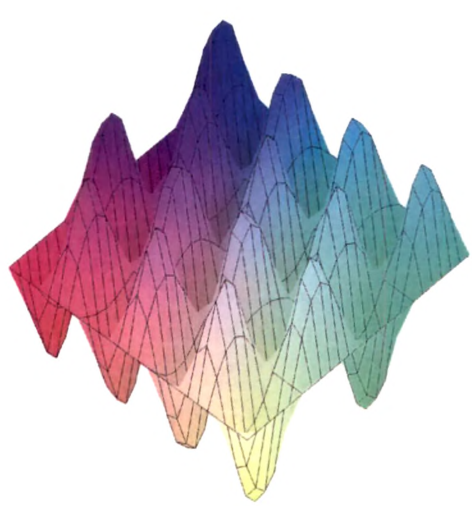
Chapter 2

PRELIMINARIES



2.1 INTRODUCTION

In this chapter, we provide the necessary basics for the tools we use for prediction of rainfall, air temperature and soil temperature and also discuss some background of the problems and data.

2.2 CATCHMENT AREA USED IN ANALYSIS

The Gujarat state of India occupies the 5 98 per cent area of the country in the western part, located at latitude of 20°1'N to 24°7' N and longitude of 68°4' E to 4°4'E, which covers 1, 95, 984 square kilometers. Western border of the state is coastline and therefore, it is one of the important maritime states in the country. The southwest monsoon brings rains in the state and this influences the atmospheric condition. The physiography and Thar Desert also influence the physical condition in the state. Climatically, most part of the state falls in sub-humid (south) semi arid (Middle) and arid (Western) zone in the northern and northwestern extremities.

In the present study, we use rainfall data series (DS) of annual rainfall from1958 to 2006, soil temperatures DS from 1982-2005, hourly air temperatures of the year 1992 are observed at Observatory, Anand agricultural University, Anand. In chapter 4, for extreme Value Analysis, we use daily data from 1901 to 1992 of 58 stations of 8 different Agro-climatic zones observed by India Meteorological department, Air Port, Ahmedabad.

2.3 DEFINITION OF WEATHER AND CLIMATE

I) WEATHER (Ghadeker, [39])

Physical state of the atmosphere at a particular place and time is called weather. Weather is a result of interactions of soil-air-water-solar radiations. The instantaneous result of these interactions can be measured by sensors/instruments in the observatory. It is measured in terms of wind speed, temperature, humidity, atmospheric pressure, cloudiness, and precipitation. In most places, weather changes from hour-to-hour, day-to-day, and season-to-season.

Thus, observed weather parameters always dynamic in nature. Weather of individual season decides the crop yield in that particular season. Adverse weather results in crop failure or loss in yield and compels short term planning Weather on micro-scale can be modified. For example mulching, solarization, covering the field with different plastic cover, irrigation, etc.

II) CLIMATE

The long period average of weather elements in a region is known as climate. Thus climate is derived information on regional basis and it constitutes geographical information Climates are classified as desert, continental, marine, savanna and tropical. Climatic conditions of a region decide the suitablity of the crop to be grown. It decides crop yield potentiality in a region and it is basis in a long term planning in the agriculture sector

2.3.1 WEATHER AND CLIMATE PREDICTION

Prediction of climate is a difficult task. Human being was trying to predict the climate since its civilization by astrological methods. Advent of computer and electronic technology facilitate the scientists for prediction of climate is a more efficient way.

There are number of large computer models called General Circulation Models, or GCMs, that have been built to study and predict climate. These models use the basic laws of science (conservation of mass, energy and momentum etc.) to represent the large-scale circulations and interactions of the earth -atmosphere system.

In order to understand and predict the climate system better, one has to understand the basic climatic processes. Many ongoing research programs, both in the United States, and elsewhere around the world, are dedicated to producing this better basic knowledge.

The climate system is extraordinarily complex and one may never be able to predict fully how it will respond. Indeed, within limits it may even be a "chaotic system," which would mean that precise predictions would never be possible.

However, the weather elements are express as climatic parameters by Mathematical terms, like normal measurements (Conard [22])

(a) Estimation of central tendency OR Mean values of the major elements.

(b) Estimation of periodic, quasi-periodic or non-periodic variability in nature/dispersion about the averages.

Climatic variability gives major problems from agriculture system and food production year-to-year.

2.3.2 WATER CYCLE

The hydrological cycle of earth atmosphere is one of the example in which water evaporate from the water body and vegetation due to incident of solar energy and thereby maintaining the temperature of atmosphere. Evaporated water in the gaseous form diffuse in atmosphere and condensed leads to waterfall that is Rainfall. This rainwater again gains as a runoff and accumulates in reservoir, infiltrate in soil and recharge groundwater level, runoff in the river and sea. These rainfall activities also control the air and earth temperature

2.4 FORMULATION OF PREDICTION PROBLEM

Man has built up his progress through information and knowledge of his surrounding regions. In many situations, he requires unavailable information for planning and acting therefore, he tries to forecast (predict) that information using his available knowledge about phenomena that have already occurred or may occur in the future.

Thus, without having data or information one can make a judgment forecast and with availability of information one can approach/make forecast using statistics or mathematics laws, which has high accuracy in composition to judgment forecasts. Types of forecasts release by India Meteorological Department (IMD) are

1) Short Range Forecasts (SRF): Covering a period of a few hours to two days

2) Medium Range Forecasts (MRF): Covering a period from 3 to 10 days.

3) Long Range Forecasts (LRF). Covering a period of a month or a season.

2.4.1 DETAILS OF THE INSTITUTION ISSUING WEATHER PREDICTION

The India Meteorological Department (IMD), India issues long-range forecasts of all - India monsoon rainfall every year. To give accurate LRF, first approach is by selecting predictors by having highest correlation coefficients (Krishna Kumar [88]) and then to develop a model like Multiple Regression Multiple Power Regression or Dynamic Stochastic Transfer Techniques etc.

Atmospheric study is global in nature, so it requires international co-operation. Therefore, there is a formation of World Meteorological Organization (WMO) as a special Agency of United Nations. Advantage of satellites, meteorology has global viewing continuous homogenous data, which was highly needed for prediction of weather parameters. The satellite programmes advance their research and adding new sensors to improve the capability of prediction or study of the atmosphere. Now many countries have their own meteorological satellites and the data's being made available to other countries through the global telecommunication network.

2.4.2 NATURE OF THE RAINFALL IN INDIA

Rainfall is a discrete variable process in any part of the India Commencement of rainfall may be early or delayed, frequent breaks in July or August, it terminate earlier or persist longer time than usual and also rainfall may be very unevenly distributed in space and time since it may be excessive in one part of the country and deficient in another part (Rao [127]) Thus distribution of rainfall in India is erratic and its behavior unpredictable as climate changes (Pramanik *et al.* [123]) due to its determination largely by the physical features and orientations of mountains and plateau with regard to prevailing winds. About 85 1 % of total rainfall of the year occurs during the southwest monsoon (June-September) season.

2.4.3 DIFFERENT METHODS USED FOR PREDICTION OF RAINFALL BY INDIA METEROLOGICAL DEPARTMENT (IMD)

There is a tremendous impact of rainfall patterns in agricultural planning (Sharma *et al.* [137]; Sharma *et al.* [138]). Long Range Forecast (LRF) is a challenging task especially in the modern world where one is facing the major environmental problems of global warming. Therefore, accurate long-range forecast of rainfall requires proper method in right direction. Most of the approaches to forecast the rainfall uses either a Dynamical Prediction Model or Statistical Methods.

In general, rainfall is highly non-linear phenomena in nature exhibiting what is known as the 'butterfly effect' i.e the fluttering wings of a butterfly at one corner of the globe may ultimately cause a tornado at another geographically far away place. Lorenz [101] discovered this phenomenon, which is popularly known as the butterfly effect.

To give accurate LRF, first approach is by selecting predictors by having highest correlation coefficients and then to develop a model like

(I) Multiple Regressions (Murphy *et al.* [110]; Walker [164], Thapliyal [153])

(II) Auto Regressive Integrated Moving Average Technique (ARIMA) (Gowariker *et al* [47] ; Vaziri [162]).

(III) Multiple Power Regression (Gowariker [47]; Paranjpe *et al* [119])

(IV) Dynamic Stochastic Transfer Techniques etc (Thapliyal [152]) and

(V) Harmonic Analysis Technique

In recent years, new modeling technique Artificial Neural Network (ANN) (Wiberg *et al* [168]) is mainly used by statistician, mathematician and meteorologist etc. as an alternative for solution of complex problem like prediction of monsoon rainfall or rainfall runoff, temperature, antifungal activity, medicine etc.

An attempt has been made in this thesis to predict the

(i) Annual Rainfall (ARF),

(ii) Weekly Rainfall (WRF) and its probabilities (PBs),

(iii) Highest One day Maximum Rainfall (HOMRF),

(iv) Hourly Air Temperatures (HAT) and weekly Soil Temperatures (ST), for Anand station of Gujarat by using the artificial neural network (ANN) techniques.

2.5 ARTIFICIAL NEURAL NETWORK (ANN)

An Artificial Neural Network (ANN) is a flexible mathematical structure, which is capable of identifying complex non-linear relationship (Jacques *et al* [72]) between input and output data set. ANN models have been found useful and efficient, particularly in problems for which the characteristics of the processes are difficult to describe using Mathematical equations (Zaldivar *et al* [173]) Thus, ANN provides input-output simulation and forecasting models in situations that do not require modeling of the internal structure of the parameters.

An Artificial Neural Network (ANN) is a massively parallel distributed processor made up of highly networked simple processing units (called neurons; Fig: 2.1) which has natural ability for storing experimental knowledge and making it available for future use. The ANN is a new information-processing paradigm that is motivated by the working of brain in human or other species for the processing of information It resembles the brain in two things:

i. Knowledge is acquired by the network from its environment through a learning process

ii. Inter-neuron connection strengths (known as synaptic weights) are used to store the acquired knowledge.

NEURON ELEMENTS

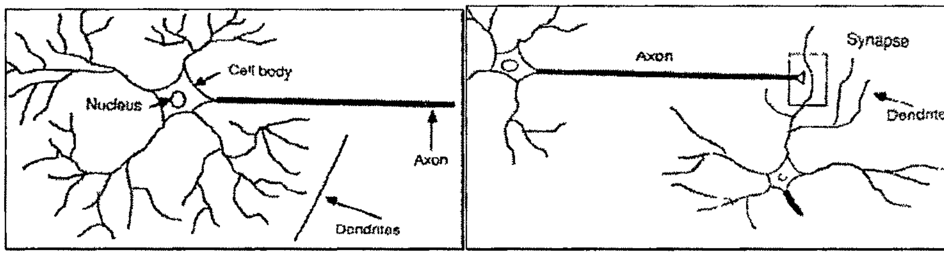


Fig: 2.1

These Artificial Neural Network can be interpreted as adaptive machines, which can store knowledge through the learning process. Neural Network methods are applied for example, in hydrological prediction (Kingston *et al* [85]), pattern recognition, vision, speech recognition, classification, and control systems.

2.5.1 A NEURON

A more sophisticated neuron (Figure 2.2 (a)) is the McCulloch and Pitts model (MCP). The first step toward artificial neural networks came in 1943 when Warren McCulloch, a neuron physiologist, and a young mathematician, Walter Pitts, wrote a paper on how neurons might work. They modeled a simple neural network with electrical circuits.

The weight associated with connection of an input is a number which when multiplied with the input gives the weighted input. These weighted inputs are then added together and if they exceed a pre-set threshold value, the neuron fires. In any other case the neuron does not fire

A MCP NEURON

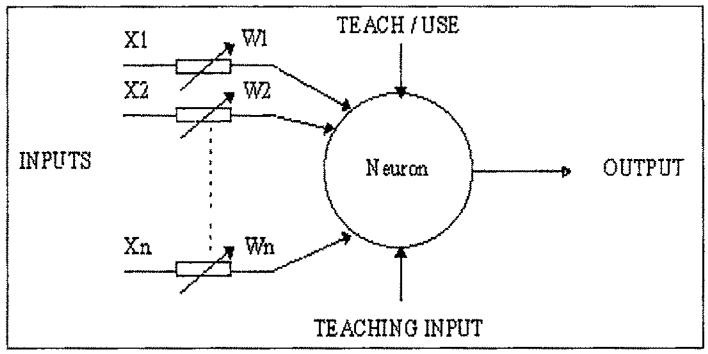


Fig: 2.2 (a)

The MCP neuron has the ability to adapt to a particular situation by changing its weights and/or threshold. Various algorithms like Delta rule and the Back error propagation rule etc. are used to train a network to assign the right values of weights.

A Single Artificial Neuron - Perceptron

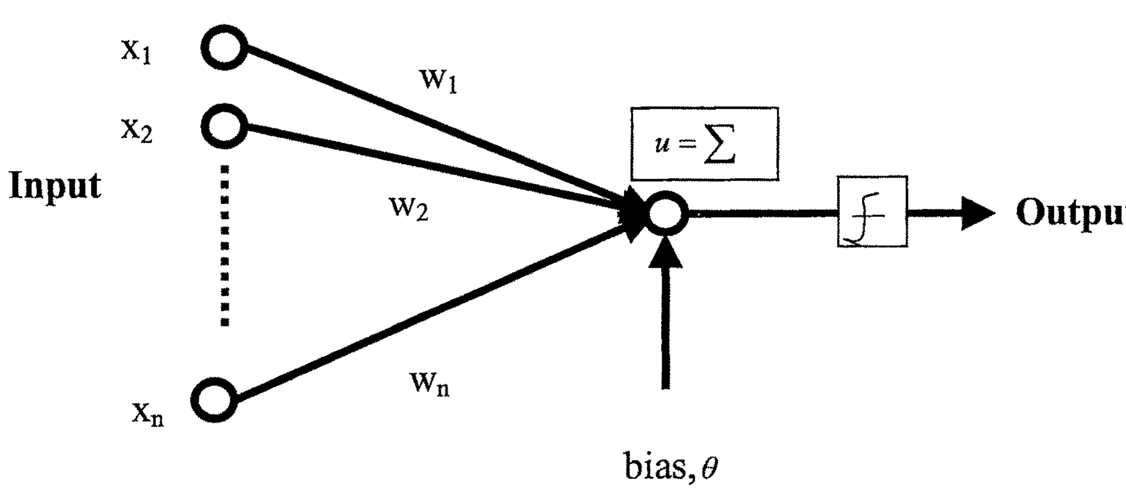


Fig: 2.2(b)

A single artificial neuron model (perceptron) consists of the following components (Fig· 2.2 (b)):

1. Input x : It is an n vector $x = [x_1, x_2, ... x_n]^T$ inputted to the neuron.
2. A set of synaptic weights w_i i= 1,2.......n, representing the connecting strength of a neuron.
3. An adder ($\sum$) for summing the weighted inputs to a neuron.
4. A transfer (activation) function f for limiting the amplitude of the output of a neuron.
5. The bias θ has the effect of increasing or lowering the net input of the activation function, depending on whether it is positive or negative respectively. The bias 'θ' has the effect of applying an affine transformation to the input of the neuron
6. Output y. The output is computed as $y = f(\sum_{i=1}^{n} w_i x_i - \theta)$.

2.5.2 TYPES OF ANN

There are many types of neural network models. The models can be categorized in many ways. One possibility is to classify them on the basis of the learning principle. There are five basic learning rules. Those are

1. Error correction learning: Rooted in optimum filtering
2. Memory based Learning: Operates by memorizing the training data explicit.

3. Hebbian Learning· Inspired by neurobiological considerations.
4 Competitive Learning: Inspired by neurobiological considerations.

Boltzmann Learning: Based on ideas borrowed from statistical mechanics

There are two types of ANN based as learning methods:

(I) Supervised learning

and

(II) Unsupervised learning.

(I) SUPERVISED LEARNING

In supervised learning network is provided with example cases and desired responses. The network weights are then adapted in order to minimize the difference between network outputs and desired outputs.

Examples of supervised learning are

1. Perceptron learning rule
2. Delta learning rule
3. Widrow-Hoff learning rule

(II) UNSUPERVISED LEARNING

In the unsupervised learning the network is given only input signals, and the network weights change through a predefined mechanism, which usually groups the data into clusters of similar data.

Examples of Unsupervised Learning are

1. Hebbian Rule
2. Winner-take-all

Artificial Neural Network are also classified by their topology like. Feed-Forward Network, Multi-layer Perceptrone Network (MLP), Recurrent Network etc.

The MLP network consists of several layers of neurons. Each neuron in a certain layer is connected to each neuron of the next layer. There are no feedback connections. The neuron weights are considered as free parameters. The most often used MLP-network consists of three layers:

(i) Input layer, (ii) Hidden layer and (iii) Output layer.

A three layers MLP network is shown in the figure 2.3.

A THREE LAYER MLP NETWORKS

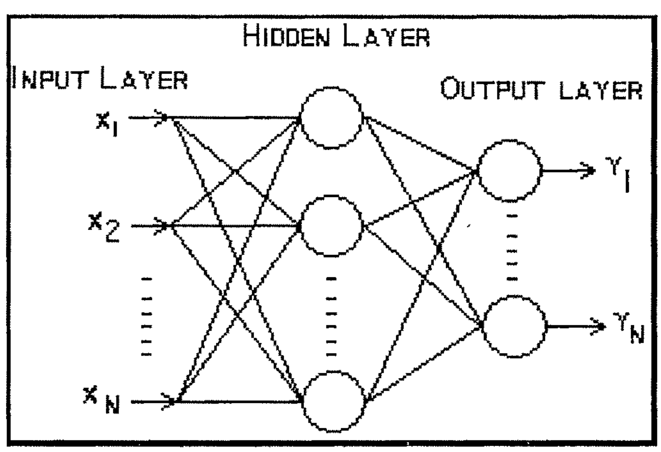


Fig: 2.3

As an N-dimensional input vector is fed to the network, an M-dimensional output vector is produced. The network can be understood as a function from the N-dimensional input space to the M-dimensional output space

The ANN models are researched in connection with many power system applications, short-term rainfall forecasting applications etc. Most of the suggested models use MLP networks. In the models, inputs variables to the network are generally the present and past data values and outputs variables are future data values. But the network is trained using actual historical data.

The basic device (neuron) of the network is a perceptron. This is a computation unit, which produces its output by taking a linear combination of the input signals and by transforming this by a function called activity function. The output of the perceptron as a function of the input signals can be given by, $y = f(\sum w_i x_i + \theta)$ where, y is the output variable; i=1,2,3...n , x_i is the input variables (signals), w_i is the synaptic weights of the neurons, θ is the bias term (another neuron weight) with fixed point 1 and f is the activity or transfer functions.

The activation function used in the hidden layer is usually nonlinear (logistic sigmoid or hyperbolic tangent) and the activation function in the output layer can be either nonlinear (a nonlinear-nonlinear network) or linear (a nonlinear-linear network). Possible forms of the activity function are linear function, step function, logistic function and hyperbolic tangent function etc. These are shown below (Cybenko [25]):

$$f(x) = \frac{1}{1+e^{-\alpha x}} \quad \textit{is the logistic Sigmoid Function.}$$

$$f(x) = \tanh(\alpha x) \quad \textit{is the Hyperbolic Tangent Function.}$$

$$f(x) = \begin{cases} x & \text{when } x>0 \\ 0 & \text{when } x \le 0 \end{cases} \quad \textit{is the Piecewise Linear Function.}$$

where, ' α ' is a constant and 'x' is the input variable to the neuron.

The inputs x_i s is multiplied by the connecting weights w_{ij}. The weighted inputs $x_i * w_{ij}$ to each of the hidden neuron is added and passed through the activation function. Input to the first neuron in the hidden layer is:

$$x_1w_{11}+x_2w_{21}+x_3w_{31}+x_4w_{41}+ \quad +x_nw_{n1}$$

The output of the hidden neuron of the j^{th}-hidden layer would be:

$$h_j = \frac{1}{(1+e^{-\sum w_{ij}x_i})}$$

The output of the hidden neurons is then multiplied by the connection weights to the output neuron, w_k and passed on to the output layer. The weighted outputs of the hidden layer are added together and then passed through the activation function. The input to the output neuron is

$$h_1w_1+h_2w_2+h_3w_3+h_4w_4+h_5w_5+h_6w_6+h_7w_7+ \quad + h_nw_n$$

The output of the network would then be:

$$h_j = \frac{1}{(1+e^{-\sum w_ix_i})}.$$

The neural network of this type can be understood as a function approximate. It has been proved that given a sufficient number of hidden layer neurons, it can approximate any continuous function from a compact region of R^N to R^M at an arbitrary accuracy.

By giving training to the network, network weights are adjusted according to the given output from Historical data.

Thus, the network learns through examples. The idea is to give the network, input variables with desired outputs. To each input signal the network produces an output signal, and the learning aims at minimizing the sum of squares of the differences between desired and actual outputs. The learning is carried out by trial and error, i.e. repeatedly feeding the input-output patterns to the network. One complete presentation of the entire training set is called an epoch. The learning process is usually performed on an epoch-by-epoch basis until the weights stabilize and the sum of squared errors converges to some minimum value The mostly used learning algorithm for the MLP-networks is the back propagation algorithm. This is a specific technique for implementing gradient descent method in the weight space, where the gradient of the sum of squared errors with respective to the weights is approximated by propagating the error signals backwards in the network. The Equation for training in the back propagation algorithm in the output layer is

$$w_j(n+1) = w_j(n) + r\ y\ *(y' - y)*(1- y)*h_j*\ w_j*(1- h_j)*x_j$$

where, y' is the actual output,

y: the output of the network ,

r: the learning factor (0.1to0.001),

n. the number of learning cycles(Epochs),

x_j· the input that corresponds to the output y′,

h_j output of the hidden neuron of the j^th^-hidden layer

2.5.3 LEARNING LAW: THE DELTA RULE

Delta rule, which is the continuous version of the discrete perceptron learning rule. Thus, it is also known as the continuous perceptron rule. The learning signal for this rule is called delta and is defined by

$$r = [d - f(w^T x)] f'(w^T x)$$

The term $f'(w^T x)$ is the derivative of the activation function f(net) computed for $net = w^T x$.

This rule is based on the simple idea of continuously modifying the strengths of the input connections to reduce the difference that is delta between the desired output value and the actual output of a processing element. This rule changes the synaptic weights in the way that minimizes the mean squared error of the network. This rule is also referred to as the Widrow-Hoff [169] Learning Rule and the Least Mean Square (LMS) Learning Rule. The way that the Delta Rule works is that the delta error in the output layer is transformed by the derivative of the transfer function and is then used in the previous neural layer to adjust input connection weights. In other words, this error is back propagated into previous layers one layer at a time. The process of back-propagating the network errors continues until the first layer is reached. The network type called Feed forward, Back-propagation derives its name from this method of computing the error term

The explanation of the delta learning rule is as shown below:

Let $x(k)=[x_1(k),x_2(k),...,x_n(k)]^T$ $,k=1,2,.p$ be p input patterns and d(k), k= 1,2,..p be the corresponding output patterns of the perceptron network. Let y(k) , k =1,2,.. ,p be the output of the network with input x(k) and weight w(k) . Let e (k) = [d(k)-y(k)] be the error when we inputted the kth pattern. The mean squared error is given by·

$$E=\frac{1}{2}(d(k)-y(k))^2$$

$$E=\frac{1}{2}[d(k)-f(w^Tx)]^2$$

We want to minimize E with respect to w. The error (or energy) E is decreasing at the highest rate in the direction of the negative gradient of E.

Highest rate minimizing direction = - gradient E.

$$\begin{aligned}&=-[d-f(w^Tx)](-f'(w^Tx)x)\\&=[d-f(w^Tx)]f'(w^Tx)x\end{aligned}$$

Let η be a positive constant and we take the increment vector as

$$\Delta w_i=\eta\ (d-f(w^Tx))f^1\ (w^Tx)x$$

Thus weight update rule to minimize the error E is given by

$$w(k+1)=w(k)+\eta\ (d-f(w^Tx))f^1\ (w^Tx)x$$

When using the delta rule, it is important to ensure that the input data set is well randomized Well-ordered or structured presentation of the training set can lead to a network, which cannot converge to the desired accuracy If that happens, then the network is incapable of learning the problem.

2.5.4 APPROXIMATION CAPABILITIES OF FEED FORWARD NEURAL NETWORK FOR CONTINUOUS FUNCTIONS

We have the well-known approximation theorem:

❖ **WEIERSTRASS'S APPROXIMATION THEOREM**

Let g be a continuous real- valued function defined on a close interval [a,b]. Then, given any ε positive, there exists a polynomial y (which depend on ε) with real coefficients such that

$$| g(x) - y(x) | < \varepsilon \qquad \text{for every } x \in [a,b]$$

We now present some fundamental results in the form of theorems, on continuous function approximation capabilities of feed forward nets. The main result is that two-layer feed-forward net with a sufficient number of hidden units, of the sigmodial activation type, and a single linear output is capable of approximating any continuous function $f : R^n \to R^n$ to any desired accuracy.

Before formally stating this result, let us consider some early observations on the implications of a classic theorem on function approximation, Kolmogorov's theorem, which motivates the use of layered feed forward nets as function approximators.

❖ **KOLMOGROV'S THEOREM**

It has been suggested (Hecht [61]) that Kolomogrov's theorem concerning the realization of arbitrary multivariate functions provides theoretical support for neural network that implement such function.

❖ **THEOREM** : Any continuous real-valued functions $f(x_1, x_2, .., x_n)$ defined on $[0,1]^n$, $n \geq 2$, can be represented in the form ·

$$f(x_1, x_2, .., x_n) = \sum_{j=1}^{2n+1} g_j [\sum_{i=1}^{n} \varphi_{ij}(x_i)]$$

where the g_j terms are properly chosen continuous functions of one variable, and the φ_{ij} functions are continuous monotonically increasing functions independent of *f*.

The theorem states that one can express a continuous multivariate function on a compact set in terms of sums and compositions of a finite number of single variable functions.

2.5.5 SINGLE-HIDDEN-LAYER NEURAL NETWORK ARE UNIVERSAL APPROXIMATORS

Rigorous mathematical proofs for the universality of feed forward layered neural nets employing continuous sigmoid type, as well as other more general, activation units were given independently , by Cybenko[25].

Cybenko's proof is distinguished by being mathematically concise and elegant [it is based on Hahn-Banach theorem (2.5.8)].

2.5.6 LIPSCHITZ CONDITION

A function f, from M to itself is said to satisfy a Lipschitz condition if there is some real number $0 < k < 1$ such that, for all x and y in M,

$$d(f(x), f(y)) \leq k\, d(x, y).$$

The smallest such value of k is called the Lipschitz constant of f.

In mathematics, functions between ordered sets are said to be monotonic (or monotone, or isotone) if they preserve the given order.

2.5.7 BANACH FIXED POINT THEOREM

The Banach fixed point theorem (known as the contraction mapping theorem) is an important tool in the theory of metric spaces, it guarantees the existence and uniqueness of fixed points of certain self maps of metric spaces

❖ **THEOREM**

Let (X, d) be a non-empty complete metric space. Let $T : X \to X$ be a contraction mapping on X, i e: there is a nonnegative real number $q < 1$ such that

$$d(Tx, Ty) \leq q \cdot d(x, y)$$

for all x, y in X. Then the map T admits one and only one fixed point x^* in X (this means $Tx^* = x^*$).

Moreover, this fixed point can be found as below:

start with an arbitrary element x_0 in X and define an iterative sequence by $x_n = Tx_{n-1}$ for $n = 1, 2, 3, \ldots$ This sequence converges and its limit is x^*

2.5.8 THE BACK-PROPAGATION ALGORITHM

Units are connected to one another. Let W_{ij} be the weight of the connection from unit u_i to unit u_j. It is then convenient to represent the pattern of connectivity in the network by a weight matrix W whose elements are the weights W_{ij}. Two types of connection are usually distinguished: excitatory and inhibitory. A positive weight represents an excitatory connection whereas a negative weight represents an inhibitory connection. The pattern of connectivity characterizes the architecture of the network.

A unit in the output layer determines its activity by following a two-step procedure.

First, it computes the total weighted input x_j, using the formula·

$$X_j = \sum_i y_i W_{ij}$$

where y_i is the activity level of the i^{th} unit in the previous layer and w_{ij} is the weight of the connection between the i^{th} and the j^{th} unit.

Next, the unit calculates the activity yj using some function of the total weighted input. Typically we use the sigmoid function·

$$y_j = \frac{1}{1+e^{\alpha x_j}}$$

Once the activities of all output units have been determined, the network computes the error E, which is defined by the expression:

$$E = \frac{1}{2}\sum_i (y_i - d_i)^2$$

where y_i is the activity level of the i^{th} unit in the top layer and d_i is the desired output of the i^{th} unit.

The back-propagation algorithm has four steps:

1. Compute how fast the error changes as the activity of an output unit is changed. This error derivative (EA) is the difference between the actual and the desired activity.

$$EA_j = \frac{\partial E}{\partial y_j} = y_j - d_j$$

2. Compute how fast the error changes as the total input received by an output unit is changed. This quantity (EI) is the answer from step 1 multiplied by the rate at which the output of a unit changes as its total input is changed.

$$EI_j = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \times \frac{dy_j}{dx_j} = EA_j y_j (1 - y_j)$$

3. Compute how fast the error changes as a weight on the connection into an output unit is changed. This quantity (EW) is the answer from step 2 multiplied by the activity level of the unit from which the connection emanates

$$EW_{ij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial W_{ij}} = EI_j y_i$$

4. Compute how fast the error changes as the activity of a unit in the previous layer is changed. This crucial step allows back propagation to be applied to multilayer networks. When the activity of a unit in the previous layer changes, it affects the activities of all the output units to which it is connected. So to compute the overall effect on the error, we add together all these separate effects on output units. But each effect is simple to calculate. It is the answer in step 2 multiplied by the weight on the connection to that output unit.

$$EA_i = \frac{\partial E}{\partial y_i} = \sum_j \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial y_i} = \sum_j EI_j W_{ij}$$

By using steps 2 and 4, we can convert the EAs of one layer of units into EAs for the previous layer This procedure can be repeated to get the EAs for as many previous layers as desired. Once we know the EA of a unit, we can use steps 2 and 3 to compute the EWs on its incoming connections

There are many variations to the learning rules for back-propagation network. Different error functions, transfer functions, and even the modifying method of the derivative of the transfer function can be used. Here, the error function, or delta weight equation, is modified so that a portion of the previous delta weight is fed through to the current delta weight. This acts, in engineering terms, as a low-pass filter on the delta weight terms since general trends are reinforced whereas oscillatory behavior is canceled out. This allows a low, normally slower, learning coefficient to be used, but creates faster learning.

The number of input-output pairs that are presented during the accumulation is referred to as an epoch. This epoch may correspond either to the complete set of training pairs or to a subset.

There are limitations to the feed forward, back-propagation architecture. Back-propagation requires lots of supervised training, with lots of input-output examples. Additionally, the internal mapping procedures are not well understood, and there is no guarantee that the system will converge to an acceptable solution.

2.5.9 APPLICATIONS OF ANN

Typical feed forward, back-propagation applications include speech synthesis from text, robot arms, evaluation of bank loans, image processing, knowledge representation, forecasting and prediction, and multi-target tracking etc. [5].

Neural network have broad applicability to real world business problems (Corne *et al.* [23]; Jorge Kazno [79] ; Raman *et al.* [124] ; Stanislaw [150]) In fact, they have already been successfully applied in many industries. Since neural networks are best at identifying patterns or trends in data, they are well suited for prediction or forecasting (Solomatine *et al.* [149]).

Their ability to learn by example makes them very flexible and powerful. There is no need to understand the internal mechanisms of the task. They are also very well suited for real time systems because of their fast response and computational times, which are due to their parallel architecture.

Basically, most applications of neural networks have following five categories:

1) Data association,
2) Data conceptualization,
3) Data filtering,
4) Prediction and Classification.

2.6 FOURIER SERIES FOR ONE VARIABLE (Grewal [48])

❖ PERIODIC FUNCTION

A function f (x) is said to be periodic if there exist a real number T for any x of the domain such that f (x+T)=f (x). P is said to be the fundamental period.

❖ REPRESENTATION OF FOUIER SERIES

Let f (x) be defined in the interval [0, P] and determined outside of this interval by its periodic extension, and let P is the period. The Fourier series corresponding to f(x) is

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(\frac{2\pi kx}{P}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2\pi kx}{P}\right) \qquad (2.11)$$

where , the Fourier coefficients a_n and b_n are

$$a_k = \frac{2}{P}\int_0^P f(x) \cos\left(\frac{2\pi kx}{P}\right) dx; \qquad k=1,2,\ldots$$

$$b_k = \frac{2}{P}\int_0^P f(x) \sin\left(\frac{2\pi kx}{P}\right) dx; \qquad k=1,2,\ldots$$

Amplitude and phase angle defined by

$$c_k = \sqrt{a_k^{\ 2} + b_k^{\ 2}} \ \ and \ \ \theta = \tan^{-1}\left(\frac{a_k}{b_k}\right); \qquad k = 1,2,..$$

2.6.1 CONDITONS FOR A FOURIER EXPANSION (Wylie, [172])

❖ **DIRICHLET'S CONDITIONS**

Any function f (x) can be represented as a Fourier series (2.11) where a_0, a_n and b_n are constants, provided:

(I) f(x) is periodic, single valued and finite;

(II) f(x) has finite number of discontinuities at any one period,

(III) f(x) has at the most a finite number of maxima and minima

2.6.2 CONVERGENCE THEOREM OF FOURIER SERIES

❖ **THEOREM: GIBB'S PHENOMENON**

Let f(x) be a function, which is twice differentiable, such that $f(x)$, $f'(x)$, and $f''(x)$ are piecewise continuous on the interval $(-\pi, \pi)$. Then, for any, $x \in (-\pi, \pi)$ the sequence of Fourier partial sums $f_N(x)$, converges to $\frac{1}{2}(f(x-) + f(x+))$, as n tends to ∞. Where, the notations f(x+) and f(x-) are the right-limit and left-limit respectively of f at the point x.

$$\text{where, } f_N(x) = \frac{a_0}{2} + \sum_{n=1}^{N} a_n \cos\left(\frac{2\pi nx}{P}\right) + \sum_{n=1}^{N} b_n \sin\left(\frac{2\pi nx}{P}\right) \tag{2.12}$$

Sum (2.12) is known as partial sum of Fourier series.

If function s(f) is defined by

$$s(f)(x) = \begin{cases} f(x) \ \textit{if } f \textit{ is continuous at } x \\ \dfrac{f(x-)\,f(x+)}{2} \ \textit{if } f \textit{ is discontinuous at } x \end{cases}$$

then from the above theorem,

$$\lim_{N\to\infty} f_N(x) = s(f)(x), \textit{for any } x \in [-\pi, \pi].$$

2.6.3 HARMONIC ANALYSIS

According to mathematical principles, any function which is defined at every point in the interval can be represented by an infinite series of sine and cosine functions. This series called a Fourier Series. The determination of a finite sum of sine and cosine terms is called "Harmonic Analysis "

Harmonic analysis which studies the representation of functions or signals as the superposition of basic waves. The basic waves are called 'harmonics', hence this analysis is known by "harmonic analysis."

Harmonic analysis is most common for the study of the periodical variations of the meteorological parameters. Harmonic analysis helps in the physical understanding of the regular fluctuations.

Harmonic analysis (HA) is used to apply for periodic variations (Achela,[2]) of the weather parameters and it is helpful to understand the regular fluctuations in the variable(Van wijk, [160]) In the case of meteorological data, observations are not continuous but exits only at discrete points Temperature observation made at equal spacing, every week, every hour, or may be given of a month. Now, if only a finite number of points exist in the interval to be analyzed, a finite number of sines and cosines terms will be able to estimate all the observations.

In harmonic analysis different harmonics are isolated and each can be treated as independent identity and each may have different physical meaning For example, it may be possible that the first harmonic of the diurnal pressure cycle may be diurnal heating by the sun, but the second harmonic may be caused by the Sun's tide-producing force. In many cases it is possible to account for the complete variation at once, but the individual harmonics can be explained. Harmonic analysis helps to give the boundary condition in the case of temperature study by certain differential equation. For example, the vertical variation of the daily temperature cycle can be expressed as a solution of the equation of heat conduction, where, the temperature variation at the surface is given in terms of harmonic analysis

Since the annual and diurnal variation of weather parameters like soil and air temperature have a tendency to interact with each other, diurnal variation is studied on the basis of observations during the one complete cycle. Therefore, to predict the soil temperature, Harmonic Analysis (HA) is applied. Liakatas[99], have pointed out that HA affords the most precise and accurate method.

2.7 DOUBLE FOURIER SERIES (DFS)

In this thesis we also employ Fourier series representation to predict the annual rainfall (ARF) in chapter 5 and hourly air temperature (HAT) in chapter 6.

First define an orthonormal to a system of continuous functions or square integrable functions defined on R, where R is a rectangle in the xy-plane

$$R : a \le x \le b, \quad c \le y \le d \qquad \text{The system} \quad \phi_n(x,y) \quad n = 0,1,2 \tag{2.3}$$

is said to be orthogonal if

$$\iint_R \phi_n(x,y)\,\phi_m(x,y)\,dx\,dy = 0;\ n \ne m.$$

Norm of $\phi_n(x,y)$ is defined by $\|\phi_n(x,y)\| = \sqrt{\iint_R \phi_n(x,y)^2 \ \ dx\,dy}$. (2.4)

The system (2.3) is said to be normal if

$$\|\phi_n(x,y)\| = 1;\ n=0,1,2$$

This implies that $\iint_R \phi_n(x,y)\,\phi_m(x,y)\,dx\,dy = 1$

2.7.1 FOURIER COEFFICIENTS (FCs)

One can associate a Fourier series with every absolutely integarble function f(x,y) defined on R. That is

$$f(x,y) \approx c_0\,\phi_0(x,y) + c_1\,\phi_1(x,y) + c_2\,\phi_2(x,y) + \ldots.. + c_n\ \phi_n(x,y) + \ldots., \tag{2.5}$$

where,

$$c_n = \frac{\iint_R f(x,y)\,\phi_m(x,y)\,dx\,dy}{\iint_R \phi_n^{\,2}(x,y)\,dx\,dy} = \frac{\iint_R f(x,y)\,\phi_m(x,y)dxdy}{\|\phi_n\|^2}$$

(2 6)

The quantities c_n given by (2 6) are known as the Fourier Coefficients (FCs) associated with the function f(x,y).

All the properties of complete systems proved for one variable is true for two variables also.

2.7.2 DOUBLE TRIGONOMETRIC FOURIER SERIES

$$\left.\begin{array}{l} 1, \cos mx, \sin mx, \cos ny, \sin ny.. \\ \cos mx \cos ny, \sin mx \cos ny, \\ \cos mx \sin ny, \sin mx \sin ny, ...m = 1,2...; n = 1,2..... \end{array}\right\} \quad (2\ 7)$$

form the basic trigonometric system in two variables. Each of these functions is of period P=2π in both x and y. The functions of the system (2.7) are orthogonal on the square K: $[a \le x \le a+2\pi, \quad b \le y \le b+2\pi \ ]$.

2.7.3 FOURIER COEFFICIECIENTS OF f(x,y)

Let f(x,y) be defined on K. Then,

$$A_{00} = \frac{\iint_K f(x,y)\, dx\, dy}{\|1\|^2} = \frac{\iint_K f(x,y)\, dx\, dy}{4\pi^2};$$

$$A_{m0} = \frac{\iint_K f(x,y) \cos mx\, dx\, dy}{\|\cos mx\|^2} = \frac{\iint_K f(x,y) \cos mx\, dx\, dy}{2\pi^2}; \ m = 1,2.$$

$$A_{0n} = \frac{\iint_K f(x,y) \cos ny\, dx\, dy}{\|\cos ny\|^2} = \frac{\iint_K f(x,y) \cos ny\, dx\, dy}{2\pi^2}; \ n = 1,2...$$

$$B_{m0} = \frac{\iint\limits_K f(x,y)\sin mx\ dx\,dy}{\|\sin mx\|^2} = \frac{\iint\limits_K f(x,y)\ \sin mx\,dx\,dy}{2\pi^2};\ m = 1,2...$$

$$B_{0n} = \frac{\iint\limits_K f(x,y)\sin my\ dx\,dy}{\|\sin my\|^2} = \frac{\iint\limits_K f(x,y)\ \sin my\,dx\,dy}{2\pi^2};\ y = 1,2...$$

and same way,

$$\left.\begin{aligned} a_{mn} &= \frac{1}{\pi^2}\iint\limits_K f(x,y)\cos mx\cos ny dx dy, \\ b_{mn} &= \frac{1}{\pi^2}\iint\limits_K f(x,y)\sin mx\cos ny dx dy, \\ c_{mn} &= \frac{1}{\pi^2}\iint\limits_K f(x,y)\cos mx\sin ny dx dy, \\ d_{mn} &= \frac{1}{\pi^2}\iint\limits_K f(x,y)\sin mx\sin ny dx dy \end{aligned}\right\};\ m = 1,2,3...\ and\ n = 1,2,3,\ .. \qquad (2\ 8)$$

with these FCs (2 8) Fourier Series of f(x,y) can be defined as below,

f(x,y) be defined on square K = { (x,y): $-\pi < x < \pi$, $-\pi < y < \pi$] $\subset R^2$, then

z=f(x,y)=

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\lambda_{mn}[a_{mn}\cos mx\cos ny + b_{mn}\sin mx\cos ny + c_{mn}\cos mx\sin ny + d_{mn}\sin mx\sin ny] \qquad (2\ 9)$$

where, $\lambda_{mn} = \begin{cases} \frac{1}{4} & for\ m = n = 0 \\ \frac{1}{2} & for\ m > 0,\ n = 0\ or\ m = 0,\ n > 0 \\ 1 & for\ m > 0,\ n > 0 \end{cases}$

Now,

f(x,y)=

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \lambda_{mn}[a_{mn} \cos mx \cos ny + b_{mn} \sin mx \cos ny + c_{mn} \cos mx \sin ny + d_{mn} \sin mx \sin ny].$$

implies that

$\lim_{\substack{m \to \infty \\ n \to \infty}} S_{mn}(x, y) = f(x, y)$ *Or for given* $\varepsilon \succ 0$ *, there exists a number N such that the inequality*

$|f(x, y) - S_{mn}(x, y)| \le \varepsilon$ *, holds for* $m \ge N, n \ge N$. *this is also applicable to every square*

$$[a \le x \le a + 2\pi, \quad b \le y \le b + 2\pi \;]$$

2.7.4 A NUMERICAL EXAMPLE ILLUSTRATING THE DFS METHOD

During the computation using Double Fourier Series , Fourier Coefficients are found from constructing a matrix equation which is given in chapter 5. There, we have used pseudoinverse of a matrix. This is defined as below

If A^+ is the unique matrix and said to be pseudoinverse of A which satisfies the following criteria:

1. $AA^+A = A$,
2. $A^+AA^+ = A^+$ (A^+ is a weak inverse for the multiplicative semigroup)
3. $(AA^+)^* = AA^+$ (That is, AA^+ is Hermitian)
4 $(A^+A)^* = A^+A$ (A^+A is also Hermitian)

Here M^* is the conjugate transpose of a matrix M whose elements are real.

Example 1:

Notations:

1) A(i,j)= Component of the i^{th} raw and j^{th} column of the matrix A

2) A(,j) = All the components of j^{th} column of the matrix A

Suppose that x =[2 3] , y=[4 1] and z=x+y.

Let m = 1, 2 and n =1, 2

Number of observation is 2.

Now, $sx = x'_{2x1} * m_{1x2} = \begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix}$ and $s_1y = y'_{2x1} * n_{1x2} = \begin{bmatrix} 4 & 8 \\ 1 & 2 \end{bmatrix}$;

$$aa = \begin{bmatrix} s1y & sx \end{bmatrix}_{2\times 4} = aa = \begin{bmatrix} 4 & 8 & 2 & 4 \\ 1 & 2 & 3 & 6 \end{bmatrix}$$

$$ac = \left[\frac{\cos(aa)}{2}\right]' \quad and \quad as = \left[\frac{\sin(aa)}{2}\right]'$$

$$ac = \begin{bmatrix} -0.3268 & 0.2702 \\ -0.0728 & -0.2081 \\ -0.2081 & -0.4950 \\ -0.3268 & 0.4801 \end{bmatrix} \qquad as = \begin{bmatrix} -0.3784 & 0.4207 \\ 0.4947 & 0.4546 \\ 0.4546 & 0.0706 \\ -0.3784 & -0.1397 \end{bmatrix}$$

$$cx = \cos(sx) \quad , \quad cx = \begin{bmatrix} -0.4161 & -0.9900 \\ -0.6536 & 0.9602 \end{bmatrix} ; \qquad cy = \begin{bmatrix} -0.6536 & -0.1455 \\ 0.5403 & -0.4161 \end{bmatrix}$$

$$ssx = \sin(sx)\ ,\quad \text{ssx} = \begin{bmatrix} 0.9093 & 0.1411 \\ -0.7568 & -0.2794 \end{bmatrix};$$

$$ssy = \sin(s1y)\ ;\quad ssy = \begin{bmatrix} -0.7568 & 0\ 8415 \\ 0.9894 & 0.9093 \end{bmatrix},$$

Now , let

$$r = \begin{bmatrix} \text{cx}(1,1).*\text{cy}(\ ,1) \\ \text{cx}(2,1)\ *\text{cy}(.,1) \end{bmatrix} = \begin{bmatrix} 0\ 2720 \\ -0\ 2248 \\ 0\ 4272 \\ -0\ 3531 \end{bmatrix};\qquad re = \begin{bmatrix} \text{cx}(1,2).*\text{cy}(\cdot,2) \\ \text{cx}(2,2).*\text{cy}(\ ,2) \end{bmatrix} = \begin{bmatrix} 0\ 1440 \\ 0\ 4119 \\ -0\ 1397 \\ -0\ 3995 \end{bmatrix},$$

$$r1 = \begin{bmatrix} \text{cx}(1,1).*\text{ssy}(:,1) \\ \text{cx}(2,1).*\text{ssy}(:,1) \end{bmatrix} = \begin{bmatrix} 0.3149 \\ -0.4117 \\ 0.4946 \\ -0.6467 \end{bmatrix};\qquad r1e = \begin{bmatrix} \text{cx}(1,2).*\text{ssy}(:,2) \\ \text{cx}(2,2).*\text{ssy}(:,2) \end{bmatrix} = \begin{bmatrix} -0\ 8331 \\ -0.9002 \\ 0.8080 \\ 0.8731 \end{bmatrix};$$

$$r2 = \begin{bmatrix} \text{ssx}(1,1).*\text{cy}(\cdot,1) \\ \text{ssx}(2,1).*\text{cy}(:,1)]; \end{bmatrix} = \begin{bmatrix} -0.5943 \\ 0.4913 \\ 0.4946 \\ -0.4089 \end{bmatrix},$$

r2e=[ssx(1,2) .*cy(:,2) ; ssx(2,2) *cy(:,2)]; r2e=[-0.0205;-0.0587;0.0407,0.1163];

r3 =[ssx(1,1) .*ssy(:,1) ; ssx(2,1) .*ssy(:,2)];r3=[-0.6882;0.8997;-0.6368;- 6882];

r3e= [ssx(1,2) .*ssy(:,2) ; ssx(2,2) *ssy(·,2)];r3e=[0.1187; 0.1283;-0.2351;-0.2541];

D=[iw1 ;ac;as; r re;r1 r1e;r2 r2e;r3 r3e]';

$$
= \begin{bmatrix}
0\,2500 & -0\,3268 & -0\,0728 & -0\,2081 & -0\,3268 \\
-0\,3784 & 0\,4947 & 0\,4546 & -0\,3784 & \\
0\,2720 & -0\,2248 & 0\,4272 & -0\,3531 & \\
0\,3149 & -0\,4117 & 0\,4946 & -0\,6467 & \\
-0\,5943 & 0\,4913 & 0\,4946 & -0\,4089 & \\
-0\,6882 & 0\,8997 & -0\,6368 & -0\,6882 & \\
 & & & & \\
0\,2500 & 0\,2702 & -0\,2081 & -0\,4950 & 0\,4801 \\
0\,4207 & 0\,4546 & 0\,0706 & -0\,1397 & \\
0\,1440 & 0\,4119 & -0\,1397 & -0\,3995 & \\
-0\,8331 & -0\,9002 & 0\,8080 & 0\,8731 & \\
-0\,0205 & -0\,0587 & 0\,0407 & 0\,1163 & \\
0\,1187 & 0\,1283 & -0\,2351 & -0\,2541 &
\end{bmatrix}_{2\times 25}
$$

u1=pinv(D)$_{25\,\mathrm{x}\,2}$; Coefficient matrix $c = \left(u_1 * \begin{bmatrix} 6 \\ 4 \end{bmatrix} \right)_{25\times 1}$; where,

$$u_1 = \begin{bmatrix} 0\ 0412 & 0\ 0519 \\ -0\ 0627 & 0\ 064 \\ -0\ 0100 & -0\ 0452 \\ 0\ 0301 & -0\ 1070 \\ -0\ 0658 & 0\ 1116 \\ -0\ 0742 & 0\ 0992 \\ 0\ 0822 & 0\ 0938 \\ 0\ 0807 & 0\ 0090 \\ -0\ 0659 & -0\ 0255 \\ 0\ 0468 & 0\ 0280 \\ 0\ 0465 & 0\ 0949 \\ 0\ 0788 & -0\ 0374 \\ -0\ 0576 & -0\ 0836 \\ 0\ 0689 & -0\ 1899 \\ -0\ 0607 & -0\ 1941 \\ 0\ 0770 & 0\ 1724 \\ -0\ 1291 & 0\ 2037 \\ -0\ 1065 & 0\ 0042 \\ 0\ 0892 & -0\ 0203 \\ 0\ 0883 & 0\ 0017 \\ -0\ 0752 & 0\ 0319 \\ 0\ 1254 & 0\ 0366 \\ 0\ 1598 & 0\ 0152 \\ -0\ 1110 & -0\ 0429 \\ -0\ 1199 & -0\ 0463 \end{bmatrix}_{25\times 2}$$

$$c = \begin{bmatrix} 0\ 4551 \\ -0\ 1167 \\ -0\ 2409 \\ -0\ 6086 \\ 0\ 0514 \\ -0\ 0488 \\ 0\ 8684 \\ 0\ 5199 \\ -0\ 4976 \\ 0\ 3926 \\ 0\ 1008 \\ 0\ 3235 \\ -0\ 6800 \\ -0\ 3463 \\ -1\ 1407 \\ 1\ 1514 \\ 0\ 0402 \\ -0\ 6222 \\ 0\ 4538 \\ 0\ 5368 \\ -0\ 3237 \\ -0\ 6064 \\ 1\ 0198 \\ -0\ 8375 \\ -0\ 9049 \end{bmatrix}_{25\times 1}$$

This coefficient matrix c is used for prediction of the rainfall with required input matrices x and y.

Now, $z' = c^{t} * D = \begin{bmatrix} 6 \\ 4 \end{bmatrix} = z$

Example 2: To predict the function f (x,y) = x^2+ y^2

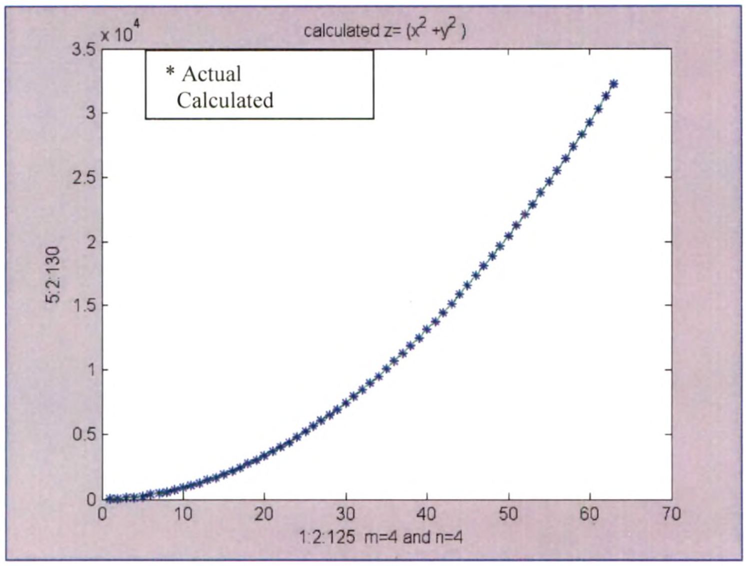


Fig: 2.4

Example 3: To predict the function f(x,y) =-3y/(x^2 +y^2+1)

Programme is given in the Appendix. Predicted values are plotted in the following graph.

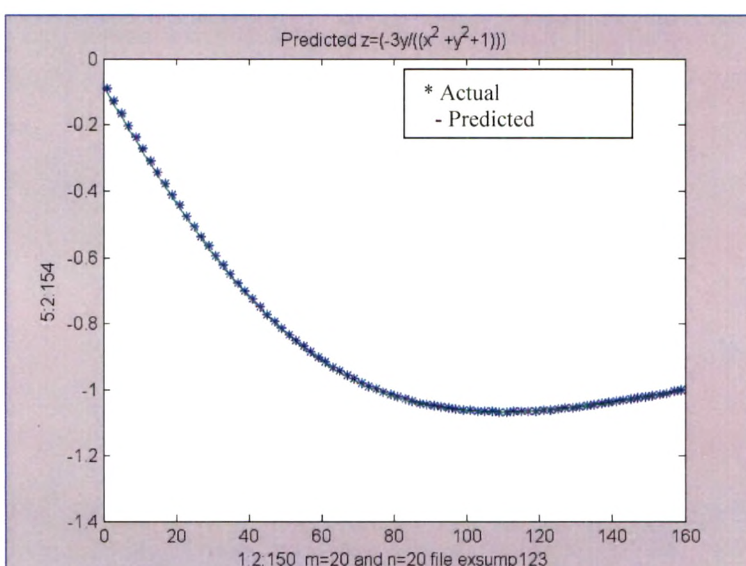


Fig: 2.5

❖