Convolution of Hyperplanes with Gaussian Kernals in a Differentially Fed Artificial Neural Network

Bangalore University
Manjunath.R, Dr K.S.Gurumurthy
Dept of EC & CSE, UVCE, Bangalore, INDIA
manju_r_99@yahoo.com

Abstract: The problem of function approximation to map a set of data with another is challenging When they are non linearly related. A variety of solutions based on Neural Networks are found in the literature. In this paper, a new class of Artificial neural networks with differential feedback are introduced. The different orders of differentials form a manifold of hyperplanes [2]. The spectral properties of these hyperplanes are explored

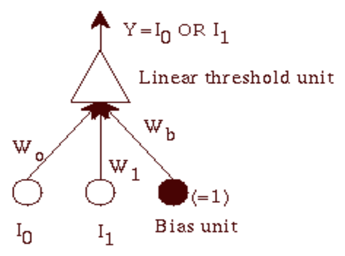


Figure.1. Artificial neural network

1 INTRODUCTION

Artificial neural network systems (ANN) are extremely helpful in modeling a system whose behaviour is unknown but for the enough non-linearly related input and output data.The structure of a simple Artificial neural network with two inputs is shown in the figure.1.The network has 2 inputs, and one output. All are binary.

The output is

1 if $W_0 * I_0 + W_1 * I_1 + W_b > 0$

0 if $W_0 * I_0 + W_1 * I_1 + W_b <= 0$ (1)

The above network can learn Logical OR: output a 1 if either I_0 or I_1 is 1. The network adapts by changing the weight by an amount proportional to the difference between the desired output and the actual output.

$\Delta w = \xi.(\text{error}).i$.(2)

here. i is the input driving the node.,ξ is the learning rate, The network functions as follows: Each neuron receives a signal from the neurons in the previous layer, andeach of those signals is multiplied by a separate weight value. The weighted inputs are summed, and passed through a limiting function (non linear in general) which scales the output to a fixed range of values.

A BP network learns by example. That is, a learning set has to be provided that consists of some input examples and the known-correct output for each case, like a look-up table. The network adapts.

The BP learning process works in small iterative steps: First, one of the example cases is applied to the network, and the network produces some output based on the current weights (initially, the output will be random). This output is compared to the known-good output, and a mean-squared error signal is calculated. The error value is then propagated backwards through the network, and small changes are made to the weights in each layer.(proportional to the error). The whole process is repeated for each of the example cases, then back to the first case again, and so on. The cycle is repeated until the overall error value drops to a predefined value. In section II differential feedback concept is developed.. In section III The spectral property of the hyperplanes is explained. The experimental results are demonstrated in section IV section V concludes

2. DIFFERENTIAL FEEDBACK METHOD

One of the major drawbacks of the conventional training methodology is that it is iterative in nature and takes a large number of cycles to converge to the prespecified error limit. By intuition, if more information is made to be hidden with the data, it takes less No. of iterations to get stabilized to the pre defined error limit. Auto regressive Moving average (ARMA) model can be conveniently used in this direction. In a typical ARMA model, the output and input are related by

y(n)=a0*x(n)+a1*x(n-1)+…+b1*y(n-1)+… .(3)

The differentials of the output (& input) can be written in to similar linear combinations of present & previous outputs (inputs).

(dy/dt)=y(n+1)-y(n) . (4)

By (3) and (4) it is clear that

y(n)=f(x(n),x(n-1),..,dy/dtt=n,d^2y/dt^2t=n,) (5)

where the ANNs are made to learn this function. Here the ARMA output is subjected to non-linear function for the same reason as in ANN which makes it possible to learn non linear functions. The output y of a neural network but for the nonlinearities can be written as

y=Σwixi. (6)

Where xi are the inputs wi, the corresponding weights. The space spanned by weight vector for different inputs is a hyperplane. The important factor is weight cannot span the entire input space [1], whatever may be the training mode. Again the linearity of the output (1) may be viewed as a particular case of ARMA

y (n+1)=b0y(n)+b1y(n-1)+…..+a0xn+… (7)

Where b0.. and a0.. are constants. The auto regressive terms b0…bn may be realized using an implied differential feedback [3]. With differential feedback it has been found out[3] that the no of iterations required for training is reduced as shown in the table I.XOR gate is considered for simulation. Gaussian distributed random input with seed value 1000 is taken as input. With I order different feedback, the output may be written as:
Σwixi +b1y1 (8)

y1 being the I order differential. This equation once again represents a plane parallel to Σwixi. Thus the set of differentially fed ANNs form a manifold of parallel planes, with ∞ order feedback being the plane with zero error.Also, simulation results of table II show that two terms of II order differential feedback i.e., y2-y1 and y1-y0 can be replaced by a single equivalent plane represented by

Weq=(w1*iextra+w2*iextra1)/y0 (9)

In II order differential feedback system, the two differential terms can be replaced by a single term. Extending this principle, the ∞ terms of ∞ order differential feedback can be replaced by a single term. This is termed as eigen plane which is the practical way of generating lowest error. Now the differential feedback becomes

dy/dt+d^2y/dt^2+… (10)

Taking Z transform, & then the inverse,

yeq=IZT{Y(z)/(1-z)} (11)

3.CONVOLUTION OF HYPERPLANES

Let yk represent the kth hyperplane corresponding to kth order of differential feedback. Hence yk= a*y0+b*y1+…
=b1*yk-1+d/dtn(y0) c1 being a constant Here the incremental portion yΔ=c1*d/dt(0) will be approximated as the convolution of some function with y0.

Ie yΔ=y0*f (12)

F may be found out by pushing both sides of 12 to frequency domain. k*y(k)=y(k)*F(f) k being the frequency index of DFT.Thus F(f) should have a linear response over a range of indices as shown in fig 2.Its equivalent time domain signal may be expressed as. F=(sin**2(ax)/Ax *x)*exp-(Nt/2) since the triangle is delayed by N/2.N being the no. of points in DFT. The signal is shown in fig 3. As evident from the figure, over a very small duration, the result may be approximated to a Gaussian pulse of scale factor l where the Gaussian kernel with dilation parameter l is given by G(l,x) an exponential function. Analytically F decays as $1/t^2$ =(1-t*t/k) . A Gaussian pulse exp(-t*t/k) also may be approximated as (1-t*t/k).Hence, the envelop of f is Gaussian if the peaks are too close ie sinc function function is large.

The table I shows that the gap between the hyperplanes decreases with increase in the order. I.e. the information content becomes more and more abstract. Hence the kernel which is convolved with the output y is scaled wth progressively increasing scale factor.

3.1.Working of the model

In the domain of learning, mixtures of Gaussians is a powerful tool for statistical modeling. Such a model can avoid the problem of overfitting if care is taken. For this class of generative models a complexity control scheme is presented in [4], to provide an effective means for avoiding the problem of over-fitting .The model calls for decrease in the variance at higher levels of resolution..

This constraint on the model is satisfied with higher order differential feedback. Table I shows that the error or variance gets reduced for a given number of iterations if differential feed back is given. The initialization problem [5], encountered with the fitting of Gaussian mixture models is remedied by choosing the zeroth ordered hyperplane as the initial approximation and giving more feedback to lift it towards the eigen plane or maximum likelihood estimator described in section I. It gradually deforms the objective function via deterministic annealing and progressively tracks the maximum likelihood solution across the decreasing noise levels as required by the model. The differential feedback architecture goes well with the model described in [4]. There the objective of model fitting is to exploit some subspace structure within the data, which often results from their concentration near some manifold of lower dimensionality. This gives rise to the following generative model

$$x=f(z)+u \qquad (13)$$

With an affine mapping F(z)=Az+μ.All components of x are linear combinations of Gaussian random variables and thus the observable can be characterized by a multivariate normal density

$$g(x|\phi)=(2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} \qquad (14)$$

Since convolution with Gaussian function may be expressed as a linear combination of scaled and shifted Gaussians, the hyperplanes are expressible as a linear combination of Gaussian variables. Also, in sec II it has been observed that the hyperplanes are linearly transported.In order to reduce the complexity of such a model, the covariance model $\Sigma=\psi+2I\sigma$ is proposed[4].Thus the model has two parts- a data dependent signal part and a noise part which corresponds to the fixed part and differential parts of the proposed model. In the model in [4] as the variance reduces, more and more subspaces in the data become visible. This is true with the proposed model where outputs corresponding to higher levels include lower level planes.

To summarize, with differential feedback, the complexity of the model is automatically controlled making it resistant for overfitting and makes the observable data Gaussian.

4. SIMULATION

The differentially fed Artificial neural networks are made to learn the psd of random data .The Normal distributed data is generated using Matlab. The error after learning and the differentials of the error are stored. The normalised PSDs of convolution of outputs without feedback and Gaussian pulse and normalized PSD of the first derivatives with feedback are shown in the fig.4..It may be seen that the derivatives are formed by the convolution of the output with Gaussian functions of different scales

5. CONCLUSIONS

Differential feedback, when applied over a neural networks leads to a manifold of affinely transported hyperplanes. These hyper planes are actually formed by the convolution of the non feedback output with Gaussian kernels of different scales.

REFERENCES

1. S.Amari, 1995, Information Geometry of the EM and em algorithms for neural networks, Neural networks, 8,No.9

2. 7.S.Amari and H.Nagaoka, 2000, Methods of information geometry, AMS and Oxford University press.

3. Manjunath.R and K.S.Gurumurthy, Oct 2002, System design using differentially fed Artificial Neural networks, TENCON'02

4. Peter Meinicke and Helge Ritter. , 1999, Resolution based complexity control for Gaussian mixture models. Technical report, Faculty of Technology, University of Bielefeld.

5. McKenzie, P. and Alder. M. , 1994, Initializing the EM algorithm for use in Gaussian mixture modelling. In Gelsema, E. S. and Kanal, L. N., editors, *Pattern Recognition in Practice IV*, pages 91–105. Amsterdam:Elsevier.

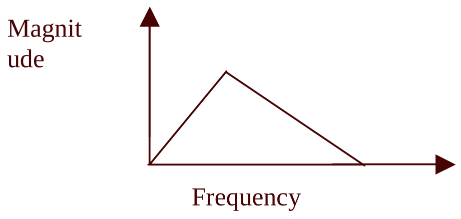
.

Table 1.Performance with feedback

Order of differential	Square error	Iterations
No feedback	18	1156
I order	18	578
II order	18	289

Table 2..Performance with II order feedback

Order of differential	Square error	Iterations
II order Feedback	18	578
Equivalent Output feedback	18	578



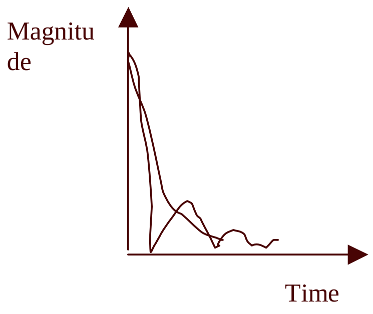
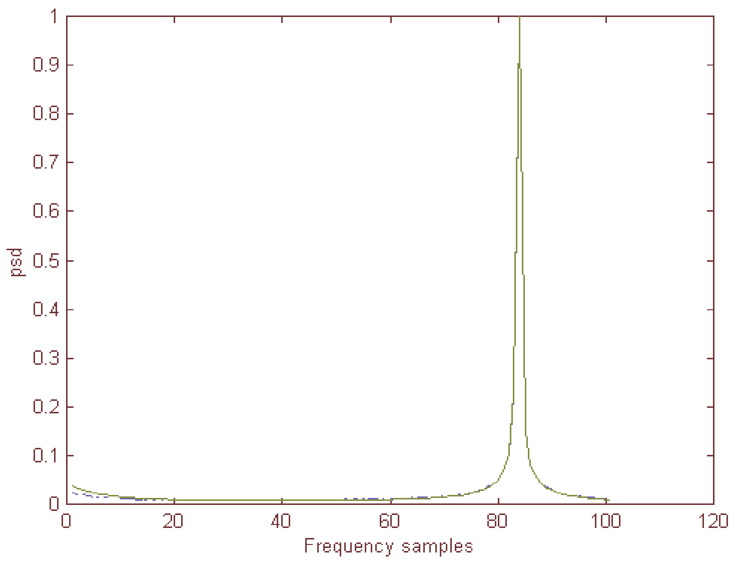


Figure 2.Desired linear response

Figure.3. The time domain plot



Legend: - PSD of non FB signal convolved with Gaussian pulse . –. PSD of the I order FB
Figure.4. PSDs of Differential signals and the convolved signals