can increase the forecasting accuracy of tourists arrivals in Ghana. The two types of ANN, i.e. the Feedforward and the Recurrent (NARX), were trained and afterwards tested on the tourist arrival data.

The recurrent network predicted better than the feedforward network. In addition, both networks did train well but the recurrent was able to generalize by predicting the test data with minimal errors. We also found out that a small learning rate is more effective since it slows down the learning process and allows the network to adapt very well to the data. With large learning rates it is very likely for the network to over fit. Recurrent networks are time consuming but they end up having the best training which can generalize when an appropriate algorithm and a large data size is fed into it. The weights (parameters) obtained in this work would have been very difficult to be obtained if we had used any of the traditional methods.

Chapter 1

Introduction

1.1 Background Study

The first usage of Artificial Neural Networks (ANN) in forecasting was reported in 1987 by Farber and Lapedes. Farber and Lapedes designed a feed forward network that could be used in predicting a dynamic nonlinear time series generated by the logistic map and the Mackey-Glass equation. The outcome of their experiment showed that ANN are very good in modeling and predicting nonlinear time series with excellent accuracy (Zhang, 1997). The results obtained from the research by Farber and Lapedes motivated other researchers to use ANN as a means of forecasting and prediction of time series data.It is in this light that has driven the motivation in the application of ANN in time series forecasting. Statistical problems in areas like medicine, finance, physics and engineering can be solved or tackled efficiently using Artificial Neural Networks. The massive interest in ANN's by many researchers are born from the fact that ANN's operate accurately on nonlinear time series data.

In reality ANN's are closely related to the traditional statistical models like regression, clustering algorithms and nonlinear filtering. The purpose of ANN's are not to eliminate or replace the traditional predicting methods but to act as a compliment to them. This work is therefore to bring to light the power and efficiency of artificial neural networks as a tool for forecasting and prediction.

1.2 Problem Statement

There are many traditional methods that can be used in the predictions and forecasting of data. However it is very clear that most of the traditional methods are inherently linear. Also the nonlinear ones that exist tend out to be very difficult to solve. Because of this reason traditional methods tend to loose track of the actual relationship in a given data and hence produce to some extent models that are not very accurate for predictions. Real world is faced with problems whereby data is highly nonlinear and incomplete. To be able to predict the missing part of the data becomes very difficult if there are no good theories that can help reconstruct the missing part of the data. Most of these problems are nonlinear in nature and hence need appropriate nonlinear models to help reconstruct or predict the missing part with a high level of accuracy. This is where Artificial Neural Networks come into play since they are able to automatically model both linear and nonlinear relationships that exist in a given dataset.

1.3 Objective

The objective of this study is to examine the strength of Artificial Neural Networks (ANN) in predicting or reconstructing a given data set using the Ghana tourist arrivals data set. To the researchers best of knowledge, ANN has not been used as a tool in predicting tourist arrivals in Ghana. In these thesis we want to demonstrate the use of ANN with its related learning algorithms most especially the backpropagation algorithm in reconstructing or predicting tourist arrivals in the tourism sector of Ghana. The results of this which we envisage to be more accurate would serve as a guide to the Ghana Tourist Authority for better planning and coming out with good policies that would help the industry.

1.4 Justification

In recent years computers are able to analyze data, take decisions and learn patterns on their own with less human efforts. These special characteristics of computers are as a result of various artificial intelligence built into programs that run on these computers. In these programs are codes which are written based on Artificial Neural Network algorithms and this accounts for the intelligence of recent computers. In some other cases identifying the structural relationships that exist in a given data set is often problematic and this sometimes hinders the predictability of that data set. ANNs are able to analyze data by learning the patterns the exist among the data set by adapting and generalizing these patterns so that whenever any other new data point is introduced to the network it would be able to predict its outcome.The time series predictions of Artificial Neural Network is non-parametric. This is because ANNs do not need to know how the time series signal is generated (Diaconescu, 2008). All these make ANN a powerful tool for predicting and reconstructing data patterns and hence the motivation of the researcher in applying it to the tourist arrivals data in the Ghana tourist industry.

1.5 Methodology

Following previous works by R. Law and R. Pine of the Hong Kong Polytechnic University in 2004, tourist arrivals could be predicted using economic indicators such as exchange rates, market expenses etc as independent variables. In this research we predicted the arrival in the fifth year using a four year lagged number of arrivals. The model used in the study is represented in equation (1). **Q**, representing the arrivals of tourists in a destination country can be expressed as a function of the lagged arrivals.

$$Q_t = f(Q_{t-1}, Q_{t-2}, ..., Q_{t-n}).......................... (1)$$

where Q is the number of tourist arrivals in the destination country, Ghana. In regression, (1) could be represented mathematically by a multivariate regression model as

$$Q_t = a + bQ_{t-1} + cQ_{t-2} + ... + zQ_{t-n}........... (2)$$

where a is a constant. b, c,...,z and g are the variable coefficients of the lagged tourist arrivals respectively. The independent variables in (1) would be gathered and simulated using ANN.

1.6 Organization of Thesis

The introduction of this thesis is presented in chapter one.Chapter two is devoted to the theoretical framework and literature review of the study of tourism demand.The methodology employed in this study is presented in chapter three.The analysis and simulation results are presented in chapter four. Chapter five encompasses the conclusion and recommendations for further research.

Chapter 2

Literature Review

2.1 Introduction

The world of Artificial Neural Networks (ANN) is endowed with a lot of literature on its theories and applications. The applications of ANN that has appeared in literature over the years and in various research fields give overwhelming and outstanding results. It has been shown that using ANN in time series forecasting is far more efficient than traditional forecasting methods like moving average (3), regression models and exponential smoothening. This chapter of the thesis seeks to bring to bear the various results that have appeared in literature concerning the applications of ANN.

2.2 Artificial Neural Networks

In 1996 a Switz financial analyst, Paolo Tenti, published an article in connection with the applications of ANN in the financial market. The article proposes the usage of recurrent neural networks (one of the many ANN architecture) in predicting the foreign exchange rates of the German Deutschemark. Tenti was motivated to do this research because financial traders at that time argued that the statistical properties that governs foreign exchange rates follows a particular trend hence systematic procedures can be used to foretell the future of that particular financial instrument. On the other hand, the academicians took a very strong stand that these statistical properties has no trend but follow a sort of random walk hypothesis. Years later exchange rates were found to be rather dependent on past changes. This led to the dismissal of the idea that exchange rate returns follow a random walk hypothesis (Tenti, 1996). Before the advent of nonlinear dynamics, statistical tests for the random walk were usually conducted by verifying that there was no linear dependence, or that autocorrelation coefficients were not statistically significant. Evidence has now shown that while there is little linear dependence, the idea of only linear dependence can strongly be rejected, demonstrating the existence of nonlinearities in exchange rates (Brock et al., 1991). In his research, Tenti developed three different types of the recurrent network and used empirical data of the exchange rates to test their performance. He developed two trading strategies and incorporated them into the three networks. Before we continue with the outcome of his research lets take these definitions out of the way. Recurrent networks like any other artificial networks are predisposed to overfitting of data. This problem occurs when the network has been able to learn the patterns in the training data and yet cannot accurately predict an out-of-sample (test) data. To tackle this anormally, the first step is to train the network model as usual with the training data set and then evaluate the network's performance with the test data set. Also one can use any of the network algorithms available to reduce the network size by reducing the number of neurons in the hidden layer and the number of parameters to be estimated (Weigend et al., 1992). In Tenti's research, he trained the network until it converged and then observed the point at which the test data error began to rise. He then took the network weights at the iterate where the test data error was minimum as the estimated parameters. In this case, since the network was evaluated on the test data set, the network generalization was accurate.

Still in the financial world pattern recognition is one of their most challenges and ANN's have been applied most extensively to obtain the solution. Recognition of handwriting can be a daunty and challenging task to people whose field are concern with numbers or handwriting. A typical example is the bank. Almost everyday bank checks are issued by different individuals and it is the duty of the bank authorities to detect the authentic handwriting of the bearer of the check. Al-Omari et al conducted a research on the recognition of Arabic numerals using a feedforward ANN with the backpropagation learning algorithm. They came out with the most popular architecture of Neural Networks that is being used. This consists of an input layer, a hidden layer and an output layer. The number of nodes in the input layer differs according to the feature vector's dimensionality i.e. the input space dimension (Al-Omari et al, 2009). Commonly neural networks are trained, so that a particular input leads to a specific target output. The network is adjusted based on the comparison between the network output and the target. This is done continuously until the network output matches the target. In a research by Al-Omari et al, they built two networks in which the first consisted of three layers i.e. the input layer, the hidden layer and the output layer. The three layers of the first network had 600, 50 and 10 neurons respectively. The input had these number of neurons because the dimension of the input vector was 600. With the output layer it had 10 neurons because the research was to recognise 10 digits i.e. [0,1,. . .,9]. The number of neurons in the hidden layer can be varied and it rests in the hands of the researcher to decide how many neurons in the hidden layer that best suits the application. The architecture of the second network had four layers. ~~It is also~~ similar to the architecture of the first network except that there was one additional hidden layer rather than one hidden layer. The input vector also consists of 600 inputs. The first hidden layer had 250 neurons and the second had 6 neurons (Al-Omar et al, 2009). They took 600 handwritings of the digits from 0 to 9 and fed it into the network. It was observed in the experiment that the second network converged more rapidly than the first network. This was the situation since the number of epochs (iterations) was greater in the first network than in the second network. The accuracy of recognition is 93 percent in the first network and 95 percent in the second network. They found that the recognition accuracy for digits (1, 2, 3, 5, 6 and 7) remains the same in both networks. And the recognition accuracy for digits (0, 4 and 8) was better in the second network, while the recognition accuracy for digit (9) was better in first network. Digit (0) achieved the best average accuracy 96.65 percent in both networks. Digit (9) achieved the least average accuracy 89.95 in both networks. They conclude that the second network with two hidden layers recognized all digits more accurately than the first network with one hidden layer except for digit (9). In conclusion they reached the computer to think or mimic the human brain by the use of neural networks. This recognition starts with acquiring the image to be preprocessed throw a number of steps. In a final conclusion, neural network seems to be better than other techniques used for recognition (Al-Omar et al, 2009).

Koskela et al, 2007, in their research compared three neural network architecture using four separate time series data in a forecasting task. The research constituted of three main ANN architectures which include the Elman, the Finite Impulse Response (FIR) and the Multilayer Perceptron (MLP) networks. These are networks which possess a special kind of properties. The results by Koskela showed that the efficiency of the learning algorithm is a more important factor than the network model used (Koskela et al, 2007). The MLP is a kind of static network which is used to predict time series data that is in a stationary mode. A time series data is said to be stationary if its statistics does not change with time (Koskela et al, 2007). In the MLP the input vector contains previous versions or values of a sample data in their respective time steps and the network is required to predict the value of the data in the next time step. The other two networks that is the Elman and FIR are dynamic networks. Dynamic or recurrent network mechanism would be explained much better in Chapter 3. But before that, the Elman and FIR networks can store past information of the network in memory during the learning process. The experiment by Koskela et al was characterized by four separate data samples. These include the load in an electric network, fluctuations in a far-infrared laser, behaviour of sunspots and an artificially generated time series. All data samples were normalized to values between [-1,1] except the behaviour of sunspots which were normalized between [0,1]. The data samples were divided into two for each sample. The first part which which is the training set starts from the beginning of each sample and the second part which is the test set starts from the bottom. The networks were trained with the backpropagation algorithm. With FIR networks a special type of backpropagation called temporal backpropagation algorithm was implemented in the training. After training they finally concluded that the efficiency of the learning algorithm is a more important factor than the neural network model used. Training of Elman networks was three to ten times slower than for MLP depending on the training data size. For FIR network training time was five to twenty times longer than for MLP. Elman network models the load in an electric network very well and its performance in other tasks is similar as MLP networks performance. For FIR networks trained with temporal backpropagation adequate performance was reached for two prediction tasks.

ANN are trained without the restrictions of a model to derive parameters and also discover relationships that are driven and shaped by the nature of the time series data or any dataset (M. P. Wallace, 2009). ANN are able to construct or build models from any complex relationships that may exist among any given data. The relationships that exist between time variant variables in the financial market can be a series of cycles, trends or even exhibit a non-stationary behaviour. Linear models have been tested on financial time series data in the past to find complex relationships but the non-linear relationships still exist between many financial variables (M. P. Wallace, 2009). Talking about nonlinearities among financial time series data brings the random walk theory into the picture. Financial time series data has been found to follow the random walk theory. But testing the randomness within financial time series data leads to many problems. Even the most advanced nonlinear models are not very efficient to model the behaviour of financial data and predict them. ANN seems to provide the insight into the nature of the relationship between the financial time series data which will be very useful for predictions and forecasting. There are a number of considerations in using neural networks for financial forecasting, however the neural network has an advanced pattern recognition technique, which makes it particularly useful in time series forecasting (M. P. Wallace, 2009).

Daily air pollution predictions are based on the statistical relationships between weather conditions and ambient air pollution concentrations (Comrie et al, 1997). Traditional statistics methods such as multivariate linear regression have been used to model these relationships and reasonable results have been achieved. However it is very clear that these relationships are highly nonlinear and their complexities are very huge. Note that regression models are inherently linear, although curvilinear relationships can be incorporated via polynomial terms in the regression, and known relationships can be pre-specified by transforming a nonlinear predictor variable into a more linear form (e.g., by taking a logarithm) before using it in the model. ANNs are highly robust with respect to underlying data distributions (non-parametric), and no assumptions are made about relationships between variables (unlike the linear or pre-specified curvilinear relationships in regression). Thus, neural networks are well-suited to modeling complex, nonlinear phenomena ~~such~~ as the stock market or tornado formation. To date, neural networks have received limited application to the problem of air pollution forecasting and no studies comparing them to regression techniques have been performed (Comrie, et al, 1997). Comrie et al, 1997 performed a comparative study involving ANN and multiple regression for the predictions of ozone in a range of eight cities. ANN was found to out perform multiple regression in all the eight cities though the results were not overwhelming. The research also showed that the performance of the ANN would have been overwhelming if persistent information had been included in the modeling.

The accuracy of predicting a particular variable in inventory drives the performance of inventory management. Most often than not traditional forecasting methods are employed in such situations. But current researches have shown that ANN most of the time presents accurate forecasting results than the traditional methods. Comparatively the performance of ANN is better than some traditional forecasting methods like Autoregressive Integrated Moving Average (ARIMA) and Moving Average (MA) (Mitrea et al., 2007). This results was established as part of an inventory forecasting by a research group consisting of Mitrea, Lee and Wu. They developed a Nonlinear Autoregressive network (NARX) model which allows exogenous inputs. The experiment consisted of a five month inventory data samples. The data were then simulated using the ARIMA, MA and the NARX. Comparing the three results, its seen that the NARX performed better than the ARIMA and MA. They concluded that, time series forecasting is well suited with ANN than the traditional methods. Since ANN is able to learn and adapt to the patterns and relationships that exist among data set. They also found out that the ability of the ANN to learn is proportional to the number of the training samples. If the number of samples is increased, the learning performance of the ANN can be improved further (Mitrea, Lee and Wu, 2007).

2.3 Tourism

The rich cultural heritage of Ghana formerly the Gold Coast of Africa has made it one of the most visited countries in the sub region by foreign nationals. Ghana's tourism industry can boast of many tourist sites including craft villages, forests and lively beaches all year round. In the whole of West Africa, Ghana is the only country endowed with a canopy walk. The tourism industry of Ghana can boast of about 42 castles and forts inclusive. All these and many more propel many foreign nationals to visit this country. Ghana has now become one of the few destinations for tourists for leisure and pleasure in the subregion. The tourism industry has become one of the major sources of foreign exchange to Ghana and has been rated the highest generator of foreign exchange in the country. This industry ranks third after gold and cocoa when it comes to commodities that generate foreign exchange (Longmatey, 2004). This puts tourism on a high scale of importance when it comes to the development of Ghana. Ghana cannot eliminate tourism if it wants to achieve high economic laurels.

As already stated Ghanaian tourism is one of the biggest foreign exchange earner for the country. Since 2001 which encompasses the scope of this study, the number of tourists arrivals have increased gradually and considerably. Tourist arrivals increased steadily from the year 2001 to 2004 and it decreased the following year which is 2005. These increase was coupled with an increase in tourism receipts. In 2005 though there was a decrease in tourist arrivals the receipts that year increased. This implies that this venture is very important to the Ghanaian economy therefore much attention should be given to it by the authorities in charge. Below are the figures of tourist arrivals, their corresponding receipts and number of hotels and rooms for the past decade i.e. 2001 - 2011.

One of the key definitions of tourism is the movement of a person or group of persons from their country of permanent residence to another country or place for at least 24 hours but not more than a year i.e 365 days (Nunoo et al, 2007). Tourism in Ghana dates back to 1970 when the then government of Ghana set up a committee (Obuarn Committee) and tasked them to classify all tourist resources and avenues for a five year tourism developmental plan. The results presented by the Obuarn Committee. 1972, was subsequently followed by major advancements in the industry. Many foreign nationals and organization saw it to be a gold mine and started investing into the tourism industry of Ghana (Teye, 2002).

In recent years there have been many researches into tourism and its sustainability in Ghana. Tourism in Ghana as an industry is faced with many problems and set backs. These set backs

Table 2.1: Tourist Arrivals and Receipts

Year	Arrivals ('000)	Receipts (US$ million)
2001	438.8	447.8
2002	482.6	519.7
2003	530.8	602.8
2004	583.8	649.4
2005	428.6	836.1
2006	497.1	986.8
2007	586.6	1,172.0
2008	698.069	1,403.1
2009	802.779	1,615.2

Source: Ghana in Figures

hinders the sustainability and growth of the industry and thereby reducing revenue generated at long run. One of the major problems faced by the tourism industry are environmental hazards. This calls for an environmental management system that would seek to address this issue faced by the tourism industry (Nunoo et al, 2007). In their research Nunoo et al, 2007 devise a means to address this pertinent issue.

A research was conducted to see the primary influence of tourism on the social and economic structures of the host country. This study was to investigate the transformation or changes that is effected by tourism. The research employs an econometric model of tourist demand in a developed country and the country under study was Greece. The econometric model aims to improve the tourist products of the Greece. The model is estimated using the Least Squares Method (LSM) (Dritsakis et al, 2004). The tourist demand of a country is dependent on certain conditions that is prevalent in that country. These conditions or variables can include the economic, demographic, technological, psychological, socio-political, etc of that

Table 2.2: Number of hotels, rooms and beds

Year	Number of Hotels	Hotel rooms	Number of beds
2001	1,053	15,453	19,648
2002	1,162	15,992	21,227
2003	1,250	17,352	22,909
2004	1,313	18,022	23,430
2005	1,341	18,675	23,828
2006	1,405	22,467	27,569
2007	1,407	18,683	26,057

Source: Ghana in Figures

country. To verify the relationships among all these variables so as to be able to carry out a complete analysis of trends in international tourism becomes very tedious (Dritsakis et al, 2004). Dritsakis et al, 2004 came out in his research that the main variable that definitely affects demand of tourists positively is Gross National Product (GNP). GNP growth increases disposable income and hence the willingness and ability to consume various goods and services including an increase in tourist demand, whether such a demand refers to number of arrivals and number of nights spent or to sums of tourist foreign exchange (Dritsakis et al, 2004). In summary nine variables were used and the model was chosen so that the variables were related by a log-linear relationship.

Dritsakis et al concluded in his research that the variable measuring the effect of political stability is an important factor in tourist demand in all countries of the model. While the probability of terrorist acts and violent crimes is low, their possible occurrence would have significant consequences on a sensitive traveling public. The general conclusion is that the tourist host-countries have to face a more demanding, more competitive, and an intensely differentiated tourist market which forces policy makers to draw and apply a tourist policy employing diligence, timely planning, responsibility, and realism (Dritsakis et al, 2004).

Tourism demand takes its basis from the consumer theory (Zhou et al, 2007). Consumer theory predicts that the optimal consumption level of a consumer is a function of the consumers income, price of goods the consumer buys, prices of related goods (substitutes and compliments) and other demand factors. Many tourism demand researches have shown that there are two groups of tourism modeling. Zhou et al, 2007, categorized these two groups as the outbound and inbound tourism modeling. With the outbound modeling attention is given to tourists from a specific country to several other countries and with the inbound priority is given to tourists from several countries to a specific country. In modeling an outbound tourism case the expenditure of the tourists is taken to be the dependent variable where as the expenditure of tourism related goods is the depend variable of an inbound tourism model (Zhou et al, 2007). But in a typical tourism model none of these are used whether inbound or outbound because it is very difficult to track data relating to the expenditure of tourists (Anastasopoulos, 1984). Based on this reason many tourism researches use tourist arrivals as the dependent variables. Of the 85 tourism studies reviewed in Crouch (1994), 63 percent chose the number of visitor arrivals as the measure of demand while 48 percent and receipts (Zhou et al, 2007).

In another study, Aslan et al, 2009 applied a dynamic approach to determine or estimate tourism for the Turkish tourism market by using panel data estimation method. Witt and Song, 2000 proposed that the traditional regression methods tend to harbour problems like structural instability of the regression model equation. Also existing empirical research of the international tourism demand in Turkey which is based on the traditional econometric techniques faces the same problem (Var et al, 1998). Based on this fact Aslan et al decided to use the panel data estimation. Traveling by a tourist is dependent on the individuals income. In effect tourist arrivals are dependent on the income of the tourist. In several studies, the national income of the origin country i.e. the country of origin of the tourist is statistically significant on the their arrivals in the host country (Akis, 1998). Alsam et al, 2009 defined the demand of tourism as the ratio of the expenditure on tourism of the origin country to the total expenditure on tourism of the host country. Mathematically demand of tourism, D can be expressed as

$$D = \frac{E_o}{E_h}$$

where E_o is expenditure on tourism of the origin country and E_h total expenditure on tourism of the host country. Aslan et al, 2009 also argued that the motivating factor affecting the traveling of a household is their real personal income which is a measure of the wealth of a household. This synchronizes with Zhou et al, 2007 relating the consumer theory. Other factors include the relative prices in the host country. The dynamic model used in this study provides short and long-run elasticity for the variables of interest. This is an additional advantage over most studies of tourism demand, which are based on static models and only estimate long-run elasticity. This is a substantial improvement, since these models are only valid for short-term predictions (Aslan et al, 2009).

Researches on the new factors that affect tourism demand have been conducted and conclusions have been drawn. The factors that influence the demand of tourism can be classified into two major groups. One of the groups is made up of the current factors that have influence on the demand of tourism and this include the motivation and ability for the tourist to travel. On the other hand is the group comprising of the emerging factors that would affect tourism demand in some time to come. The fact that the consumer or the tourist can decide to change his or her host country or place is an emerging factor. In addition there is an endogenous dynamic within the tourists behaviour and this would characterize the tourism consumer's or the tourists behaviour within the next few years (Lohmann, 2004).

The current factors mentioned above which include external factors like the ability and motivation to travel is a very core of the tourism business. These can cripple the whole idea of tourism demand. For if a tourist does not have the freedom, time and money to travel or does not have the motivation to do so guess what will happen to the demand of tourism. Consumer Behaviour is not a reaction on a single factor but on the whole set of influencing external factors. In addition it is driven by internal factors (e.g.motives, abilities etc.). Thus, the impact of a change in a single external factor is not limited (Lohmann, 2004).

The tourist arrivals variable is still the most popular measure of tourism demand over the past few years. Specifically, this variable was measured by total tourist arrivals from an origin to a host destination, which could be decomposed further into holiday tourist arrivals, business tourist arrivals, tourist arrivals for visiting friends and relatives purposes (Song and Li, 2008).

in the second network. The accuracy of recognition is 93 percent in the first network and 95 percent in the second network. They found that the recognition accuracy for digits (1, 2, 3, 5, 6 and 7) remains the same in both networks. And the recognition accuracy for digits (0, 4 and 8) was better in the second network, while the recognition accuracy for digit (9) was better in first network. Digit (0) achieved the best average accuracy 96.65 percent in both networks. Digit (9) achieved the least average accuracy 89.95 in both networks. They conclude that the second network with two hidden layers recognized all digits more accurately than the first network with one hidden layer except for digit (9). In conclusion they reached the computer to think or mimic the human brain by the use of neural networks. This recognition starts with acquiring the image to be preprocessed throw a number of steps. In a final conclusion, neural network seems to be better than other techniques used for recognition (Al-Omar et al, 2009).

Koskela et al, 2007, in their research compared three neural network architecture using four separate time series data in a forecasting task. The research constituted of three main ANN architectures which include the Elman, the Finite Impulse Response (FIR) and the Multilayer Perceptron (MLP) networks. These are networks which possess a special kind of properties. The results by Koskela showed that the efficiency of the learning algorithm is a more important factor than the network model used (Koskela et al, 2007). The MLP is a kind of static network which is used to predict time series data that is in a stationary mode. A time series data is said to be stationary if its statistics does not change with time (Koskela et al, 2007). In the MLP the input vector contains previous versions or values of a sample data in their respective time steps and the network is required to predict the value of the data in the next time step. The other two networks that is the Elman and FIR are dynamic networks. Dynamic or recurrent network mechanism would be explained much better in Chapter 3. But before that, the Elman and FIR networks can store past information of the network in memory during the learning process. The experiment by Koskela et al was characterized by four separate data samples. These include the load in an electric network, fluctuations in a far-infrared laser, behaviour of sunspots and an artificially generated time series. All data samples were normalized to values between [-1,1] except the behaviour of sunspots which were normalized between [0,1]. The data samples were divided into two for each sample. The first part which which is the training set starts from the beginning of each sample and the second part which is the test set starts from the bottom. The networks were trained with the backpropagation algorithm. With FIR networks a special type of backpropagation called temporal backpropagation algorithm was implemented in the training. After training they finally concluded that the efficiency of the learning algorithm is a more important factor than the neural network model used. Training of Elman networks was three to ten times slower than for MLP depending on the training data size. For FIR network training time was five to twenty times longer than for MLP. Elman network models the load in an electric network very well and its performance in other tasks is similar as MLP networks performance. For FIR networks trained with temporal backpropagation adequate performance was reached for two prediction tasks.

ANN are trained without the restrictions of a model to derive parameters and also discover relationships that are driven and shaped by the nature of the time series data or any dataset (M. P. Wallace, 2009). ANN are able to construct or build models from any complex relationships that ~~may~~ exist ~~among any~~ given data. The relationships that exist between time variant variables in the financial market can be a series of cycles, trends or even exhibit a non-stationary behaviour. Linear models have been tested on financial time series data in the past to find complex relationships but the non-linear relationships still exist between many financial variables (M. P. Wallace, 2009). Talking about nonlinearities among financial time series data brings the random walk theory into the picture. Financial time series data has been found to follow the random walk theory. But testing the randomness within financial time series data leads to many problems. Even the most advanced nonlinear models are not very efficient to model the behaviour of financial data and predict them. ANN seems to provide the insight into the nature of the relationship between the financial time series data which will be very useful for predictions and forecasting. There are a number of considerations in using neural networks for financial forecasting, however the neural network has an advanced pattern recognition technique, which makes it particularly useful in time series forecasting (M. P. Wallace, 2009).

Daily air pollution predictions are based on the statistical relationships between weather conditions and ambient air pollution concentrations (Comrie et al, 1997). Traditional statistics methods such as multivariate linear regression have been used to model these relationships and reasonable results have been achieved. However it is very clear that these relationships are highly nonlinear and their complexities are very huge. Note that regression models are inherently linear, although curvilinear relationships can be incorporated via polynomial terms in the regression, and known relationships can be pre-specified by transforming a nonlinear predictor variable into a more linear form (e.g., by taking a logarithm) before using it in the model. ANNs are highly robust with respect to underlying data distributions (non-parametric), and no assumptions are made about relationships between variables (unlike the linear or pre-specified curvilinear relationships in regression). Thus, neural networks are well-suited to modeling complex, nonlinear phenomena such as the stock market or tornado formation. To date, neural networks have received limited application to the problem of air pollution forecasting and no studies comparing them to regression techniques have been performed (Comrie, et al, 1997). Comrie et al, 1997 performed a comparative study involving ANN and multiple regression for the predictions of ozone in a range of eight cities. ANN was found to out perform multiple regression in all the eight cities though the results were not overwhelming. The research also showed that the performance of the ANN would have been overwhelming if persistent information had been included in the modeling.

The accuracy of predicting a particular variable in inventory drives the performance of inventory management. Most often than not traditional forecasting methods are employed in such situations. But current researches have shown that ANN most of the time presents accurate forecasting results than the traditional methods. Comparatively the performance of ANN is better than some traditional forecasting methods like Autoregressive Integrated Moving Average (ARIMA) and Moving Average (MA) (Mitrea et al., 2007). This results was established as part of an inventory forecasting by a research group consisting of Mitrea, Lee and Wu. They developed a Nonlinear Autoregressive network (NARX) model which allows exogenous inputs. The experiment consisted of a five month inventory data samples. The data were then simulated using the ARIMA, MA and the NARX. Comparing the three results, its seen that the NARX performed better than the ARIMA and MA. They concluded that, time series forecasting is well suited with ANN than the traditional methods. Since ANN is able to learn and adapt to the patterns and relationships that exist among data set. They also found out that the ability of the ANN to learn is proportional to the number of the training samples. If the number of samples is increased, the learning performance of the ANN can be improved further (Mitrea, Lee and Wu, 2007).

2.3 Tourism

The rich cultural heritage of Ghana formerly the Gold Coast of Africa has made it one of the most visited countries in the sub region by foreign nationals. Ghana's tourism industry can boast of many tourist sites including craft villages, forests and lively beaches all year round. In the whole of West Africa, Ghana is the only country endowed with a canopy walk. The tourism industry of Ghana can boast of about 42 castles and forts inclusive. All these and many more propel many foreign nationals to visit this country. Ghana has now become one of the few destinations for tourists for leisure and pleasure in the subregion. The tourism industry has become one of the major sources of foreign exchange to Ghana and has been rated the highest generator of foreign exchange in the country. This industry ranks third after gold and cocoa when it comes to commodities that generate foreign exchange (Longmatey, 2004). This puts tourism on a high scale of importance when it comes to the development of Ghana. Ghana cannot eliminate tourism if it wants to achieve high economic laurels.

As already stated Ghanaian tourism is one of the biggest foreign exchange earner for the country. Since 2001 which encompasses the scope of this study, the number of tourists arrivals have increased gradually and considerably. Tourist arrivals increased steadily from the year 2001 to 2004 and it decreased the following year which is 2005. These increase was coupled with an increase in tourism receipts. In 2005 though there was a decrease in tourist arrivals the receipts that year increased. This implies that this venture is very important to the Ghanaian economy therefore much attention should be given to it by the authorities in charge. Below are the figures of tourist arrivals, their corresponding receipts and number of hotels and rooms for the past decade i.e. 2001 - 2011.

One of the key definitions of tourism is the movement of a person or group of persons from their country of permanent residence to another country or place for at least 24 hours but not more than a year i.e 365 days (Nunoo et al, 2007). Tourism in Ghana dates back to 1970 when the then government of Ghana set up a committee (Obuarn Committee) and tasked them to classify all tourist resources and avenues for a five year tourism developmental plan. The results presented by the Obuarn Committee. 1972, was subsequently followed by major advancements in the industry. Many foreign nationals and organization saw it to be a gold mine and started investing into the tourism industry of Ghana (Teye, 2002).

In recent years there have been many researches into tourism and its sustainability in Ghana. Tourism in Ghana as an industry is faced with many problems and set backs. These set backs

Year	Arrivals ('000)	Receipts (US$ million)
2001	438.8	447.8
2002	482.6	519.7
2003	530.8	602.8
2004	583.8	649.4
2005	428.6	836.1
2006	497.1	986.8
2007	586.6	1,172.0
2008	698.069	1,403.1
2009	802.779	1,615.2

Table 2.1: Tourist Arrivals and Receipts

Source: Ghana in Figures

hinders the sustainability and growth of the industry and thereby reducing revenue generated at long run. One of the major problems faced by the tourism industry are environmental hazards. This calls for an environmental management system that would seek to address this issue faced by the tourism industry (Nunoo et al, 2007). In their research Nunoo et al, 2007 devise a means to address this pertinent issue.

A research was conducted to see the primary influence of tourism on the social and economic structures of the host country. This study was to investigate the transformation or changes that is effected by tourism. The research employs an econometric model of tourist demand in a developed country and the country under study was Greece. The econometric model aims to improve the tourist products of the Greece. The model is estimated using the Least Squares Method (LSM) (Dritsakis et al, 2004). The tourist demand of a country is dependent on certain conditions that is prevalent in that country. These conditions or variables can include the economic, demographic, technological, psychological, socio-political, etc of that

Table 2.2: Number of hotels, rooms and beds

Year	Number of Hotels	Hotel rooms	Number of beds
2001	1,053	15,453	19,648
2002	1,162	15,992	21,227
2003	1,250	17,352	22,909
2004	1,313	18,022	23,430
2005	1,341	18,675	23,828
2006	1,405	22,467	27,569
2007	1,407	18,683	26,057

Source: Ghana in Figures

country. To verify the relationships among all these variables so as to be able to carry out a complete analysis of trends in international tourism becomes very tedious (Dritsakis et al, 2004). Dritsakis et al, 2004 came out in his research that the main variable that definitely affects demand of tourists positively is Gross National Product (GNP). GNP growth increases disposable income and hence the willingness and ability to consume various goods and services including an increase in tourist demand, whether such a demand refers to number of arrivals and number of nights spent or to sums of tourist foreign exchange (Dritsakis et al, 2004). In summary nine variables were used and the model was chosen so that the variables were related by a log-linear relationship.

Dritsakis et al concluded in his research that the variable measuring the effect of political stability is an important factor in tourist demand in all countries of the model. While the probability of terrorist acts and violent crimes is low, their possible occurrence would have significant consequences on a sensitive traveling public. The general conclusion is that the tourist host-countries have to face a more demanding, more competitive, and an intensely differentiated tourist market which forces policy makers to draw and apply a tourist policy employing diligence, timely planning, responsibility, and realism (Dritsakis et al, 2004).

Tourism demand takes its basis from the consumer theory (Zhou et al, 2007). Consumer theory predicts that the optimal consumption level of a consumer is a function of the consumers income, price of goods the consumer buys, prices of related goods (substitutes and compliments) and other demand factors. Many tourism demand researches have shown that there are two groups of tourism modeling. Zhou et al, 2007, categorized these two groups as the outbound and inbound tourism modeling. With the outbound modeling attention is given to tourists from a specific country to several other countries and with the inbound priority is given to tourists from several countries to a specific country. In modeling an outbound tourism case the expenditure of the tourists is taken to be the dependent variable where as the expenditure of tourism related goods is the depend variable of an inbound tourism model (Zhou et al, 2007). But in a typical tourism model none of these are used whether inbound or outbound because it is very difficult to track data relating to the expenditure of tourists (Anastasopoulos, 1984). Based on this reason many tourism researches use tourist arrivals as the dependent variables. Of the 85 tourism studies reviewed in Crouch (1994), 63 percent chose the number of visitor arrivals as the measure of demand while 48 percent and receipts (Zhou et al, 2007).

In another study, Aslan et al, 2009 applied a dynamic approach to determine or estimate tourism for the Turkish tourism market by using panel data estimation method. Witt and Song, 2000 proposed that the traditional regression methods tend to harbour problems like structural instability of the regression model equation. Also existing empirical research of the international tourism demand in Turkey which is based on the traditional econometric techniques faces the same problem (Var et al, 1998). Based on this fact Aslan et al decided to use the panel data estimation. Traveling by a tourist is dependent on the individuals income. In effect tourist arrivals are dependent on the income of the tourist. In several studies, the national income of the origin country i.e. the country of origin of the tourist is statistically significant on the their arrivals in the host country (Akis, 1998). Alsam et al, 2009 defined the demand of tourism as the ratio of the expenditure on tourism of the origin country to the total expenditure on tourism of the host country. Mathematically demand of tourism, D can be expressed as

$$D = \frac{E_o}{E_h}$$

where E_o is expenditure on tourism of the origin country and E_h total expenditure on tourism of the host country. Aslan et al, 2009 also argued that the motivating factor affecting the traveling of a household is their real personal income which is a measure of the wealth of a household. This synchronizes with Zhou et al, 2007 relating the consumer theory. Other factors include the relative prices in the host country. The dynamic model used in this study provides short and long-run elasticity for the variables of interest. This is an additional advantage over most studies of tourism demand, which are based on static models and only estimate long-run elasticity. This is a substantial improvement, since these models are only valid for short-term predictions (Aslan et al, 2009).

Researches on the new factors that affect tourism demand have been conducted and conclusions have been drawn. The factors that influence the demand of tourism can be classified into two major groups. One of the groups is made up of the current factors that have influence on the demand of tourism and this include the motivation and ability for the tourist to travel. On the other hand is the group comprising of the emerging factors that would affect tourism demand in some time to come. The fact that the consumer or the tourist can decide to change his or her host country or place is an emerging factor. In addition there is an endogenous dynamic within the tourists behaviour and this would characterize the tourism consumer's or the tourists behaviour within the next few years (Lohmann, 2004).

The current factors mentioned above which include external factors like the ability and motivation to travel is a very core of the tourism business. These can cripple the whole idea of tourism demand. For if a tourist does not have the freedom, time and money to travel or does not have the motivation to do so guess what will happen to the demand of tourism. Consumer Behaviour is not a reaction on a single factor but on the whole set of influencing external factors. In addition it is driven by internal factors (e.g.motives, abilities etc.). Thus, the impact of a change in a single external factor is not limited (Lohmann, 2004).

The tourist arrivals variable is still the most popular measure of tourism demand over the past few years. Specifically, this variable was measured by total tourist arrivals from an origin to a host destination, which could be decomposed further into holiday tourist arrivals, business tourist arrivals, tourist arrivals for visiting friends and relatives purposes (Song and Li, 2008).

Chapter 3

Methodology

3.1 Introduction

This chapter of the thesis will center on the methodology used in the thesis. The main theme of the thesis i.e Artificial Neural Networks (ANN) would also be treated into much details in this chapter. Artificial Neural Networks are very powerful when it comes to dealing with imperfect and nonlinear dataset hence chosen as the method in this thesis.

3.2 The Human Brain

Artificial Neural Networks are inspired by biological neural networks that exit in the human brain and they mimic the learning behaviour of the human brain. The term derives its origin from the human brain or the human nervous system which consists of large amount of biological neurons massively connected in a parallel mode. This inspired researchers to make an effort to design a prototype network that could mimic these large amount of interconnections that exist within the human brain. The human brain can be said to be a highly complex, nonlinear and massively parallel computer that can organize its structural constituents called neurons very well. The number of neurons found in the human brain is about 10-100 billion (World Book

Inc,2001). These neurons are connected with each other in a very complex manner by the aid of synaptic nerves. The complex network of neurons is about 60 trillion and over. The figure below shows a typical human brain with the complex network of neurons.

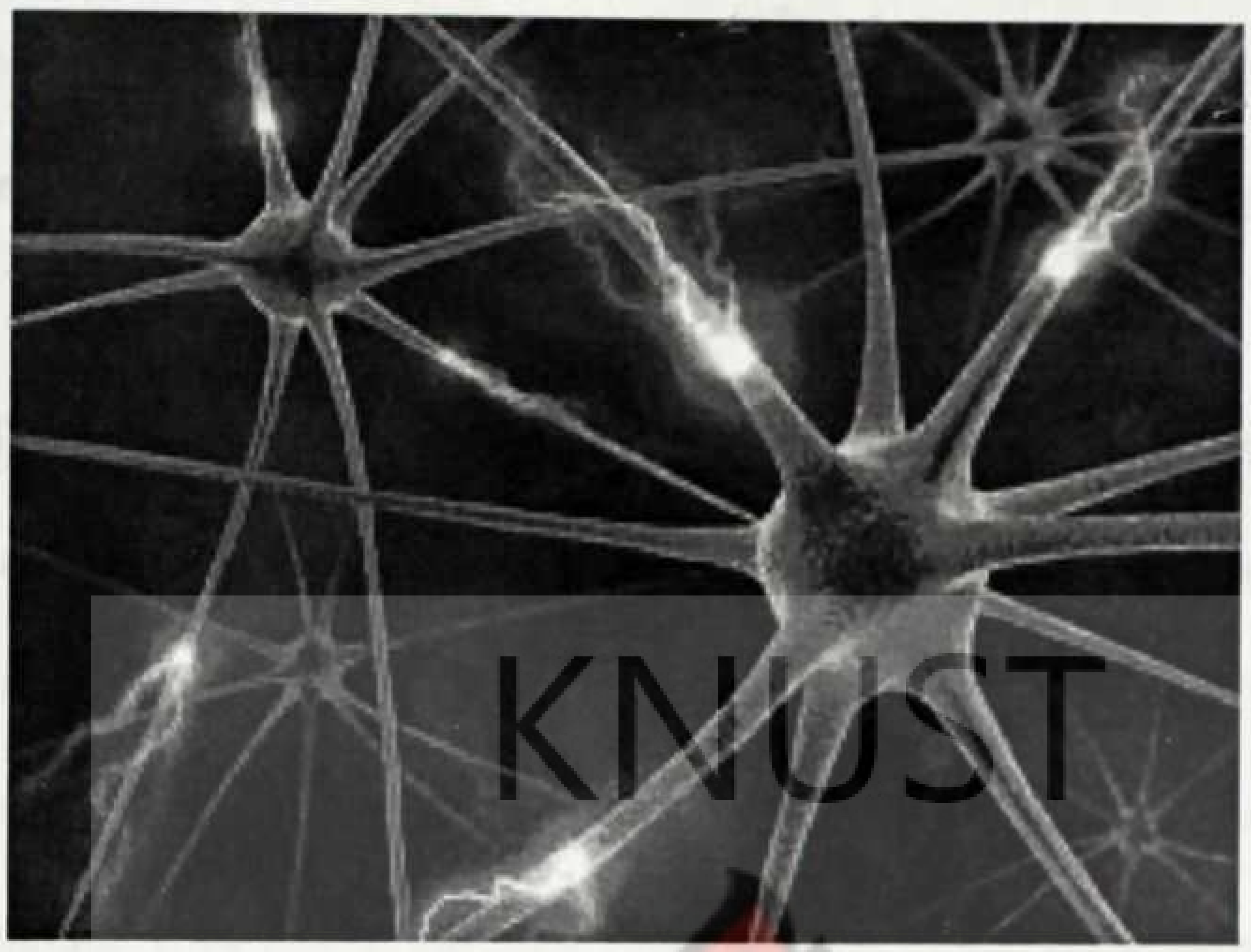

Fig. 3.0: Complex connections between neurons

To give a much deaper illustration to the diagram above, we present in this section a model of the human brain.

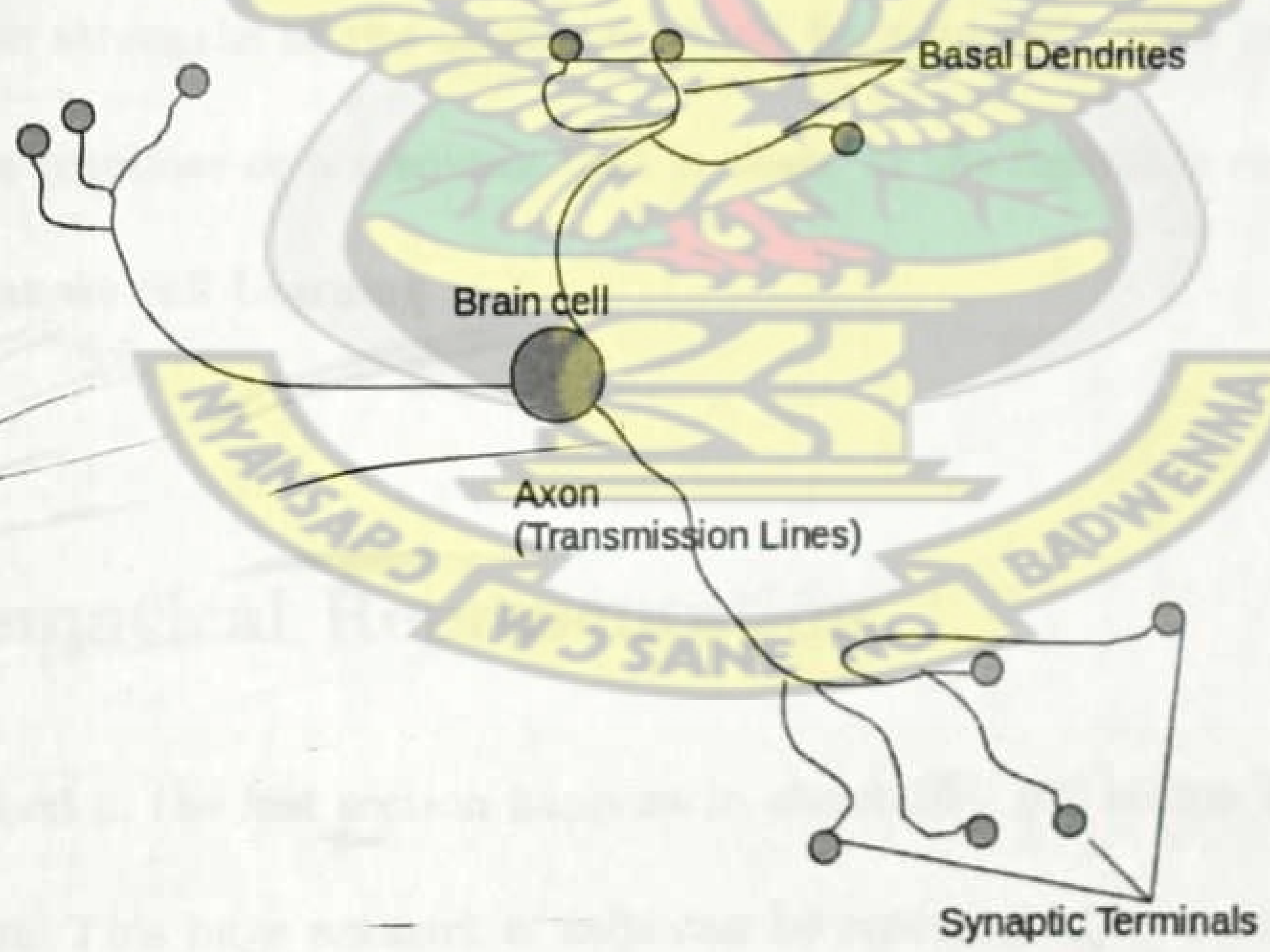

Fig. 3.1: Human Brain

In the diagram above the axons are the lines or routes that carry electrical signals from and to the brain cell. These axons or transmission lines end up to what we call the synaptic terminals. These terminals are used to make connections to other brain cells or cell bodies. The functionality of the dendrites are similar to the axons. The only difference between the two is that the axons are much longer than the dendrites. The basal dendrites acts as the receptive portion or zone to the brain cell. Information to the brain is fed through this part. When information is sent to the brain through the dendrites, they are relayed through the axons to the cell in the form of electrical signals. These signals are then processed in the brain cell after which they are relayed back to the other brain cells through the same media. The amount of signal a dendrite sends depends on its connection strength. If the connection strength of a dendrite is strong then the amount of signal that passes through it would be much more that a dendrite with weak connection strength. Since the connection strength of the dendrites are not the same the signals that passes through them are not of the same strength as explained. Before a signal is sent to the cell, these dendrites already have initial connection strengths and they use that to send the signal to the cell. After processing the signal and the response would be relayed. If the response is not desirable then this same signal would be sent again but this time the connection strengths of the dendrites would be adjusted. This process goes on and on until a desirable response or a response that is closer to the desirable response is achieved. This process is what we call Learning.

3.3 Mathematical Representation

The process described in the last section happens in about 10 - 100 billion network of neurons or cells in the brain. This huge network of cells can be represented by a single simple neuron which will perform the same function as the network. Below is a simple neuron model of the network.

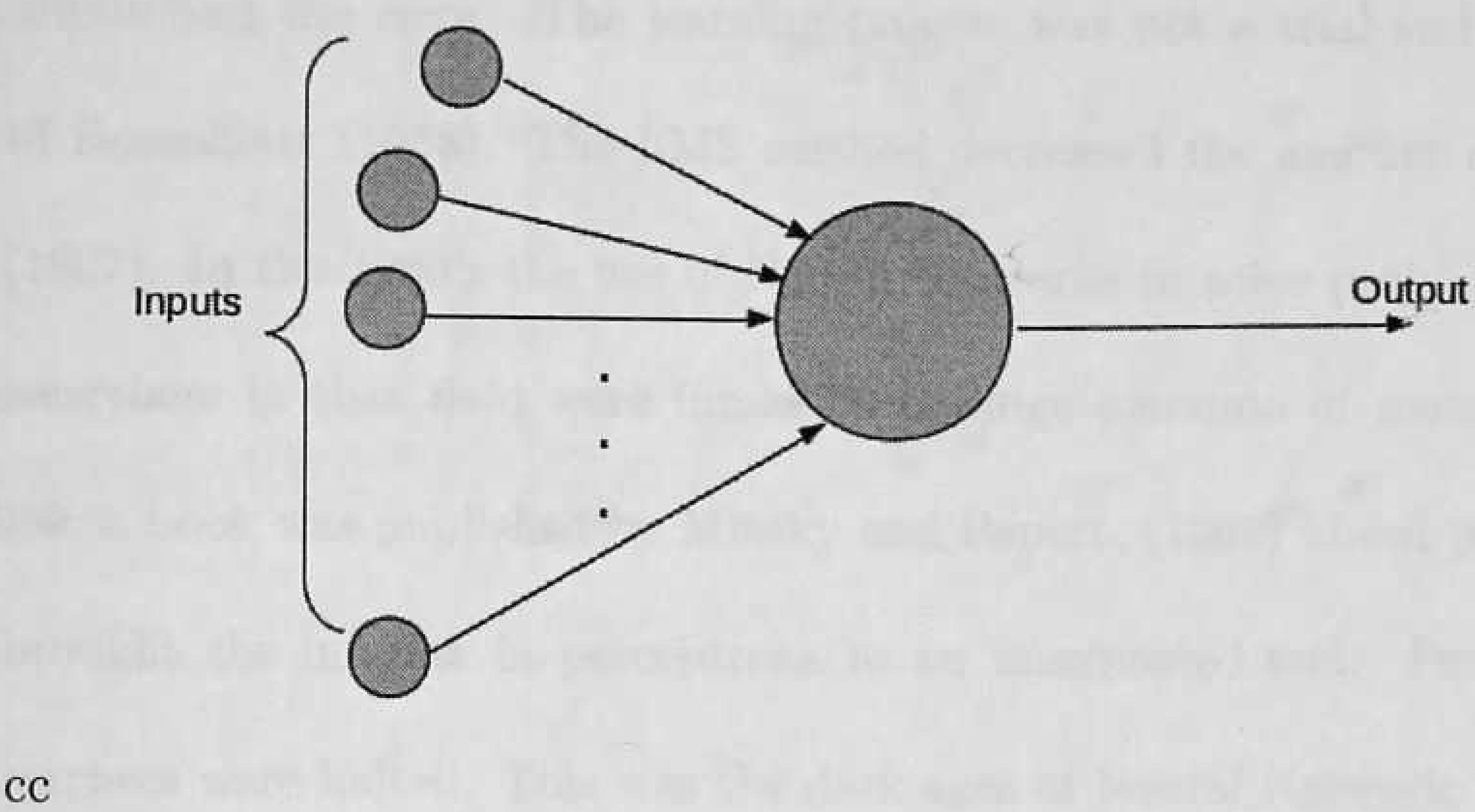

Fig. 3.2: Simple Neuron Model

The first ever mathematical model of Neural Network was developed by McCulloch and Pits (1943). The model had two nodes (neurons) at the input section and one node (neuron) at the output section. The weight of each node was the same and the output was set to binary. Hence the model outputs either 0 or 1 depending on the value of the summed inputs. One can say that the output of the network remains 0 until a certain threshold value of the network is attained. After McCulloch and Pitts (1943) had set the pace, many researchers took interest in the field of Neural Network. Rosenblatt (1958) developed the next Neural Network which he called the perceptron. Rosenblatt (1958), who was a physiologist, randomly interconnected the perceptrons and used trial and error to randomly change the weights in order to achieve "learning." Ironically, McCulloch and Pitts' model of Neural Network is a much better model for the process that goes on inside the neuron than the perceptron, which is the basis for the modern day field of Neural Networks (Anderson and Rosenfeld, 1987). Widrow and Hoff (1960) developed a mathematical method for adapting the weights. Assuming that a desired response existed, a gradient search method was implemented, which was based on minimizing the error squared. This algorithm later become known as LMS, or Least Mean Squares. LMS, and its variations, has been used extensively in a variety of applications, especially in the last few years. This gradient search method provided a mathematical method for finding an answer that minimized the error. The learning process was not a trial-and-error process as was in that of Rosenblatt (1958). The LMS method decreased the amount of computational time Smith (1997). In the 1960's the use of Neural Networks to solve problems became more popular. Researchers in that field were funded with huge amounts of money until the late 1960's. In 1969, a book was published by Minsky and Papert, (1969) about perceptrons. This publication brought the interest in perceptrons to an unexpected end. Funding for Neural Network researchers were halted. This was the dark ages of Neural Network. The publication was able to establish that perceptrons could only be used to solve problems which are linear in nature. Minsky and Papert (1969) concluded that the mere fact that perceptrons cannot solve non-linear problems, its credibility when it comes to non-linear situations was doubted. In the early of the 1970's the interest for Neural Network started developing. This was due to the fact that a new approach or architecture for implementing Neural Network was discovered. This approach or algorithm was known as the back propagation method and it was developed by Werbos (1974). The back propagation was then independently rediscovered by Parker (1985) and by Rumelhart and McClelland (1986), simultaneously (Smith, 1997). The central idea behind this solution is that the errors for the units of the hidden layer are determined by back-propagating the errors of the units of the output layer. For this reason the method is often called the back-propagation learning rule.

As indicated in the figure above the inputs are represented by the $x_1, x_2, x_3, ..., x_n$ and their respective weights are $w_{k1}, w_{k2}, w_{k3}, ..., w_{kn}$. The zone labeled S_k is where the processing of the inputs occurs. With this simple model, S_k is given by the weighted sum of all the inputs. i.e. $S_k = \sum_{i=1}^{n} x_i w_{ki} + b_o$. Where b_o is a bias input which is given a value of 1. Its weight also has a value of unity. To get the final network output, the weighted sum, S_k is passed through an activation or threshold function. This activation of threshold function would then transform its input i.e S_k into the network output. There are several of these activation functions and the choice depends on the type of application of the network but the most flexible is the hyperbolic tangent function (tanh).

3.4 Activation Functions

Activation functions in Neural Network are some special functions that remove linearity from the network. In a typical ANN, the weighted sum, S_k of the inputs are computed then the result is send as input to the activation function. The activation function then refines or normalizes its inputs before sending it out of the network as output. Several activation functions are used in Neural Networks depending on the objective of the network. The activation functions that are widely used in Neural Networks are the sigmoid and the tan hyperbolic functions.The output of the sigmoid function ranges between [0,1] and it is represented as follows:

$$F(S_k) = 1/(1 + exp(-S_k))$$

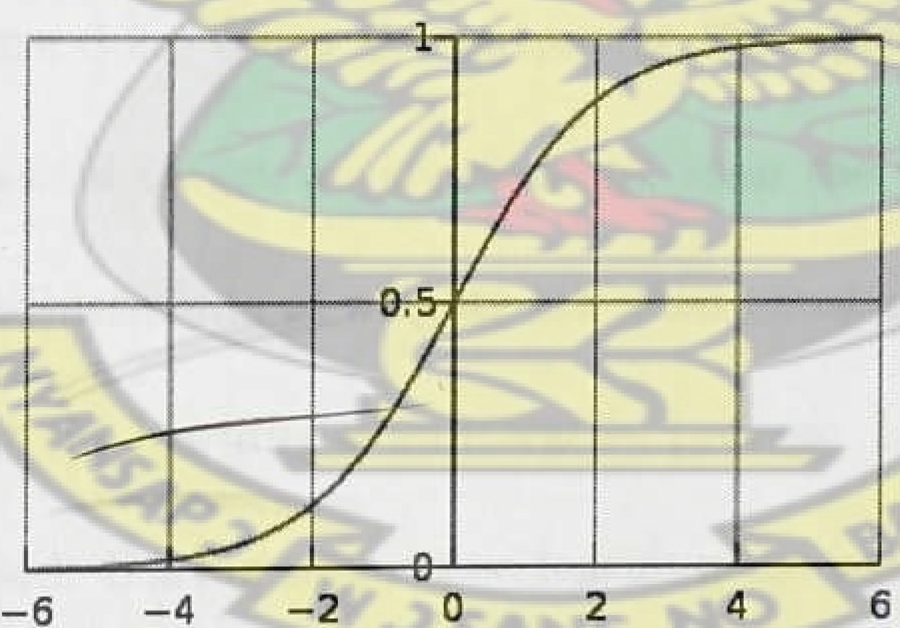

Fig. 3.3: Sigmoid Function

The output range of the tan hyperbolic function is between $[-1, 1]$ and it is also represented as follows:

$$F(S_k) = (1 - exp(-S_k))/(1 + exp(-S_k))$$

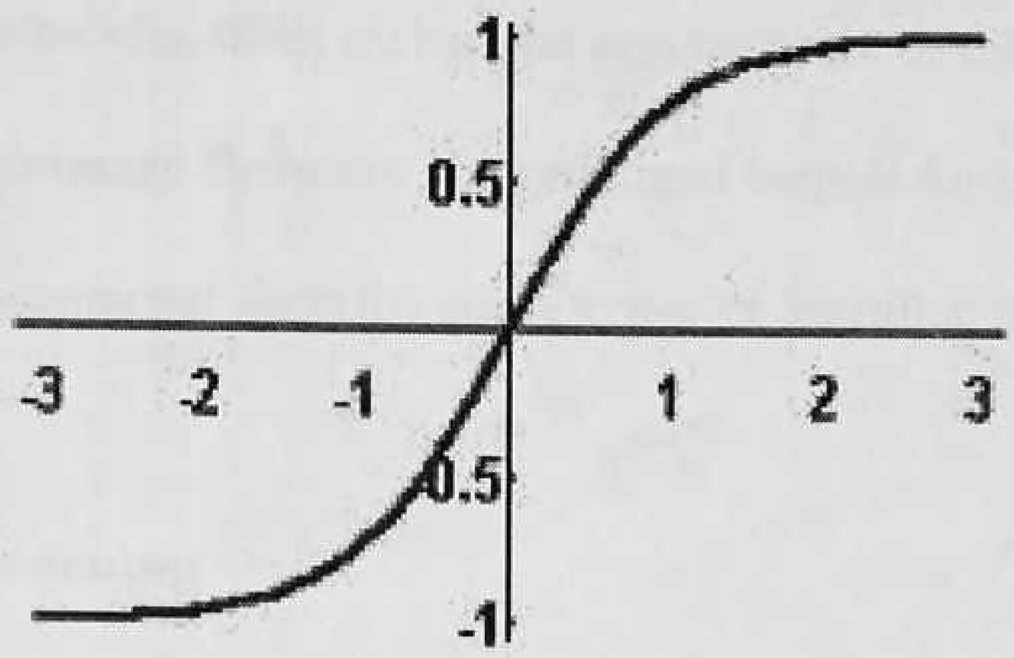

Fig. 3.4 Tanh Function

3.5 Learning Mechanisms

In section 3.2 we discussed how the processes that goes on in the human brain before it finally learns. We explained that before the brain can learn the dendrites have to adjust their strength until a desired response is reached for a signal to be sent. This same process is mimiced in the artificial neuron in figure 3.2. When the output in the simple neuron we created is not the same or closer to the actual target we are desiring, then the weights $w_{k1}, w_{k2}, w_{k3}, ..., w_{kn}$ would be adjusted. These adjustments of weights are what constitutes the learning process. It is sufficient that ~~we~~ say that ~~the learning~~ is an iterative process. Learning processes can either be supervised or unsupervised. During supervised learning the network is trained by providing it with an input and a target pairs of patterns. These input-target pairs can be provided by an external teacher, or by the system which contains the neural network. At every time step the network output is compare with the target provided by the external teacher. When the difference or error measured is negligible then learning stops. With the unsupervised learning or self-organization the network is trained to respond to clusters of pattern within the input. In this paradigm the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no priori set of categories into which the patterns are to be classified; rather the system must develop its own representation of the input stimuli. In summary there are no predefined targets for the unsupervised learning. Under these two broad groups we describe some types of learning mechanisms in the section that follows.

1. Error-Correction Learning

In the error-correction learning, the error between the target and the network output is computed. The target is what we want from the network and the network output is the output the neural network gives us (Jain et al, 1996). Suppose the target is d and the network output is y, the error, e between them is given by

$$e = d - y$$

.

Iteratively it can be represented as

$$e(n) = d(n) - y(n)\text{..........}(1.1)$$

where n is the discrete time-step. Ideally this error should be very small. We can now represent the instantaneous error as

$$Minimize\ E(n) = \tfrac{1}{2}\sum_k e_k^2(n)\text{.......................}(1.2)$$

where k is each output neuron. After the error has been computed the weight can be updated by

$$\Delta w_{ki}(n) = \eta * e_k(n) * x_i(n)$$

where x_i is the input vector and η is the learning rate which is chosen arbitrarily. The value of the new weights will then be computed by

$$w_{ki}(n+1) = w_{ki}(n) + \Delta\ w_{ki}(n)\text{...................}(1.3)$$

The error-correction learning fall under supervised learning.

2. Memory Based Learning (MBL)

Memory based learning is one type of learning that is associated with classification. In this type of learning the network is allowed to memorize the association between its input pattern and its corresponding output pattern. The input patterns and their associated outputs are stored in a large memory created in the network. When a new input pattern that the network has not seen before is sent to the network, the network would compare this new input pattern with the ones already stored in memory of the network. The network would then classify the new input pattern according to its closeness to the stored input patterns. For a simple demonstration, suppose x_i is an input pattern in a vector form and d_i which is a scalar is the corresponding output pattern. The association between the pair (x_i, d_i) are stored in the network. If a new input pattern say x_{test} is introduced to the network, the network automatically searches through the stored patterns and picks the one that is closest to the x_{test}. This is done by finding the smallest euclidean distance between the x_{test} pattern and the stored patterns. The x_{test} input pattern is then classified with that particular stored pattern which x_{test} is closer to. The MBL can be summarized based on the fact that if we can find x_{new} in $(x_1, x_2, x_3, ..., x_n)$ being the nearest neighbour of the x_{test}. This can be represented as:

$$\text{If } min\ d(x_i, x_{test}) = d(x_{new}, x_{test}) \text{ then } x_{test} \text{ is closer to } x_{new}.$$

The memory based learning is the principle guiding Self-Organizing Maps.

3. Hebbian Learning

The Hebbian Learning is basically associated with weight adjustments. It takes its basis from biological neurons in the sense that when a brain cell fires another brain cell the metabolic process should be such that the synaptic connection of the brain cells are strengthened (Jain et al, 1996). The mathematical model of Hebbian Modification or Learning or the Hebbian hypothesis can be expressed as

$$\Delta w_{ki}(n) = \eta * y_k(n) * x_i(n)$$

where η is the learning rate, $y_k(n)$ is the output and $x_i(n)$ the input.

From the above mathematical model, since at a point in time the input $x_i(n)$ will be constant, a plot of $\Delta w_{ki}(n)$ against $y_k(n)$ will produce a straight line with gradient $\eta * x_i(n)$.In this case as y(n) is increased, $\Delta w_{ki}(n)$ will also increase unbounded. But $\Delta w_{ki}(n)$ must have a limit. Because of this reason the covariance hypothesis is a better option. The covariance hypothesis is expressed mathematically as:

$$\Delta w_{ki}(n) = \eta * (x_i(n) - \bar{x}) * (y_k(n) - \bar{y})$$

where $\bar{x}$ and $\bar{y}$ are the time-averaged values of $x_i(n)$ and $y_k(n)$ respectively.

With this type of learning, the $\Delta w_{ki}(n)$ has a limit or a saturation point. It will not increase unbounded. A plot of $\Delta w_{ki}(n)$ against $y_k(n)$ would yield $\eta * (x_i(n) - \bar{x})$ as the slope and $-\eta * (x_i(n) - \bar{x}) * \bar{y}$ as intercept on the $\Delta w_{ki}(n)$ axis. This point on the $\Delta w_{ki}(n)$ axis is the max depression point. In summary the weight $w_{ki}(n)$ increases if $x_i(n) > \bar{x}$ and $y_k(n) > \bar{y}$ or $x_i(n) < \bar{x}$ and $y_k(n) < \bar{y}$. Also the weight $w_{ki}(n)$ decreases if either $x_i(n) > \bar{x}$ and $y_k(n) < \bar{y}$ or $x_i(n) < \bar{x}$ and $y_k(n) > \bar{y}$.

3.6 Components of Artificial Neural Networks (ANN)

An Artificial Neural Network is composed of many simple elements (units) interconnected with each other in parallel. The elements found in an ANN is similar to that of biological neurons. Actually a Neural Network System is a prototype of the human brain system. The human brain is made up of neurons that are connected to each other by string-like substances called synapses. In the quest of a neuron(sender) relaying a message or signal to another neuron(receiver), the message or signal is passed through the synapses from the sender to the target receiver. The strength of the signal or message sent depends on the synapses. The same way in an ANN, the units (neurons) involved are connected to each other by weights which relay signals from one unit to the other.

ANN's are trained to perform complex functions in various fields of application including pattern recognition, speech recognition, identification, forecasting, classification and control systems. In general Artificial Neural Networks are computational models that possesses a property of learning and adaptation. A typical Neural Network consists of a pool of simple processing units (neurons) which communicate by sending signals to each other over a massive number of weighted connections. The components of an Artificial Neural Network include:

- A set of processing units (neurons)
- Connections between the processing units by weights (synapses). This determines the strength of the signal a unit sends to another unit to which it is connected to.
- A propagation rule, which determines the input of a unit from its external output.
- An activation function which determines the output of the network.
- An external input or offset normally called the bias.

3.7 Artificial Neural Network Architectures

Every Neural Network has three basic types of layers. There is the layer (input layer) that receives input data from the exterior of the network. There is also the output layer which has a main duty of sending data out of the network and the hidden layer has its input and output remaining in the network. In fact there are a whole lot of network architectures that one can choose when applying ANN in any field of study. The most widely used in forecasting applications are the feedforward and recurrent networks. In the next section we will discuss into details these two types of network architectures.

3.8 Feedforward Networks

In a feedforward network information in the network flows in just one direction i.e. a forward direction with respect to the inputs. The flow occurs along one pathway from the input layer to the hidden layer then to the output layer. Clearly there is no feedback loop in this type of architecture (Girish, 2007). Figure 3.5 shows a three layer feedforward network with one hidden layer.

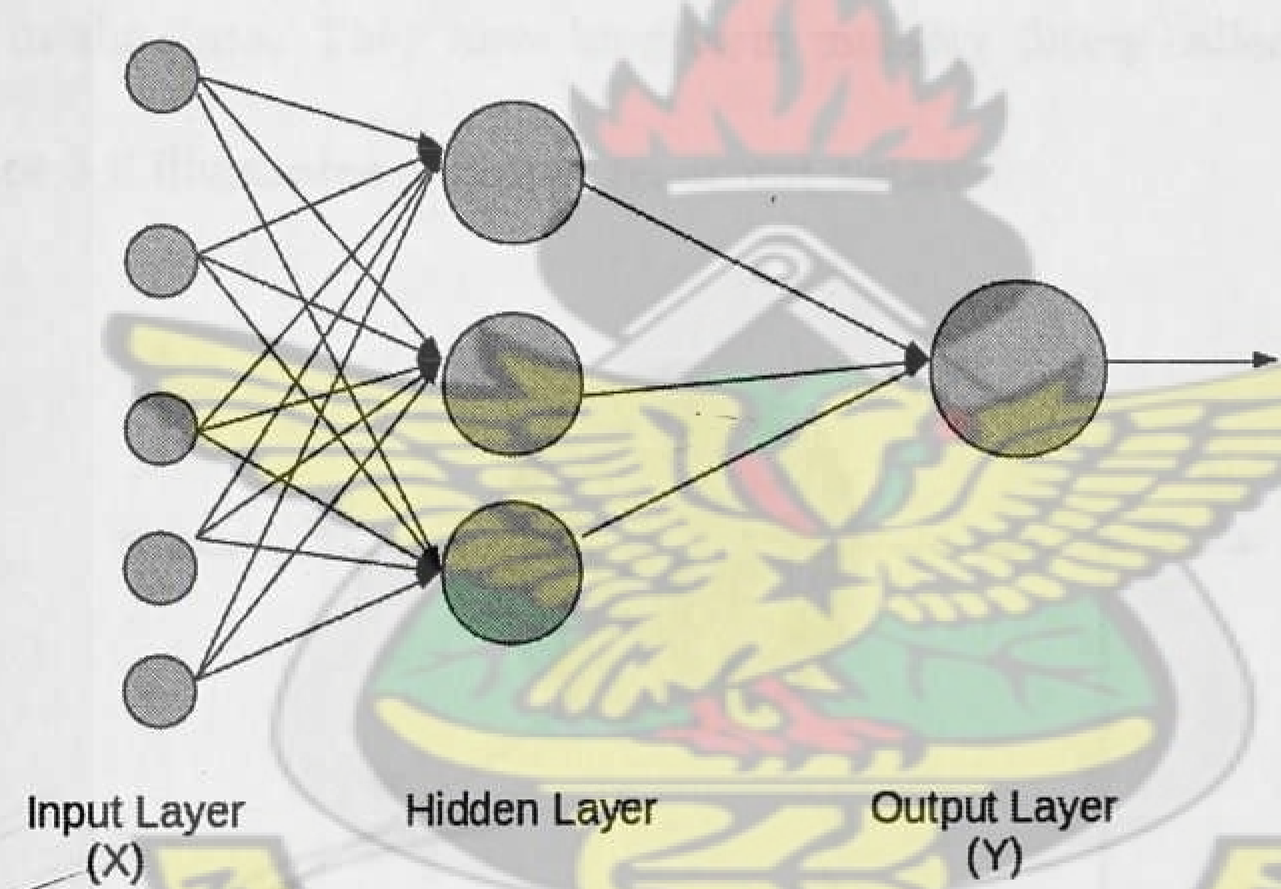

Fig. 3.5: A simple feedforward network

This feedforward network has an input layer with 5 input neurons represented by a vector X, a hidden layer with 3 neurons and an output layer with 1 neuron. The output of the network is represent by a scalar Y. In a typical forecasting problem, the vector X consists the independent variables and they a used to predict or forecast the dependent variable, Y. In feedforward ANNs, the information flow is purely one directional from the input layer to the hidden layer then to the output layer without any feedback from the any of the layers.

3.9 Recurrent Networks

These are special types of network architecture constructed from feedforward networks by the introduction a feedback loop in any of the layers. The feedback loop can be more than one but as we have already discussed the architecture has to fit the application. The feedback loops could connect the hidden layer back to the input layer, the output layer back to the hidden layer or even a layer connected back to itself. In fact we can have different types of feedback connections in a recurrent network. The feedback loops are placed in these type of network so that the past state of the network could be remembered and fed back into the network recursively. In this manner a memory trace of all previous inputs are presented to the network. Dynamically driven recurrent networks as sometimes called, tend to capture a long-term history of the series found in the data. They have long-term memory filters called context unit which does this job. Figure 3.6 illustrates a simple recurrent network.

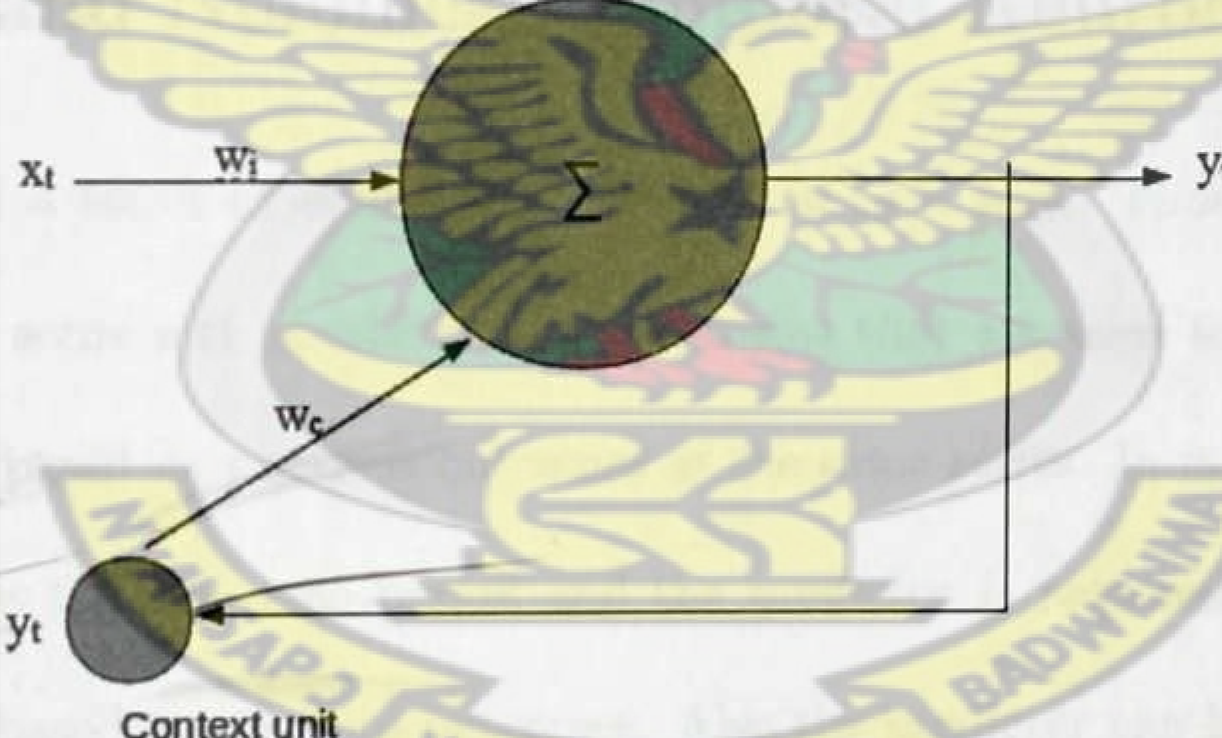

Fig. 3.6: A simple Recurrent network

At time step t the neuron receives an input, x_t and a delayed output of the neuron y_t then predicts the state of the system, y_{t+1} at the time step $t+1$. The w_i and w_c are the weights of the input and the context unit respectively. With a little demonstration let's see how the recurrent network stores past information in its memory. At time step t, the network output is

$$y_{t+1} = w_i x_t + w_c y_t \text{ .... } (1)$$

At time step $t+1$ we have the output to be

$$y_{t+2} = w_i x_{t+1} + w_c y_{t+1} \text{ ....} (2)$$

Now (2) can be written as

$$y_{t+2} = w_i x_{t+1} + w_c(w_i x_t + w_c y_t) \text{ ....} (3)$$

which will finally result in

$$y_{t+2} = (w_c)^2 y_t + w_c w_i x_t + w_i x_{t+1} \text{ ....} (3)$$

It can be seen clear from (3) that the network has been able to build in its structure the previous inputs x_t, x_{t+1} and the last output y_t.

3.10 Linear and Nonlinear Temporal Neuron Models

Time-series consists of a series of events that are related in a way. Times-series prediction or forecasting involves some sort of relational mechanisms that are used to fore tell the next outcome in that series based on previous outcomes of the same series. In most cases the time-series consists of events for a specific period and the predictor is required to tell the future outcome of that event based on previous outcomes. Also the predictor can be told to tell some of the events that can our within the time-series. The relationship found in any time-series are either considered to be linear or nonlinear. In the linear case the next outcome of the series is linear related to the previous and future outcomes. This also applies to the nonlinear case but here, the outcomes in the series are nonlinearly correlated or related. This correlation or relation between the series of outcomes is termed as autocorrelation. As the name suggests every time-series is time dependent. The number of time steps in the series that are significantly related in a time-series are called time lags or lags for simplicity. When the outcomes of an event in a time-series are used to predict future outcomes of that same event then we have what we call the temporal forecasting. In a case where an independent variable(s) is/are added to forecast the future outcomes of a particular event then we have the spatio-temporal forecasting. In the spatio-temporal model the independent variable(s) that are added are in a way related to that event outcomes that we want to predict. The model we shall be constructing in this thesis is that of a spatio-temporal model but first let us show the mathematical model of a linear temporal forecasting model and its neural network equivalent. In a simple linear time-series, the next observation is linearly related to the past observation(s) and is governed by the mathematical model

$$x_t = b_o + b_1 x_{t-1} + \epsilon_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(4)}$$

where x_t is the observation at time t which we want to forecast, x_{t-1} is the observation at time $t-1$ and b_o, b_1 are the model parameters which is estimated using the method of linear least squares (LSE). ϵ_t is the error term and it is assumed to be from an independent and a normal distribution with zero mean and a variance of σ^2. Extending equation (4) we get equation (5) which is the general form of an Autoregressive model which is the underlining equation for the formation of a time-series.

$$x_t = b_o + b_1 x_{t-1} + b_2 x_{t-2} + b_3 x_{t-3} + \ldots + b_p x_{t-p} + \epsilon_t \ldots\ldots\ldots \text{(5)}$$

where x_t is the observation at time t is to be predicted, x_{t-1}, x_{t-2}, ...,x_{t-p} are the observations at time $t-1$, $t-2$,...,$t-p$ and b_o, b_1,...,b_p are the model parameters which are estimated using the method of least squares (LSE). ϵ_t is the error term and it shows the deviation of the predicted value from the actual value. The p in the equation is the time lag. The neural model of the above linear temporal model is as shown in figure 3.7.

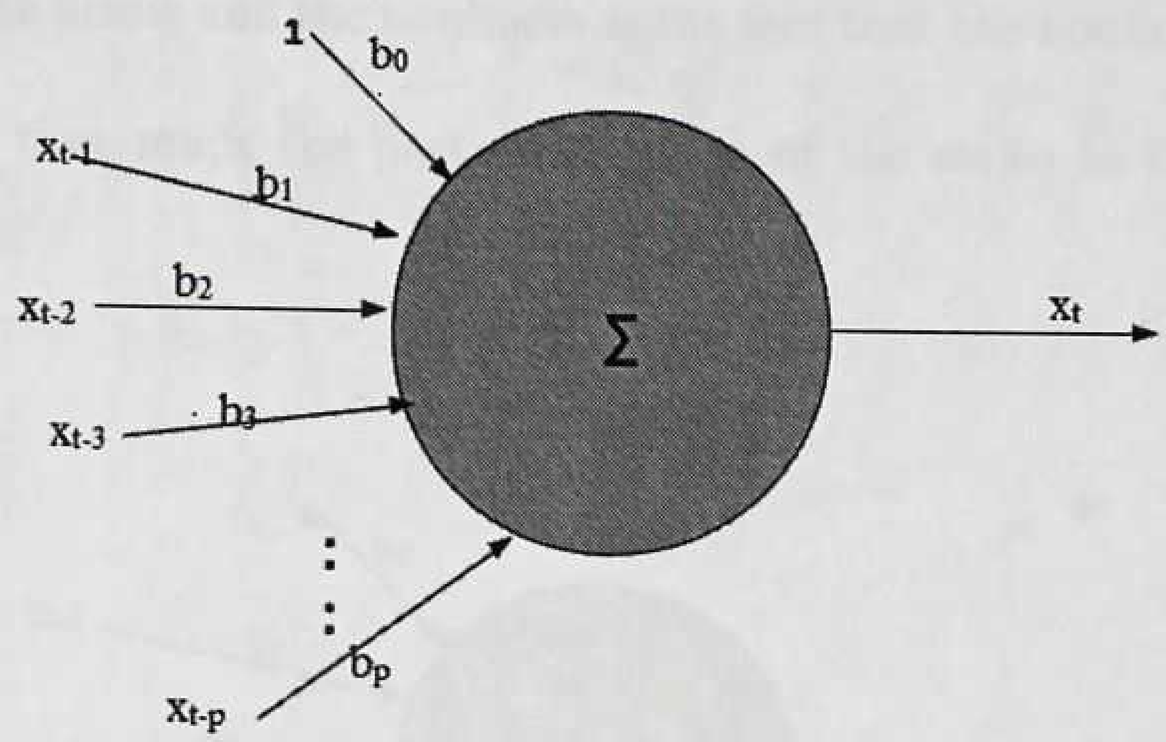

Fig. 3.7: A neural linear temporal autoregressive model

In the neural autoregressive model, the model parameters b_o, b_1,...,b_p representing the weights are estimated using the delta rule. This delta rule is a learning algorithm which would be discussed later in this chapter with other learning algorithms. The variant of the

linear temporal autoregressive model is its linear spatio-temporal autoregressive model which is most often referred as autoregressive model with exogenous inputs. The exogenous inputs or variables are the influencing independent factors which are combined with the used to predict the variable of interest.

As we have already discussed, linear models are best for time-series which have their observations being linearly related. The same scenario happens in a nonlinear time-series. Most the real world problems are nonlinear in nature and it suffices to build nonlinear models for this purpose and that is what Artificial Neural Networks is best at. The nonlinear model that governs any time-series is given by:

$$x_t = f(x_{t-1}, x_{t-2}, x_{t-3}, ..., x_{t-p}) + \epsilon_t ....(6)$$

where $f(x_{t-1}, x_{t-2}, x_{t-3}, ..., x_{t-p})$ is the nonlinear function that relates the past outcome of the time-series to its future outcome. ϵ_t is the error term and it is assumed to be from a normal distribution with $N(0,1)$. Figure 3.8 shows the autoregressive nonlinear temporal model. The only difference between the linear and the nonlinear is the fact that the nonlinear network model has a nonlinear function that maps the past observation of the series to its future predicted value.

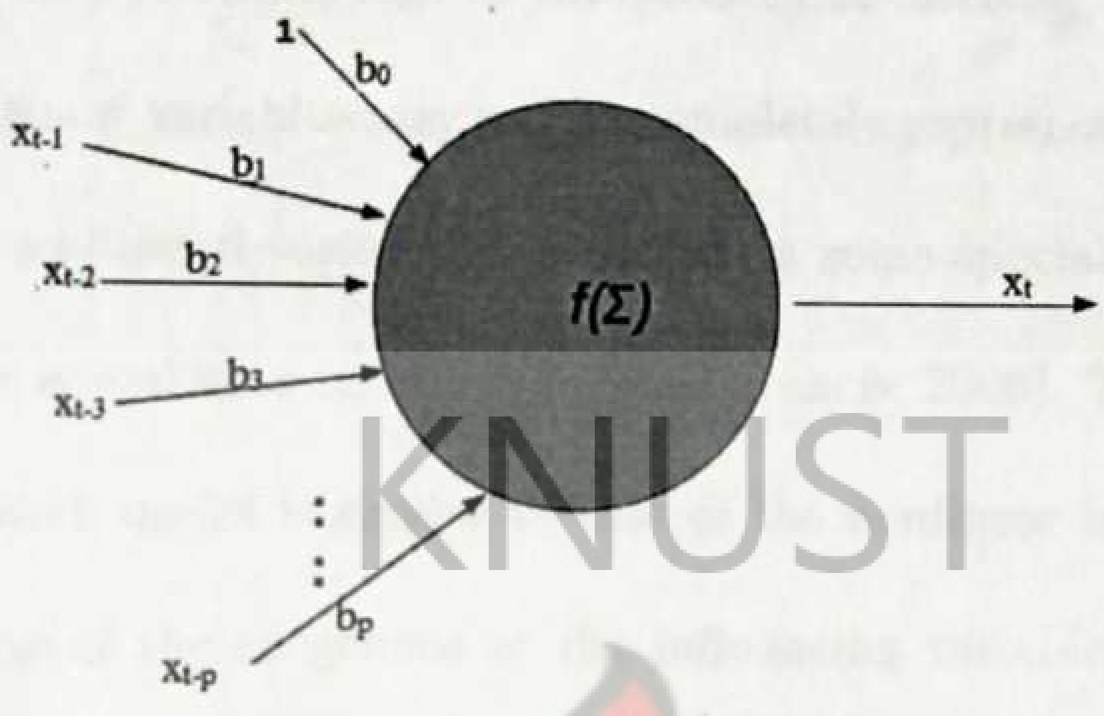

Fig. 3.8: A nonlinear temporal autoregressive model

These networks explained in this section are all temporal feedforward networks because the two models utilizes only the time lagged values of the dependent variable to predict of forecast the dependent variable. The network used in this study is shown in figure 3.9.

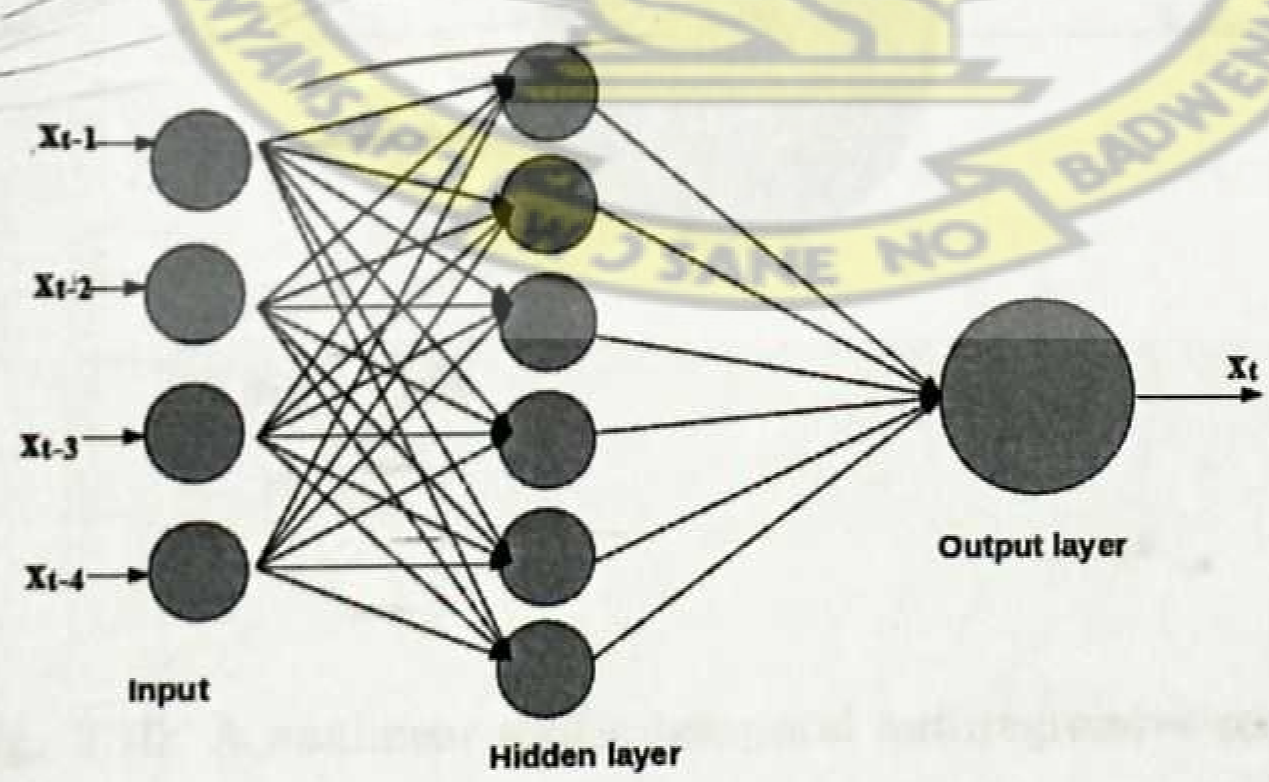

Fig. 3.9: A nonlinear temporal autoregressive model

3.11 Nonlinear Spatio-temporal Neuron Model

A spatio-temporal model incorporates not only time lags of the variable whose future observations are of interest but also other influencing variables, and possibly their time lags. This is because the model can extract information from the influencing variables that is not part of the variable of interest and the time lags of the variable of interest. In other words, the total influence of the significant variables may not be completely represented in the time-series observation and therefore additional variables are needed in some special case to capture the underlying dynamics of the actual time-series (Taylor and Francis, 2006). The nonlinear spatio-temporal feedforward network model is similar to that of the nonlinear temporal feedforward network but it has the lags of the exogenous or the influencing variables embedded in it as shown below in equation (7) and figure 3.9.

$$x_t = f(x_{t-1}, x_{t-2}, ..., x_{t-k}, y_{t-1} \cdot y_{t-2}, .., y_{t-p}) + \epsilon_t ....(7)$$

where $y_{t-1} \cdot y_{t-2}, .., y_{t-p}$ are the lags of the exogenous variable up to the *pth* lag.

Fig. 3.10: A nonlinear spatio-temporal autoregressive model

The exogenous variables can be as many as possible depending on the problem one wants to solve.

3.12 Recurrent Network with Hidden Neuron Feedback

As we have already discussed in section 3.7.2, recurrent networks have feedback loops and a context memory which are embedded into the network architecture to act as memory filters that would remember the past history of the network. In tis section we will discuss the two recurrent neuron model with hidden feedback i.e. Jordan and Elman networks. In the Jordan Network the output of the output layer is fed back into the hidden layer at the next time step. A pictorial view of this network is shown below in figure 4.0.

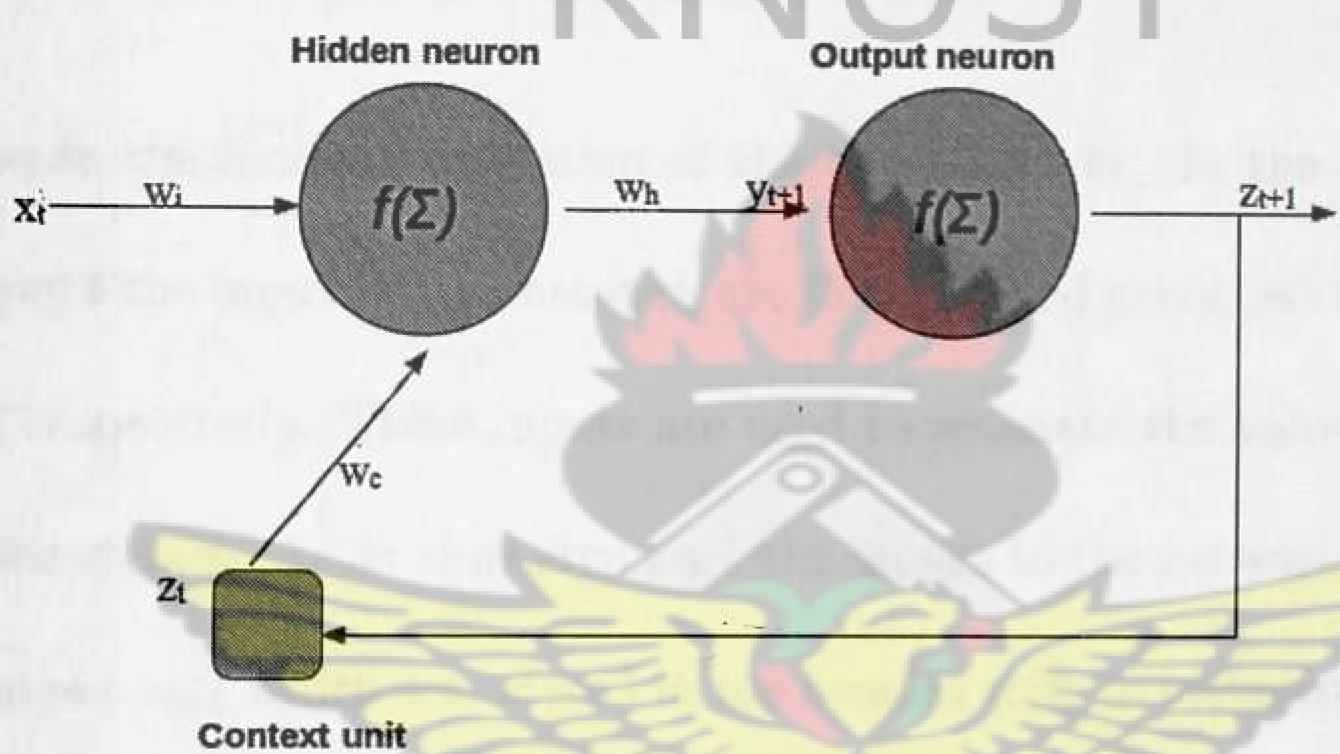

Fig. 3.11: Jordan Network

Another type of a recurrent network is the Elman network which feeds the output of its hidden layer back into itself at the next time step. This network is shown in figure 3.11.

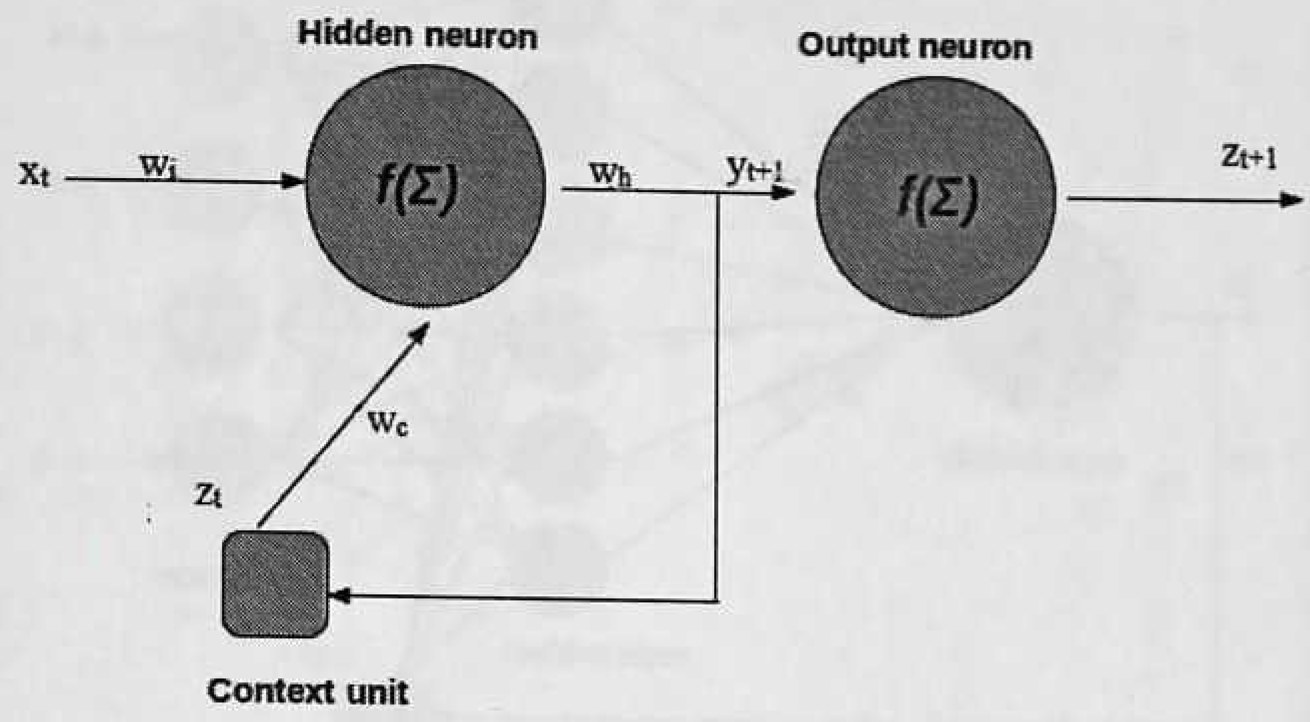

Fig. 3.12: Elman Network

Now let us discuss the mode of operation of the two networks. In the case of the Jordan network, at time step t the inputs to the network are the weighted averages of x_t and z_t with the weights w_i and w_c respectively. These inputs are used to estimate the value of the time-series z_{t+1} at the next time step, $t+1$. At time step $t+1$ the inputs to the network now becomes x_{t+1} and the delayed output z_{t+1} which was stored in the context unit to estimate the time-series at $t+2$. The same process occurs in the Elman network. The only difference is that the hidden output y_{t+1} is fed back into the hidden layer at the next time-step. Figure 3.13 shows the recurrent network that was used in this study.

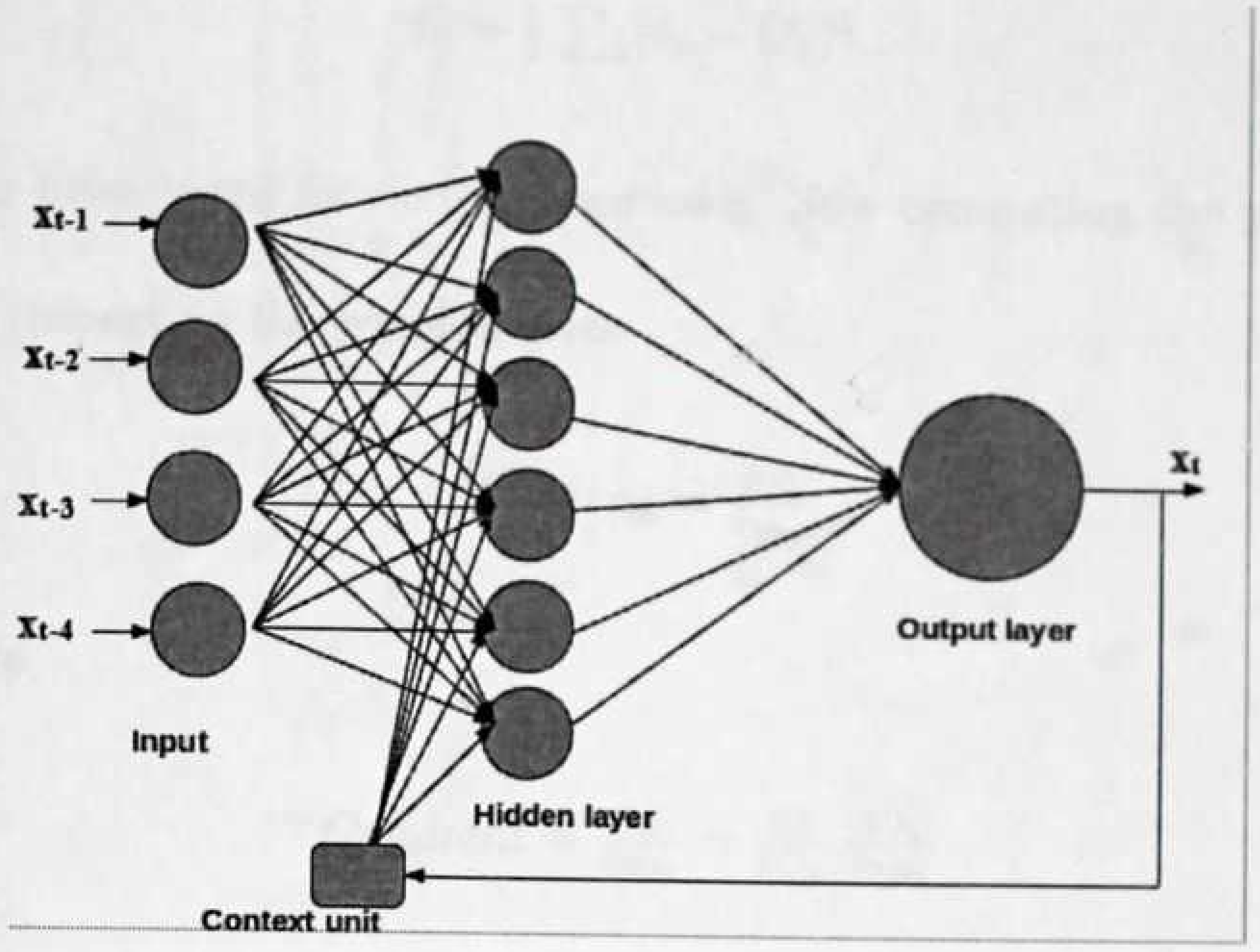

Fig. 3.13: A recurrent network model

The main process for every network is the training. Training a network is basically adjusting the weights at every time step as discussed already in this work. In this part of the thesis we would take a look at the training algorithms that can be used in Artificial Neural Networks. The two most common training algorithms which would be studied in this section are the backpropagation and gradient descent algorithms.

3.13 The Gradient Descent Algorithm

The gradient descent method is a ~~general~~ minimization method which updates parameter values in the direction opposite to the gradient of the objective function. In our case the updates or corrections we would be applying to the parameters (weights) would be in the opposite direction of the gradient of the error. Consider a linear network with x_i input patterns and the network output O_k, the error at any output neuron at any given time step is computed as

$$E_k = t_k - O_k$$

where O_k is the target output. The sum of error squared for all the input patterns can be computed as

$$E = \frac{1}{2}\sum_k (t_k - O_k)^2$$

The constant $\frac{1}{2}$ was introduced for convenience sake. Now computing the gradient of the sum squared error with respect to the weights gives

$$Gradient = \frac{\partial E}{\partial w_{ki}}$$

Using the chain rule

$$Gradient = \frac{\partial E}{\partial w_{ki}} = \frac{\partial E}{\partial O_k} . \frac{\partial O_k}{\partial w_{ki}}$$

$$\frac{\partial E}{\partial y_k} = -(t_k - O_k)$$

$$y_k = \sum_i w_{ki} x_i$$

which yields

$$\frac{\partial O_k}{\partial w_{ki}} = x_i$$

finally

$$\frac{\partial E}{\partial w_{ki}} = -(t_k - O_k)x_i$$

this can be rewritten as

$$-\frac{\partial E}{\partial w_{ki}} = (t_k - O_k)x_i$$

It can be seen from this equation that the correction we have to apply is in the opposite direction of the gradient with respect to the weights. This is what is termed as the gradient descent. Now change in weight is

$$\Delta w_{ki} = -\eta(t_k - O_k)x_i$$

where η is the learning rate. In effect

$$w_{new} = w_{old} + \Delta w_{ki}$$

3.14 The Backpropagation Algorithm

Considering a single input neuron with input x, weight w, an activation function $\phi()$, a bias θ and an output neuron O as shown in figure 3.12, the output, O from the output the neuron is given by

$$O = \phi(xw + \theta)$$

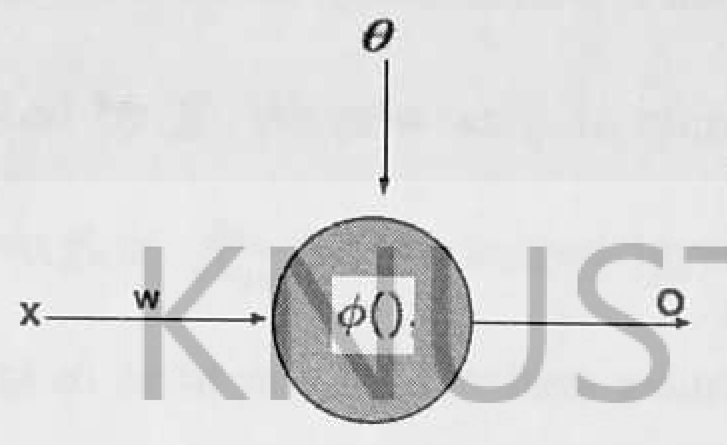

Fig. 3.12: Single input neuron

Now consider a network with input layer I, a hidden layer J and an output layer K we can define the following notations which would be inhibited by the network for carrying out the backpropagation algorithm.

- x_j : input to node j of the hidden layer
- w_{ij} : weight from node i to node j of the hidden layer
- x_k : input to node k of the output layer
- w_{jk} : weight from node j to node k of the output layer
- $\phi(x) = \frac{1}{1+e^{-x}}$: the sigmoid activation or transfer function
- θ : the bias in any layer l
- O_k : the output of node k in output layer and
- T_k : the target value of node k of the output layer

With these notations defined we can now generate the backpropagation algorithm. When an input signal is sent to the network, its corresponding output is generated through the output neuron or layer (Chakraborty,2010). The squared error (SE) produced between the target and network output is given by

$$E = \frac{1}{2}\sum_k (T_k - O_k)^2$$

The $\frac{1}{2}$ multiplying the error is introduced for convenience. This is the error of the network for a single iteration and it is denoted by E. We now want to compute the rate of change of the error with respect to the given weights, $\frac{\partial E}{\partial w_{jk}}$ for the output layer and $\frac{\partial E}{\partial w_{ij}}$ for the hidden layer. We compute these error gradients so as to minimized them eventually. Since there is no output from the input layer I we will only consider the nodes from the hidden layer J and the output layer K. Starting with the output layer weights, the rate of change of the error, E, is

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial}{\partial w_{jk}} \frac{1}{2} \sum_k (T_k - O_k)^2 \tag{3.1}$$

$$= (T_k - O_k) \frac{\partial}{\partial w_{jk}} O_k \tag{3.2}$$

but $O_k = \phi(x_k)$, the transfer function evaluated at the output node k

$$= (T_k - O_k)\phi(x_k)(1 - \phi(x_k)) \frac{\partial}{\partial w_{jk}} x_k \tag{3.3}$$

but x_k is the output of the jth node multiplied by the weight connecting that node to the kth output node. So we can say that it is the output of the jth node, O_j. Finally we obtain the rate of change of the error with respect to the output node as

$$\frac{\partial E}{\partial w_{jk}} = (T_k - O_k)O_k(1 - O_k)O_j$$

Now defining

$$\delta_k = (T_k - O_k)O_k(1 - O_k)$$

So for an output layer node $k \varepsilon K$ the rate of change of the error is given by

$$\frac{\partial E}{\partial w_{jk}} = O_j \delta_k$$

In the hidden layer the rate of change of the error is

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \frac{1}{2} \sum_k (T_k - O_k)^2 \tag{3.4}$$

$$= \sum_k (T_k - O_k) \frac{\partial}{\partial w_{ij}} O_k \tag{3.5}$$

$$= \sum_k (T_k - O_k) \phi(x_k)(1 - \phi(x_k)) \frac{\partial}{\partial w_{ij}} x_k \tag{3.6}$$

$$= \sum_k (T_k - O_k) O_k (1 - O_k) \frac{\partial x_k}{\partial O_j} \cdot \frac{\partial O_j}{\partial w_{ij}} \tag{3.7}$$

$$NB : x_k = O_j w_{jk} \tag{3.8}$$

$$= \frac{\partial O_j}{\partial w_{ij}} \sum_k (T_k - O_k) O_k (1 - O_k) w_{jk} \tag{3.9}$$

$$= O_j(1 - O_j) \frac{\partial x_j}{\partial w_{ij}} \sum_k (T_k - O_k) O_k (1 - O_k) w_{jk} \tag{3.10}$$

$$= O_j(1 - O_j) O_i \sum_k \delta_k w_{jk} \tag{3.11}$$

Defining δ_j,

$$\delta_j = O_j(1 - O_j) \textstyle\sum_k \delta_k w_{jk}$$

the rate of change of error occurring at a input layer $j \varepsilon J$

$$\frac{\partial E}{\partial w_{ij}} = O_i \delta_j$$

where O_i is the output of node i in the input layer.

With all these definitions and notations the backpropagation algorithm can implemented in learning or training a network.

Summary of the Backpropagation Algorithm

1. Run the the network with the first set of input data set to obtain the network output.

2. For each output node compute

$$\delta_k = (T_k - O_k)O_k(1 - O_k)$$

3. For each hidden node compute

$$\delta_j = O_j(1 - O_j)\sum_k \delta_k w_{jk}$$

4. Update the weights as follows

$$\Delta w = \eta \delta_l O_{l-1}$$

where η is the learning rate, O_{l-1} is the output of the previous layer node and δ_l is the δ concerned with layer l.

5. Now

$$w_{new} = w_{old} + \Delta w$$

3.15 Steps in using Backpropagation Algorithm

In this section we will demonstrate how the backpropagation algorithm can be used to train a simple feedforward network by hand. The backpropagation algorithm can be summarized as follows:

Initialize the weights

For each training pattern

""

End

Until the error is acceptably low.

End

3.16 Example Training using Backpropagation Algorithm

In this section we would use the backpropagation algorithm to train a network designed to work like the Exclusive OR operator. The is shown in figure 3.13.

Input 1	Input 2	Target
0	0	0
0	1	1
1	0	1
1	1	0

Table 3.1 Exclusive OR

In this example we have four training sets consisting of two inputs and a corresponding target. The network used in this example has two input neurons in the input layer, two neurons in the hidden layer and one neuron in the output layer. The hidden neurons and the output neuron has a sigmoid transfer function as their activation scheme. The inputs are denoted by $x = [input1, input2]$. The weights connecting the input neurons to the hidden neurons (w_H) and those connecting the hidden neurons to the output neuron (w_O) are initialized as shown below.

$$w_H = \begin{pmatrix} -0.07 & 0.94 \\ 0.22 & 0.46 \end{pmatrix}; w_O = \begin{pmatrix} -0.22 & 0.58 \end{pmatrix}; b_H = \begin{pmatrix} -0.46 & 0.1 \end{pmatrix}; b_O = \begin{pmatrix} 0.78 \end{pmatrix}$$

where b_H and b_O are the initialized biases connected to the hidden and output layers respectively. The learning rate is also set to $\eta = 0.7$.

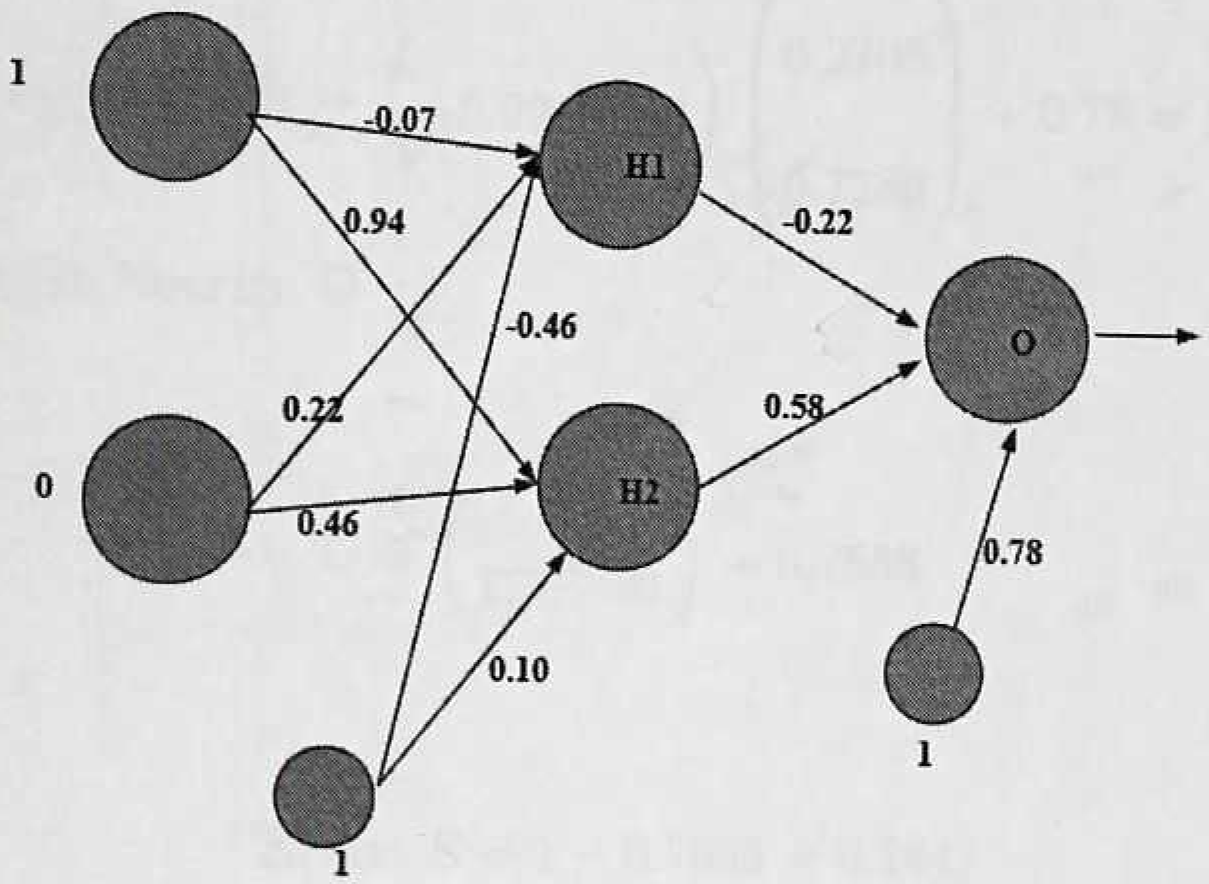

Fig. 3.13 Exclusive OR Operator

At time step=1 we compute the following.

Hidden Layer

Input into H1 and H2

$$\begin{pmatrix} H1 \\ H2 \end{pmatrix} = (w_H)'x' = \begin{pmatrix} -0.07 & 0.22 \\ 0.94 & 0.46 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -0.46 \\ 0.1 \end{pmatrix} = \begin{pmatrix} -0.53 \\ 1.04 \end{pmatrix}$$

Output from H1 and H2

$$\begin{pmatrix} H1 \\ H2 \end{pmatrix} = \begin{pmatrix} \frac{1}{1+e^{0.53}} \\ \frac{1}{1+e^{-1.04}} \end{pmatrix} = \begin{pmatrix} 0.3705 \\ 0.7389 \end{pmatrix}$$

Output layer

Input to Output Neuron, O

$$O = w_O(H1, H2)' = \begin{pmatrix} -0.22 & 0.58 \end{pmatrix} \begin{pmatrix} 0.3705 \\ 0.7389 \end{pmatrix} + 0.78 = 1.1270$$

Output from Output Neuron, O

$$O = \left(\frac{1}{1+e^{-1.1270}} \right) = 0.7553$$

Error

$$Error, E = 1 - 0.7553 = 0.2447$$

Weight adjustments

δ_k for the output neuron

$$\delta_{1O} = (T_1 - O_1)O_1(1 - O_1) = (1 - 0.7553)0.7553(1 - 0.7553) = 0.0452$$

δ_j for the hidden neurons

$$\delta_j = O_j(1 - O_j) \sum_k \delta_k w_{jk}$$

$$\delta_{H1} = O_{H1}(1 - O_{H1}) \sum_1 \delta_{1O} w_{H1O}$$

$$\delta_{H1} = 0.3705(0.0452)(-0.22) = -0.0037$$

$$\delta_{H2} = O_{H2}(1 - O_{H2}) \sum_1 \delta_{1O} w_{H2O}$$

$$\delta_{H2} = 0.7389(0.0452)(0.58) = 0.0194$$

Error Gradients

$$\frac{\partial E}{\partial w_{jk}} = O_j \delta_k$$

$$\frac{\partial E}{\partial w_{H1O}} = O_{H1} \delta_{1O}$$

$$\frac{\partial E}{\partial w_{H1O}} = 0.3705(0.0452) = 0.0167$$

$$\frac{\partial E}{\partial w_{H2O}} = O_{H2} \delta_{1O}$$

$$\frac{\partial E}{\partial w_{H2O}} = 0.7389(0.0452) = 0.0334$$

Now we compute Δ for the weights connecting the hidden layer to the output layer as

$$\Delta w = \eta \delta_l O_{l-1}$$

For H1

$$\Delta w = 0.7(0.0167) = 0.01169$$

$$w_{new} = \Delta w + w_{old}$$

$$w_{new} = 0.01169 - 0.22 = -0.20831$$

For H2

$$\Delta w = 0.7(0.0334) = 0.02338$$

$$w_{new} = \Delta w + w_{old}$$

$$w_{new} = 0.02338 + 0.58 = 0.60338$$

These same computations would be done for the weights connecting the inputs to the hidden layer and their respect bases. All these would be done during the first time step. At the second time step, the next set of training data is used with the new weights and bases that were computed in iteration one i.e. the first time step. This process is repeated until the network has learned adequately after which the network is used to predict or forecast new objects.

Chapter 4

Results and Discussion

4.1 Introduction

This chapter of the thesis concentrates on the how the data was acquired and prepared for training. The chapter discusses the network architectures that were created and the various activation functions and algorithms used in achieving the results. The results and discussions are also treated in this chapter.

4.2 Data Acquisition and Preparation

The main objective of this thesis is to investigate whether Artificial Neural Networks could be used to generate a forecasting model that would help in predicting tourist arrivals in Ghana. To investigate the power and ability of the so called Artificial Neural Network in this study, data were acquired from the World Tourism Organization(UNWTO) stationed at Spain through email correspondence. The dataset spanned from 1988 to 2009 and it consisted solely of the number of tourist arrivals. To obtain an efficient performance of the network the dataset(inputs and targets) must be preprocessed. Preprocessing of dataset before train help the network to be able to understand and use the dataset in a better way. This process in usually termed as normalization. Normalization in neural network is the process of transforming the dataset (inputs and targets) in way that the network would perform efficiently. The essence of normalization is that it removes or minimizes biasness in the dataset and also act as a catalyst in the training process. This becomes possible since all elements in the dataset are transformed to the same numerical range or scale. There are a number of normalization processes and techniques e.g. the Min-Max normalization technique. In this process the normalized data are rescaled to the range of values between 0 to 1 or -1 to 1. The technique has the ability of preserving the relationship that exist among the dataset after normalizing (Jayalakshmi and Santhakumaran, 2011). The min-max normalization technique is performed using the equation

$$\overline{x}_i = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} + 0.1$$

where $\overline{x}_i$ is the new data point after normalization. The normalization or standardization technique used in this work is the mean and standard deviation normalization process. In this procedure of normalization, the mean and standard deviation of the data set is computed and the normalized data points are obtained by using the following equation.

$$\overline{x}_i = \frac{(x_i - \mu)}{\sigma}$$

This process normalizes the dataset so that they will have zero mean and unity standard deviation and ~~it also~~ relationship preserving method i.e. the relationship between the actual dataset and the normalized dataset are the same. Figures 4.1a shows the actual arrival while 4.1b and 4.1c show the relationship that has been preserved using the Min-Max and Mean-Standard Deviation normalizing methods respectively.

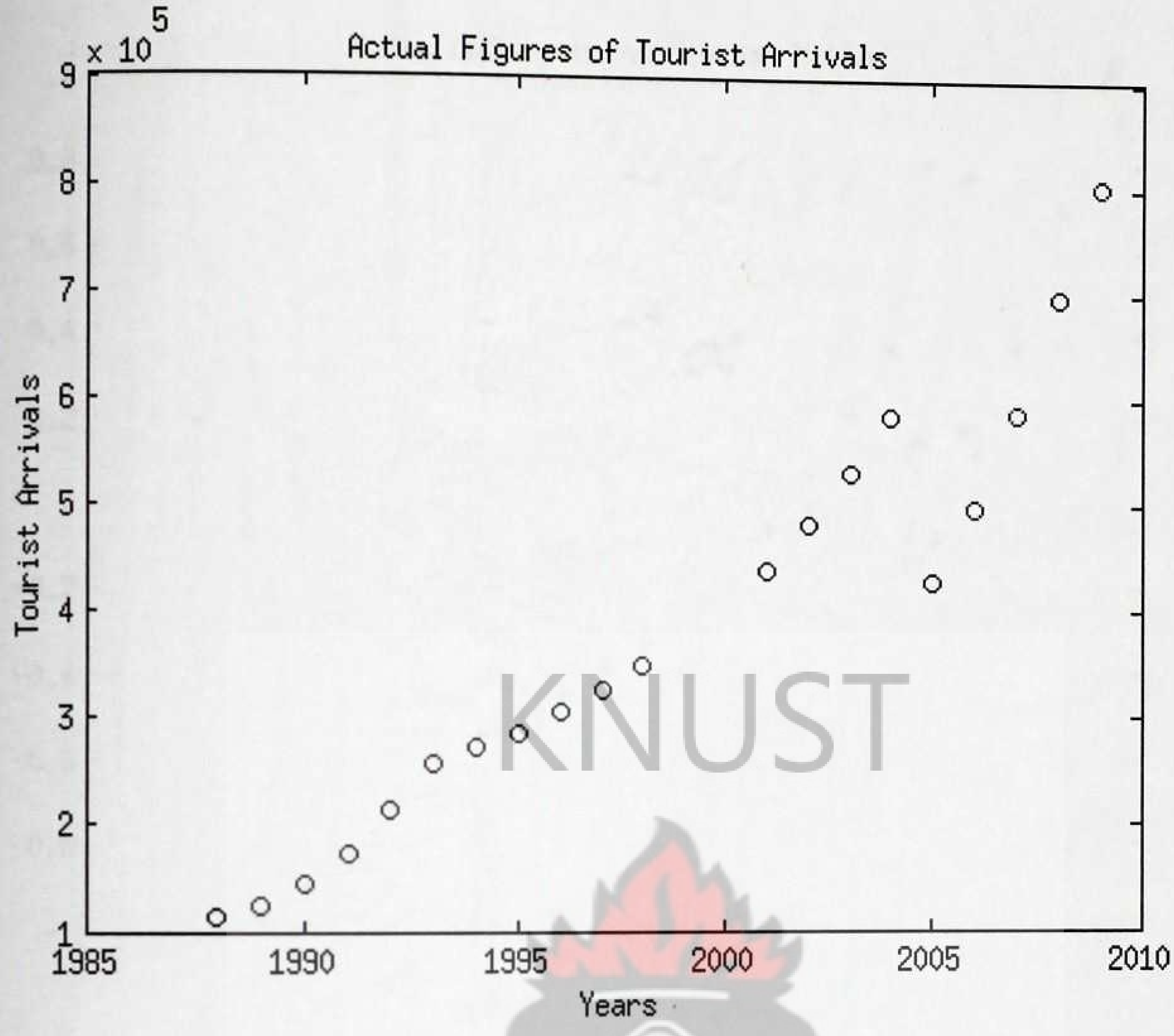

Fig. 4.1a Actual Tourist Arrival Values

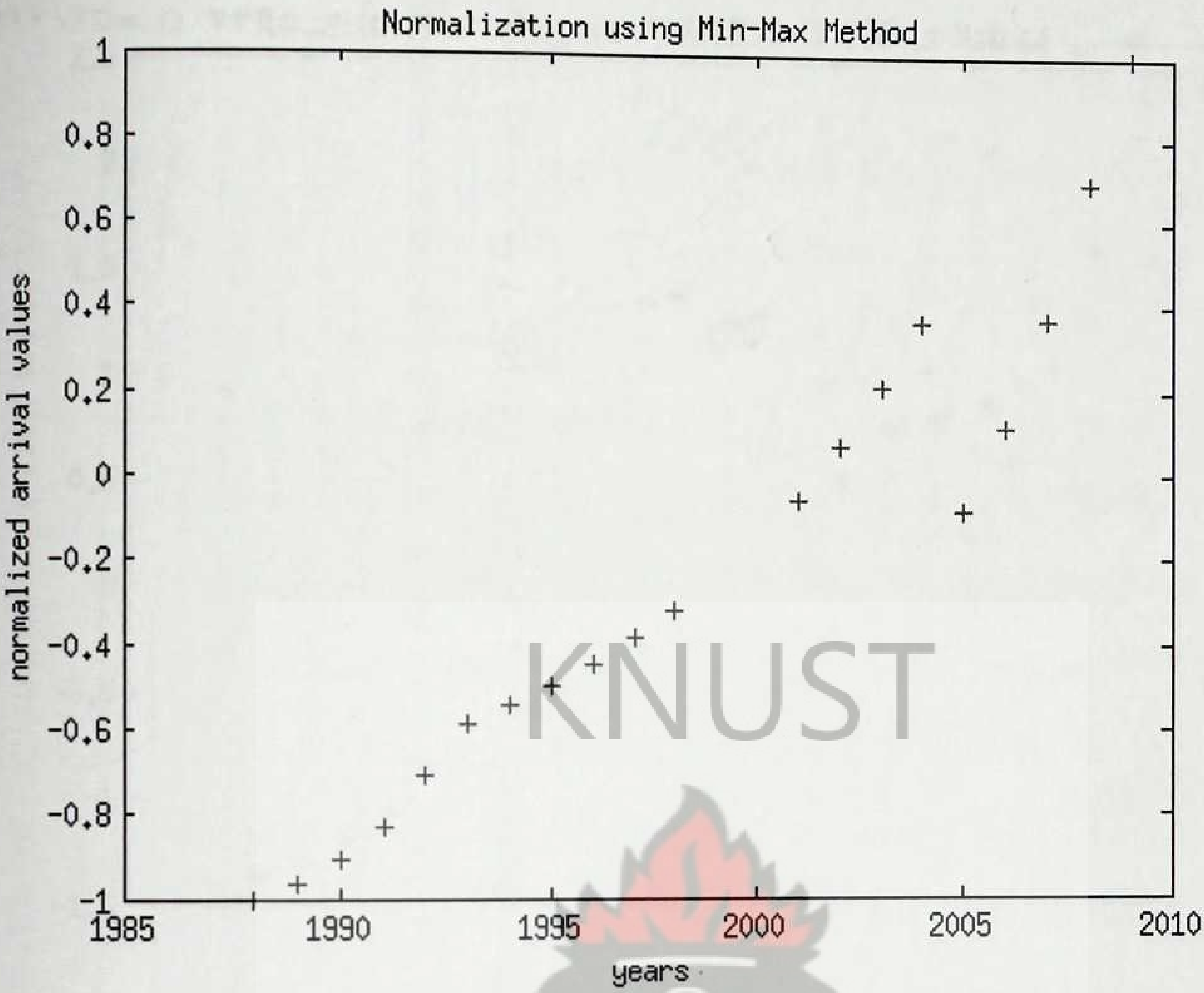

Fig. 4.1b Normalized Tourist Arrival Values

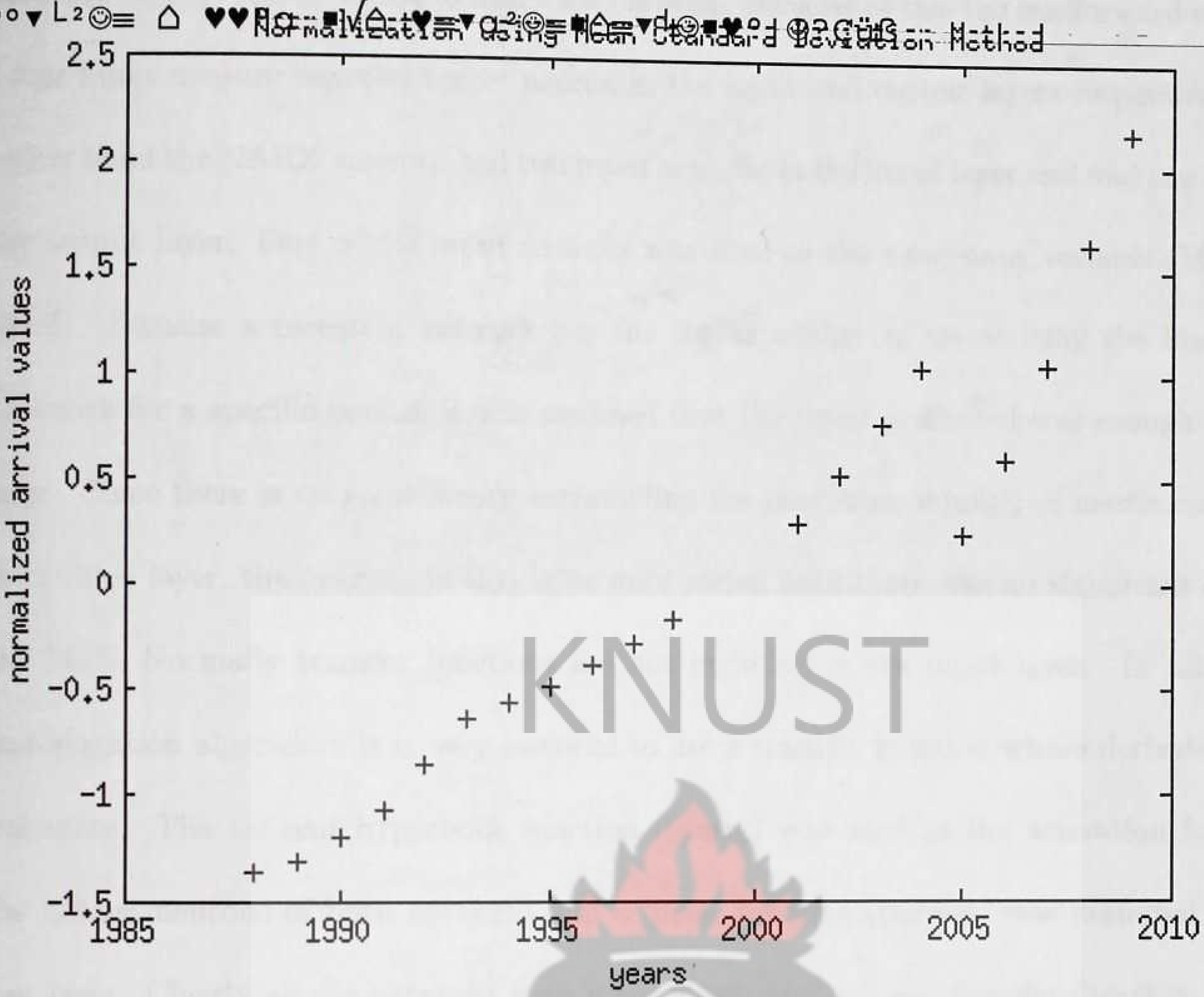

Fig. 4.1c Normalized Tourist Arrival Values

Because the data obtained were not enough, an input window consisting of four consecutive years from the normalized dataset were selected to form the inputs and the targets. This was to create enough data for the neural network for training. Actually the four year data window works in the sense that combining inputs from any four consecutive years would help predict the value of arrival in the following year.

4.3 Creating the Network Architectures

The network architecture that were used in the training were the feedforward and the NARX networks with the latter being a recurrent network. Both networks consist of three layers which include the input, hidden and output layers respectively. The four year data window that was created was used as inputs to the feedforward network. Because of this the feedforward network had four input neurons and one target neuron in the input and output layers respectively. On the other hand the NARX network had two input neurons in the input layer and and one neuron in the output layer. One of the input neurons was used as the exogenous variable (Maria et al, 2007). Because a recurrent network has the the capability of memorizing the history of the network for a specific period, it was assumed that the input presented was enough for the training. Since there is no good theory surrounding the maximum number of neurons allowed in the hidden layer, the neurons in this layer were varied until there was no significant change in the MSE. Normally transfer functions are not included in the input layer. In using the backpropagation algorithm, it is very essential to use a transfer function whose derivate could be calculate. The tangent hyperbolic function (tansig) was used as the activation function in the hidden neurons of both networks and a linear function (purelin) was preferred in the output layer. Clearly all the networks have a non linear and a linear transfer function. These network architectures allow the network to learn the nonlinear relationships that exist among the dataset. Normally a bias input is added to the input and its value is always set to 1. The weights of the input neurons are represented in the weight matrix below and are initialized to arbitrary values before training begins. The number of input weights depends on the number of hidden neurons.

$$w_{ij} = \begin{pmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1n} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{26} \\ w_{31} & w_{32} & w_{33} & \cdots & w_{36} \\ \vdots & \vdots & \vdots & & \vdots \\ w_{m1} & w_{m2} & w_{m3} & \cdots & w_{mn} \end{pmatrix}$$

4.4 Training of the Networks

After creating the networks, training can now start. The networks are trained with the back-propagation training algorithm. But before the training, the only initialization required is the input weights which is initialized to 0 in this experiment. Using the 2011 version of the Matrix Laboratory (MatLab) software, scripts and functions were written and executed to train the networks. The 20 data points where divided such the first 16 were used to for the training and the remaining 4 were used for the testing. The performance measure which is used to check the accuracy of the trained network is the Mean Squared Error (MSE) and the coefficient of correlation (R) (Diaconescu, 2008). The MSE and R are given as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(T_k - O_k)^2$$

and

$$R = \frac{\sum_{i=1}^{n}(x(t) - \overline{x})(x(t-k) - \overline{x})}{\sum_{i=1}^{n}(x(t) - \overline{x})^2}$$

where $x(t)$ is the data value, k is the lag and $\overline{x}=\frac{1}{n}\sum_{i=1}^{n} x(t)$

During training, when the number of weights (parameters) to be estimated are very few compared to the total number of training data then there would be no need to worry about over-fitting. Overfitting occurs when the network fails to generalize and adapt to the dataset. This means the network is able to train very well but fails to accurately predict a test data. To prevent such a situation there is the need to increase the training data set so that the network would have enough data for training and testing. In an event where the data set involved are very few, there are two basic methods that can be considered. The first technique is the regularization process (Backpropagation Algorithm Handbook, 2006). This involves the modification of the mean squared formula by adding a term that consists of the mean of the sum of squares of the network weights. In this way the network attains very small weights, and this forces the network response to be very smooth and less likely to over fit.The modified Mean Squared Error (MSE) formula becomes

$$MSE = \gamma MSE + (1-\gamma)MSW$$

where $MSW = \frac{1}{n}\sum_{j=1}^{n} w_j^2$ is the sum of squares of the network weights and γ is the performance ratio. The performance ratio is an arbitrary weight that is placed on the original MSE and the MSW. In our experiment its value was set to 0.5 in other to the MSE and MSW equal weights. The other technique is known as the early stopping method. In this technique the available data set is divided into three groups i.e the training set, validation set and the test set. As usual the training set is used to train the network so that all the weights would be adjusted and updated appropriately. The validation set is used to monitor the training process. During training, the error on the validation set is computed at each epoch and checked whether is has increased or decreased. In an event where there is a continuous decrease in the validation error, training continuous without any interruption. When the validation error increases for a set of consecutive epochs or iterations then training would stop and the values of the weights at the minimum validation error would be returned. During the training the neurons in the hidden layer were varied between 1 and 10 so as to determine the best combination that is perfect for the dataset.

4.5 Results and Discussion

This part of the thesis shows the various results that were obtain from the experiments that were performed. It also includes figures and tables for easy interpretation of the results. The training of both networks were done at a 0.5 learning rate. After the training though the data points were few, the two networks were able to learn and adapt to the patterns that existed among the data set. This can be seen from the various performance and regression plots that were obtained. Also the mean squared errors and the coefficient of correlations that were measured depicts this results. In general the recurrent network performed better than the feedforward network. This confirms the various results that have been achieved in literature in relation to recurrent networks and their applications to time series data. Figures 4.2 and 4.3 illustrates the performance plot for the feedforward and the recurrent networks respectively. In the feedforward network the training lasted for 19 epochs or iterations and the best validation error occurred at epoch 13. It can be seen clearly from the figure that after that point the validation error did not increase or decrease till the final epoch. The goal of the network was pegged at 10^-5 but the network was able to learn when the value reached 0.23457. That was why the training was stopped and the measure at that point was returned as 0.23457. From the same figure, it can also be seen that the network was able to generalize since the test line (red) follows the training line (blue) with very small deviations.

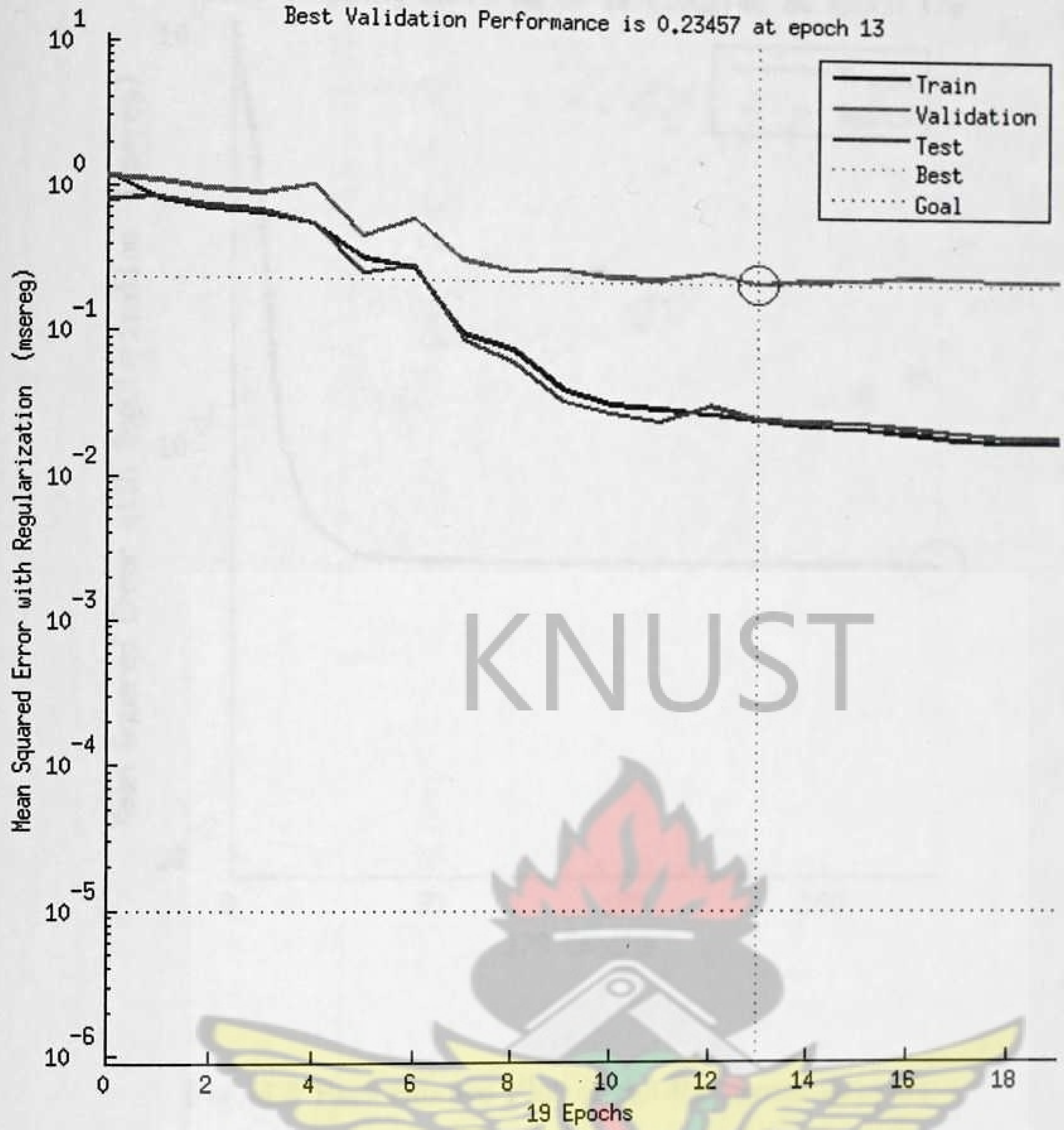

Fig. 4.2 Performance Plot of the Feedforward network

On the other hand the recurrent ~~network~~ trained up to epochs 178 and returned a performance of 0.052746. The mean squared error associated with the recurrent network is smaller than that of the feedforward. This can be attributed to the fact that the recurrent network stores in its memory the previous states of the network and hence predicts the future state of the network accurately.

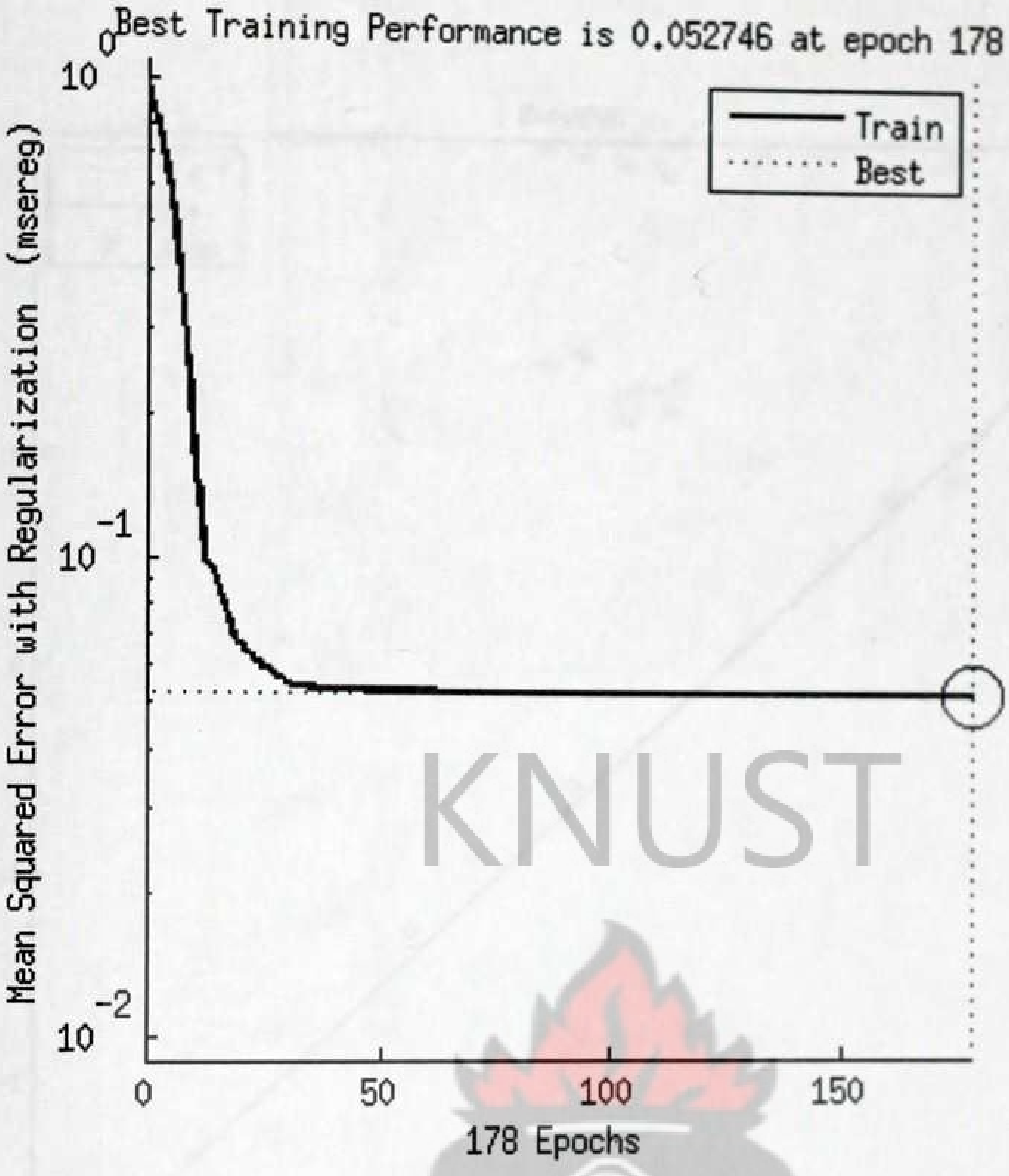

Fig. 4.3 Performance Plot of the Recurrent network

Using the test correlation of both networks, the recurrent network still does better than the feedforward network. The difference between the correlation coefficients is 0.00323 though not all that significant. This is shown in figures 4.4 and 4.5 below.

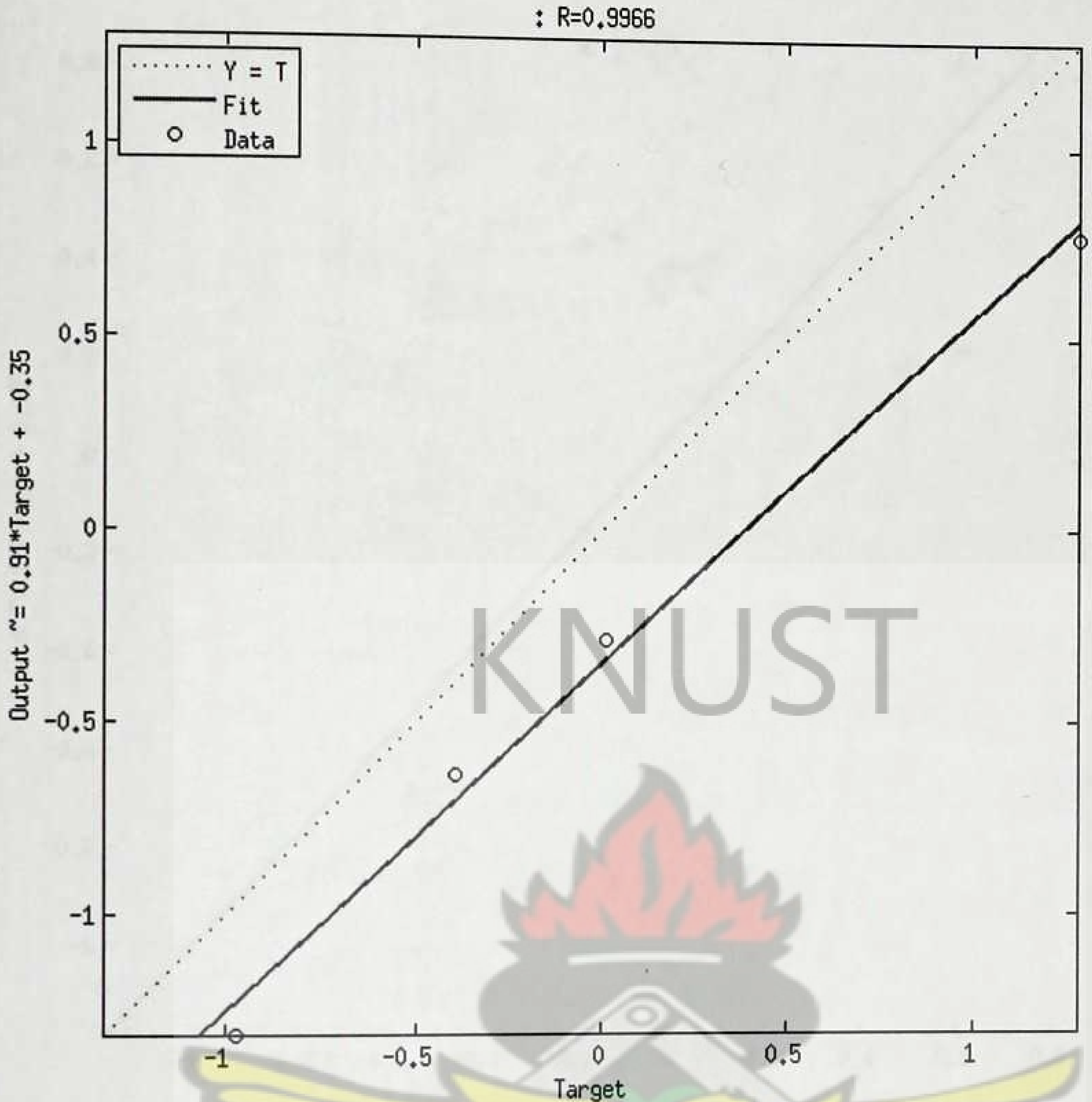

Fig. 4.4 Test for Recurrent Network

During the training process the ~~best regression~~ line relating targets to the network outputs was measured to find if there were a perfect fit i.e.outputs exactly equal to targets. If there were any perfect fit the correlation coefficient would be 1.

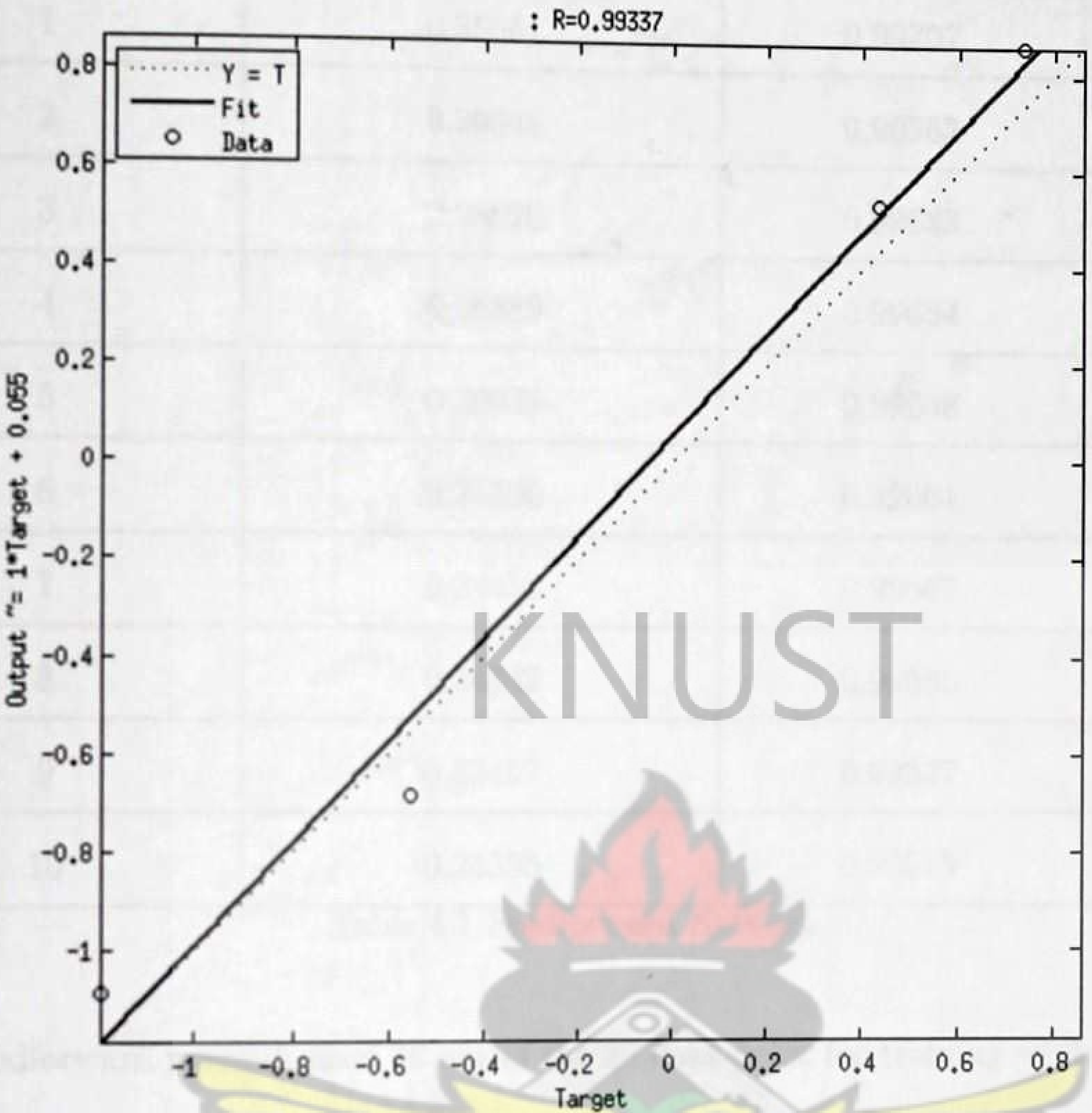

Fig. 4.5 Test for Feedforward Network

At the training sections discussed above the correlation coefficient for both the feedforward and recurrent networks are in figures 4.4 and 4.5 and in tables 4.1 and 4.2.

Hidden Layer Neuron	Mean Squared Error (MSE)	Test Correlation (R)	Time (sec)
1	0.35661	0.99707	1.065591
2	0.29948	0.99553	1.693822
3	0.28620	0.99683	1.755107
4	0.26889	0.99634	1.507852
5	0.26626	0.99658	1.492586
6	0.25396	0.99661	1.314332
7	0.24452	0.99667	1.624355
8	0.24342	0.99585	1.858545
9	0.23457	0.99337	1.797303
10	0.24356	0.99575	1.776525

Table 4.1 Feedforward Network

Also the feedforward network used 16 out of the 20 data point for training while the recurrent used 12. This suggests that the recurrent neural network did not train well but was able to predict well and vice versa in the case of the feedforward network. When the learning rate was increased during training, the training was not affected in any significant way. It is the performance ratio that rather affected the training significantly. Decreasing the performance ratio implies that less weight would be placed on the mean squared error (MSE) than the mean squared weights (MSW). A performance ratio of 0.05 was used to train both networks with 10 neurons each at the hidden layer. It was found that the performance of both networks increased showing a small mean squared error (MSE).

Hidden Layer Neuron	Mean Squared Error (MSE)	Test Correlation (R)	Time (sec)
1	0.149420	0.86424	1.684969
2	0.098796	0.87091	3.347047
3	0.079510	0.87430	3.060683
4	0.070121	0.87641	4.419511
5	0.064414	0.87642	4.373735
6	0.060569	0.87887	5.471413
7	0.057799	0.87963	5.933884
8	0.055705	0.99146	8.344863
9	0.054066	0.99244	5.770926
10	0.052746	0.99660	9.019175

Table 4.2 Recurrent Network (NARX)

This is summarized in table 4.3 and figures 4.6 and 4.7 below. From these results we can now build the complex nonlinear model that would be obtained from the training of the networks. Our purpose is to build a nonlinear model which uses the number of arrivals for four consecutive past years to predict the number of arrivals in the immediate future. The mathematical model for this can be expressed as

$$Q_t = w_{1j}Q_{t-1} + w_{2j}Q_{t-2} + w_{3j}Q_{t-3} + w_{4j}Q_{t-4} + \theta ....(8)$$

where w_{1j}, w_{2j}, w_{3j} and w_{4j} are the coefficients (weights) of the lagged number of tourist arrivals Q_{t-1}, Q_{t-2}, Q_{t-3} and Q_{t-4} respectively.

Equation 8 can be written as

$$Q_t = \sum_{i=1}^{4} w_{ij}Q_i + \theta ....(9)$$

where t = 5,6,... is the year of the prediction,j = 1,2,3,... depicting the number of hidden neurons used and θ is the bias.

Hidden Layer Neurons	Mean Squared Error (MSE)	Test Correlation	Network
10	0.24919	0.9952	Feedforward
10	0.07371	0.9999	Recurrent

Table 4.3 Performance of both networks at performance ratio 0.05

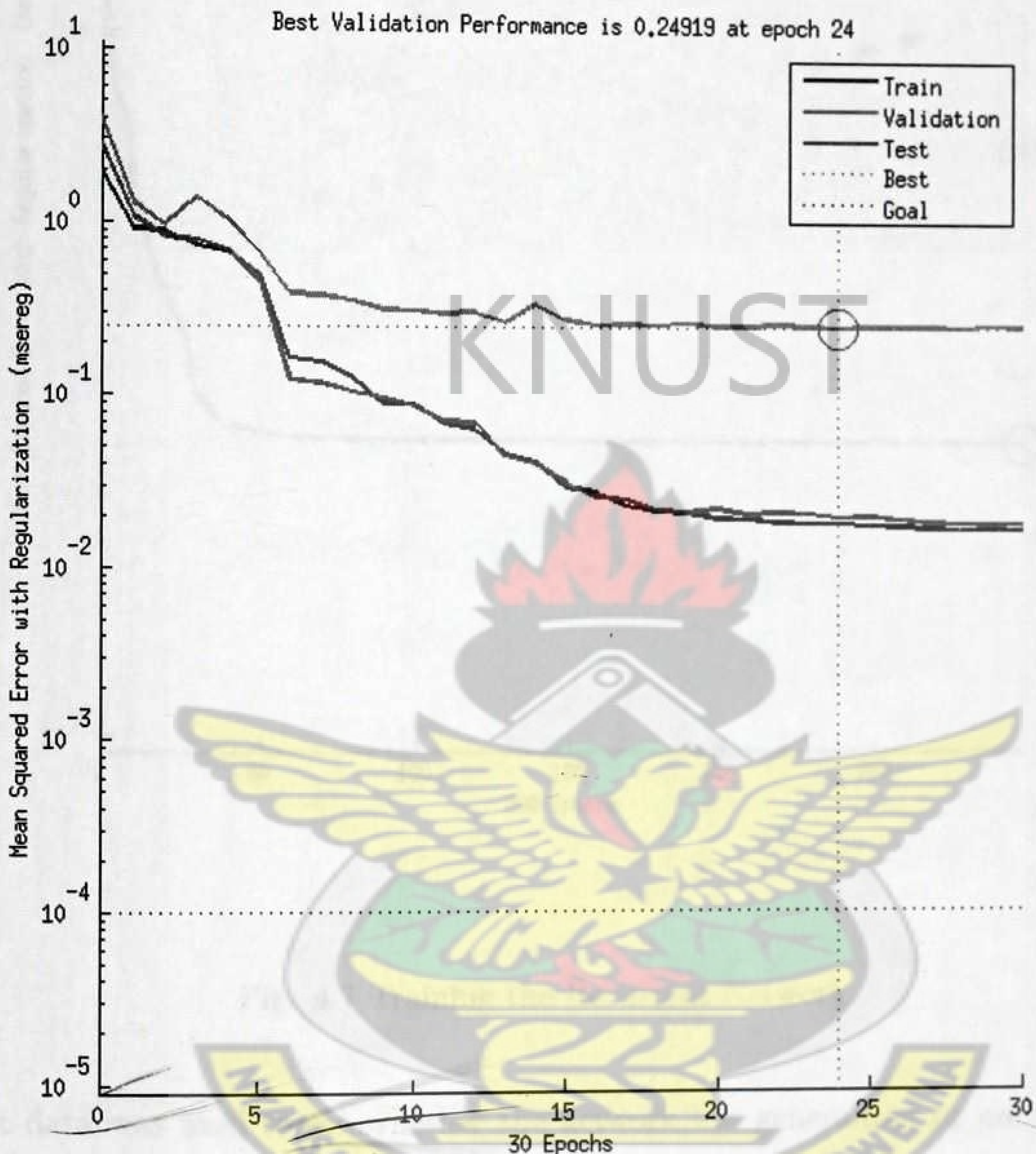

Fig. 4.6 Training the Feedforward Network

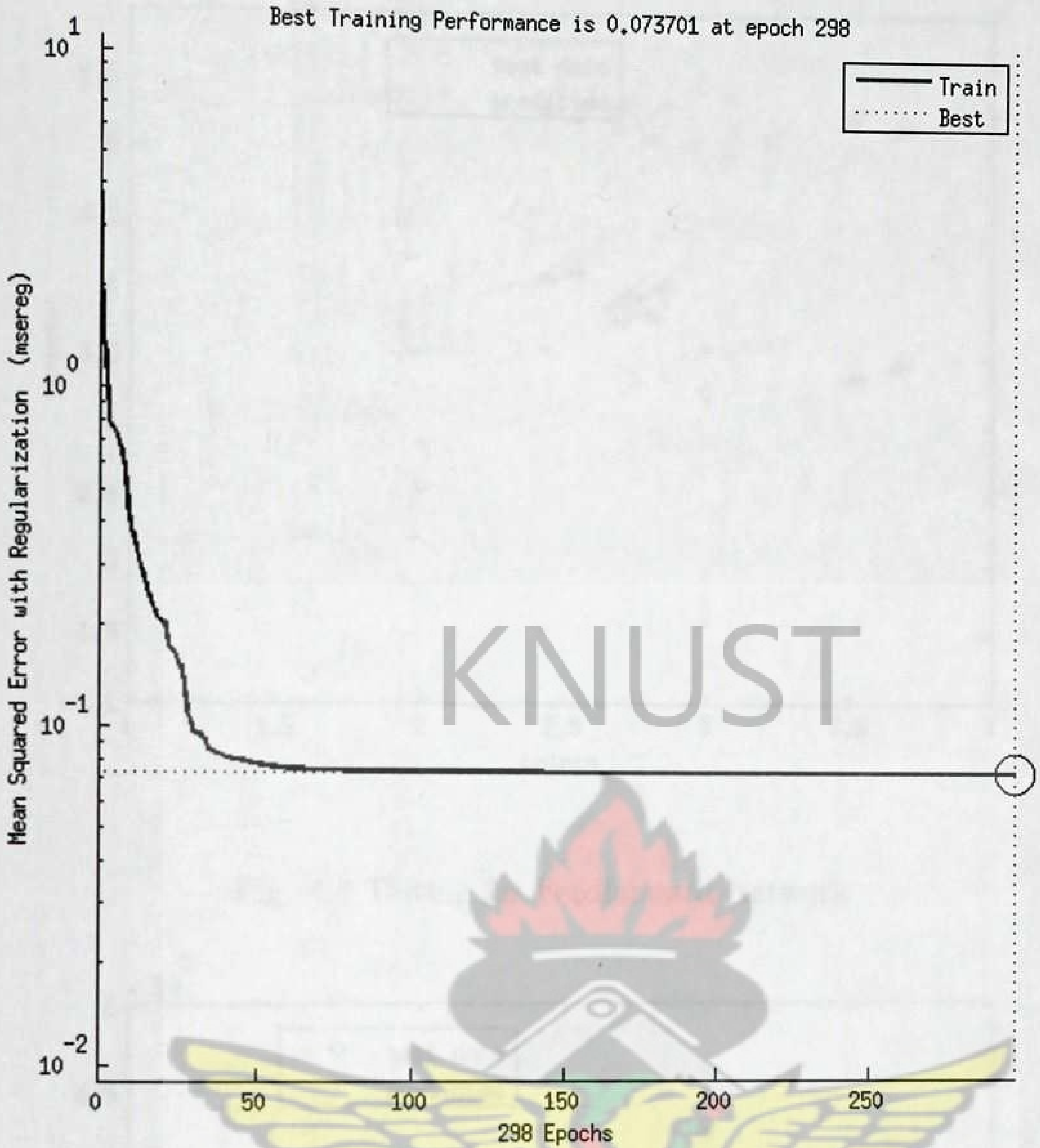

Fig. 4.7 Training the Recurrent Network

The test data was used to see whether the network has generalized or not. i.e. whether the network can accurately predict data that it has not seen before. It was discovered that the recurrent network had a greater generalization than the feedforward network. This can be seen in the figures 4.8 and 4.9 below. The recurrent network was able to predict the test data with minimal error than the feedforward network.

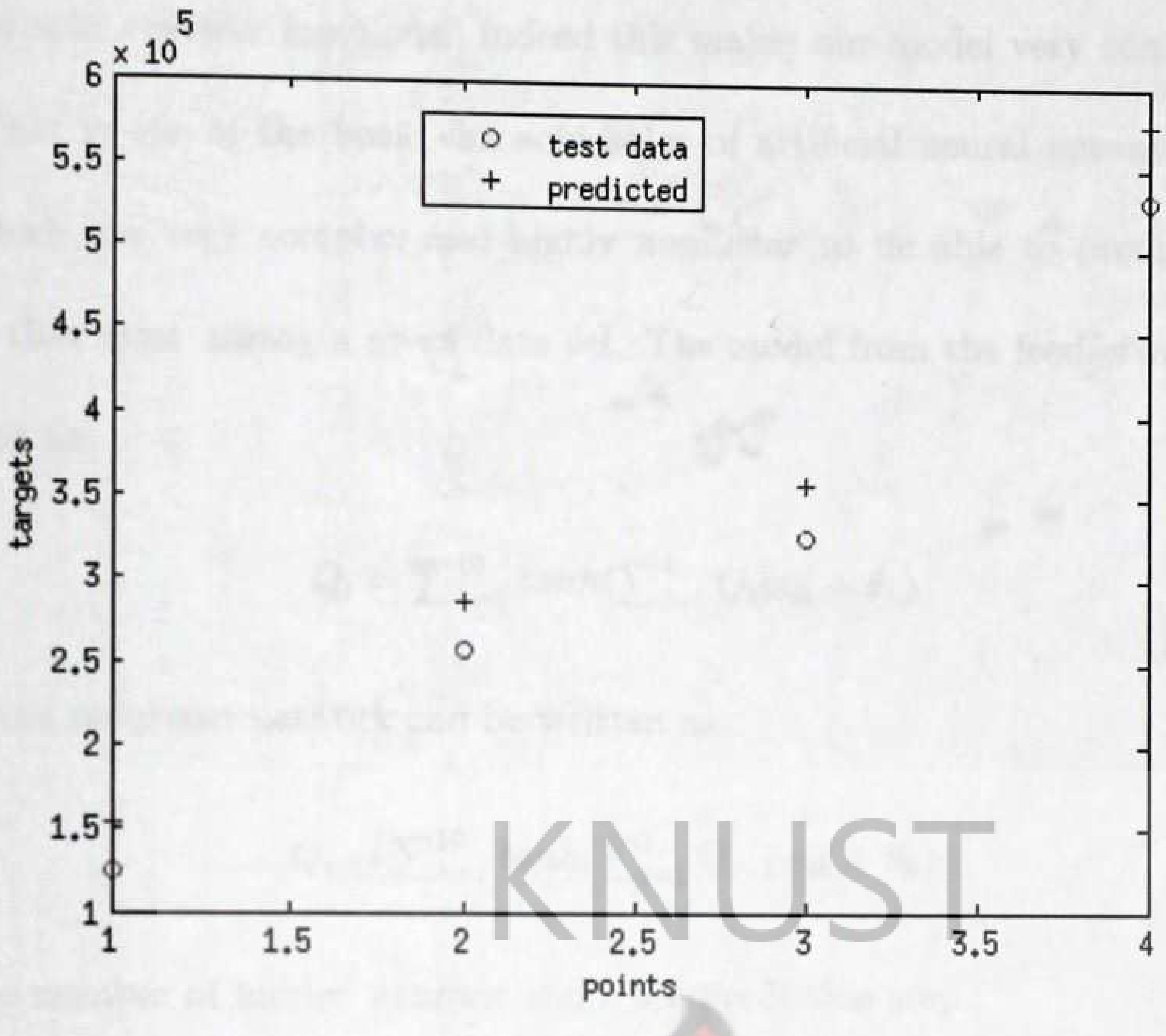

Fig. 4.8 Testing for Feedforward Network

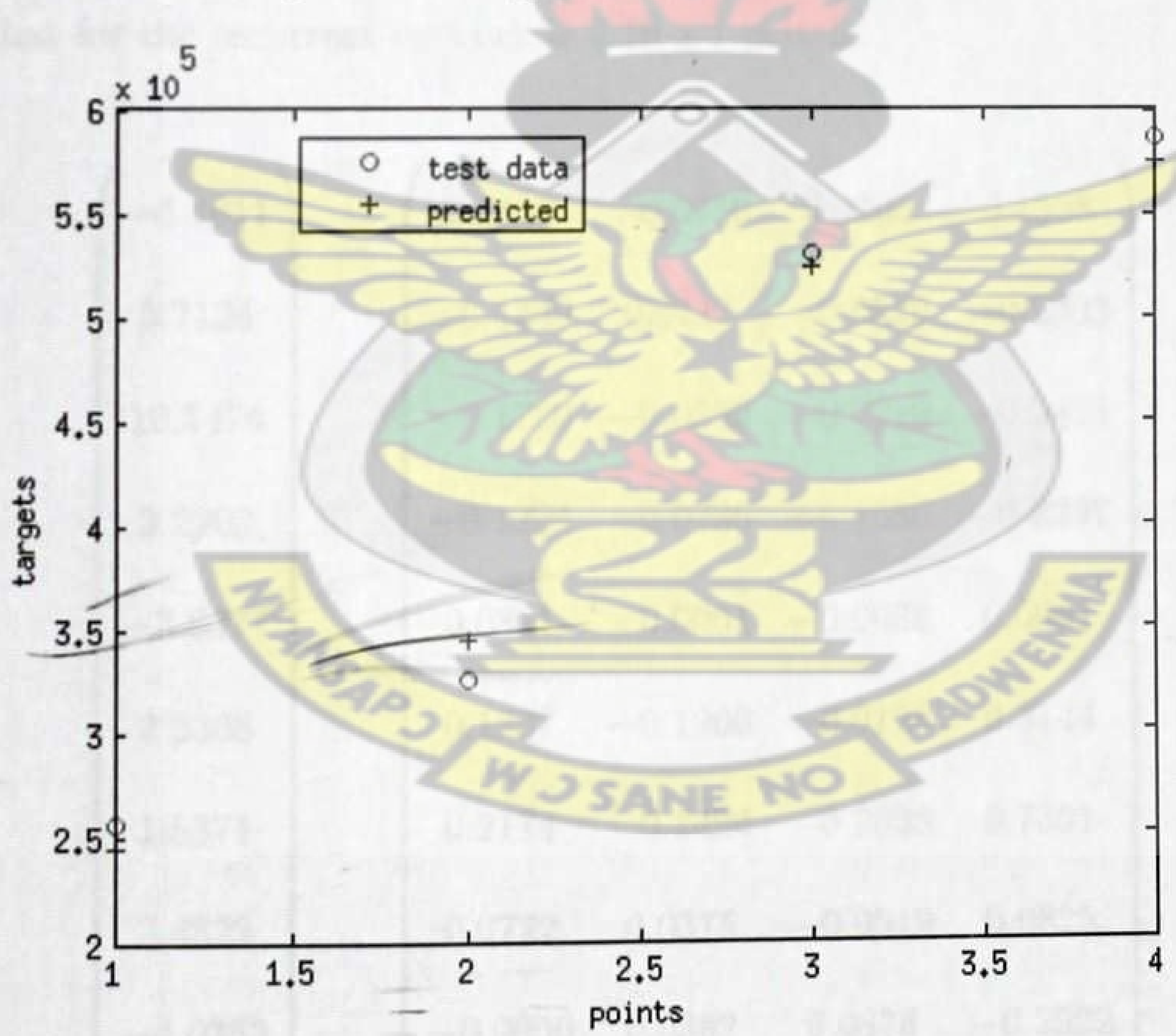

Fig. 4.9 Testing for Recurrent Network

In summary two models from both networks were obtained with 10 hidden neurons each in the layer. This can be seen from the two tables 4.3. These hidden neurons are equivalent to 10 tangent hyperbolic transfer functions. Indeed this makes our model very complex and highly nonlinear. That is one of the basic characteristics of artificial neural networks. They result in models which are very complex and highly nonlinear to be able to predict patterns and relationships that exist among a given data set. The model from the feedforward network can be represented as:

$$Q_t = \sum_{k=1}^{10} tanh(\sum_{i=1}^{4} Q_i w_{ik} + \theta_k)$$

and that for the recurrent network can be written as

$$Q_t = \sum_{k=1}^{10} tanh(\sum_{i=1}^{1} Q_{t-1} w_{ik} + \theta_k)$$

where k is the number of hidden neurons and t the prediction step.

The associated weights for the feedforward network is given by the highly dimensioned 10 x 4 matrix and that for the recurrent network is a 10 x 1 matrix.

$$\begin{pmatrix} -4.1611 \\ 5.7124 \\ 10.5474 \\ 3.2903 \\ -3.6377 \\ 2.3308 \\ 1.5371 \\ 3.4822 \\ -4.0353 \\ -0.5226 \end{pmatrix} and \begin{pmatrix} 0.0427 & -0.0750 & 0.0204 & 0.1538 \\ -0.1549 & 0.0112 & -0.0539 & -0.2303 \\ -0.1789 & -0.0359 & -0.0219 & -0.3435 \\ -0.1294 & -0.0735 & -0.1636 & -0.2297 \\ 0.0821 & -0.0061 & -0.0028 & 0.0981 \\ 0.1337 & -0.1200 & -0.0126 & 0.3144 \\ 0.2174 & -0.2424 & -0.2638 & 0.7391 \\ 0.0722 & 0.0375 & -0.0519 & 0.0825 \\ -0.0006 & 0.0382 & 0.0876 & -0.2522 \\ -0.0529 & -0.0414 & -0.0487 & -0.1424 \end{pmatrix}$$

Chapter 5

Conclusion

5.1 Introduction

This chapter is concerned with the major findings of the study. The findings are outlined in direct response to the objectives of this study. Also this chapter proposes future researches that can be carried out using Artificial Neural Networks. It also gives some recommendations on how Artificial Neural Networks could be used efficiently.

5.2 Summary of Findings

In this study we have used two types of ANN architectures to forecast the yearly tourist demands of Ghana. We did this analysis with the purpose of applying two types of ANN networks that can increase the forecasting accuracy of tourists arrivals data. The two types of ANN, i.e. the Feedforward and the Recurrent (NARX), were trained. After training and testing, the recurrent network predicted better that the feedforward network. Added to this, both networks did train well but the recurrent was able to generalize by predicting the test data with minimal errors. We also found out that a small learning rate is more effective since it slows down the learning process and allows the network to adapt well. With large learning rates it is very likely for the network to over fit. Recurrent networks are time consuming but they end up having the best training which can generalize when an appropriate algorithm and a large data size is fed into it. The weights (parameters) obtained in this work would be very difficult to be obtained if one tries to use any traditional method.

5.3 Recommendation

The ability of Artificial Neural Network to learn is proportional to the number of the training samples. In this thesis, the tourist data available is only on 20 years, which makes the training of the networks very difficult. It is recommended that the data which will be used in a neural network should be subdivided in other to attain a large sample data. In this case tourist arrivals forecasting can be better improved by using data acquired on a monthly or daily bases from the Ghana Tourist Authority. In future research other learning algorithms and types of network should be used in the prediction of tourist arrivals. Also we recommend that an automated procedure is used to obtain the optimal learning rate and performance ratio. Because these two parameters are very key in the network training.

5.4 Conclusion

In conclusion, in a time series forecasting, in the case of the Ghanaian tourist arrival data, best results is obtained using a recurrent network with 10 hidden hyperbolic tangent neurons. In our approach artificial neural network can be used for tourists arrival forecast in order to increase forecasting accuracy for the future. This can improve tourist management for a better performance by the industry. The ability of the artificial neural network to learn is proportional to the number of the training samples as already stated so the tourist authority should device a means to improve their data collection on tourists. If the number of samples is increased, the learning performance can be further improved. In conclusion, in a time series forecasting, best result is obtained using a NARX Neural Network. In Ghana's case an Artificial Neural Network approach can be used for a whole lot of forecasting application in order to increase forecasting accuracy for any data set. Artificial Neural Networks can be very useful to Ghana not only in predicting tourist arrivals but in any field where data is available.

5.5 Appendix

Table 5.1: Tourist Arrivals and Receipts

Year	Arrivals ('000)	Receipts (US$ million)
2001	438.8	447.8
2002	482.6	519.7
2003	530.8	602.8
2004	583.8	649.4
2005	428.6	836.1
2006	497.1	986.8
2007	586.6	1,172.0
2008	698.069	1,403.1
2009	802.779	1,615.2

Source: Ghana in Figures.

Table 5.2: Number of hotels, rooms and beds

Year	Number of Hotels	Hotel rooms	Number of beds
2001	1,053	15,453	19,648
2002	1,162	15,992	21,227
2003	1,250	17,352	22,909
2004	1,313	18,022	23,430
2005	1,341	18,675	23,828
2006	1,405	22,467	27,569
2007	1,407	18,683	26,057

Source: Ghana in Figures

5.6 Reference

Girish Kumar Jha,(2007). Artificial Neural Networks and its Applications. I.A.R.I., New Delhi-110 012.

T. Koskela, M. Lehtokangas, J. Saarinen, and K. Kaski, (2007). Time Series Prediction with Multilayer Perceptron, FIR and Elman Neural Networks. Tampere University of Technology Electronics Laboratory, FIN-33101 Tampere, Finland.

J. Maria, P. Junior and Guilherme A. Barreto, (2007). Long-Term Time Series Prediction with the NARX Network: An Empirical Evaluation. Department of Teleinformatics Engineering Federal University of Ceara, Av. Mister Hull, S/N CP 6005, CEP 60455-760, Fortaleza-CE, Brazil.

S. Ali, K. Al-Omari, Putra Sumari, S. A. Al-Taweel and A. J.A. Husain, (2009). Journal of Computer Science 427-434, ISSN 1549-3636 2009 Science Publications Digital Recognition using Neural Network.

Eugen Diaconescu,PhD (2008) The use of NARX Neural Networks to predict Chaotic Time Series Electronics, Communications and Computer Science Faculty, University of Pitesti, Romania.

T.Jayalakshmi, A.Santhakumaran,PHD (2011). Statistical Normalization and Back Propagation for Classification. International Journal of Computer Theory and Engineering, Vol.3, No.1.

F. Jurado, A. Caño and M. Ortega "Neural networks and fuzzy logic in electrical engineering control courses", International Journal of Electrical Engineering Education, Department of Electrical Engineering, University of Jaén, Linares (Jaén), Spain.

L. Smith. "An Introduction to Neural Networks". Centre for Cognitive and Computational Neuroscience. April 2, 2003. Retrieved on June 24, 2011.

NeuroAI, Artificial Neural Networks, 2007; Digital Signal Processing, Algorithms and Applications. Extracted on June 24, 2011.

R. C. Chakraborty (2010); “Back Propagation Network”; Soft Computing Course Lecture 15-20, notes, slides, August 10, 2010.

K. Mehrotra , K. C. Mohan, S. Ranka (2000); “Elements of Artificial Neural Networks”.

M. P. Wallace (PHD) (2008); “Neural Networks and their Application to Finance”.

L. Fu, Neural Networks in Computer Intelligence (McGraw-Hill, New York, 1994).

V. B. Teye, 2002. Tourism Development Experience in Ghana.

Ghana Tourist Authority, (2009). Tourism Statistical Fact Sheet on Ghana.

H. K. D. H. Bhadeshia, (2008). Neural Networks and Information in Materials Science Department of Materials Science and Metallurgy, University of Cambridge, Cambridge CB2 3QZ, U.K.

G. P. Zhang (2004). Neural Networks in Business Forecasting,Georgia State University, USA.

H. Demuth and M. Beale (1992). Neural Network Toolbox For Use with MATLAB.

M. H. Al Shamisi, A. H. Assi and H. A. N. Hejase, (2011).Using MATLAB to Develop Artificial Neural Network Models for Predicting Global Solar Radiation in Al Ain City – UAE,United Arab Emirates University.

C. A. Mitrea, C. K. M. Lee and Z. Wu (2009).A Comparison between Neural Networks and Traditional Forecasting Methods.

T.Jayalakshmi and A.Santhakumaran, (2011). Statistical Normalization and Back Propagation for Classification, International Journal of Computer Theory and Engineering, Vol.3, No.1, February, 2011 1793-8201.

A. K. Kain, J. Mao and K. M. Mohiuddin, Artificial Neural Network: A Tutorial, March 1996.

Taylor and Francis, Fundamentals of Neural Networks and Models for Linear Data Analysis, 2006.

Taylor and Francis, Neural Networks for Time-Series Forecasting, 2006.

Table 5.3: Number of hotels, rooms and beds

Year	Arrivals
1988	113784
1989	125162
1990	145780
1991	172464
1992	213316
1993	256680
1994	271310
1995	286000
1996	304860
1997	325438
1998	347952
2001	438800
2002	482600
2003	530800
2004	583800
2005	428600
2006	497100
2007	586600
2008	698069
2009	802779

Source: Ghana in Figures