ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 237-244

Evaluating the impact of multiplicative input perturbations on radial basis function networks

J.L.Bernier, J.González, A.Cañas, A.F.Díaz, F.J.Fernández, J.Ortega

Dpto. Arquitectura y Tecnología de Computadores
Universidad de Granada, E18071-Granada (Spain)

Abstract. In previous papers we have introduced the Mean Squared Sensitivity (MSS) as an approximation of the performance degradation of a MLP or a RBFN affected by perturbations in different parameters. In the present paper, we consider that the inputs of a RBF network are affected by noise, using a multiplicative model for such perturbations. We have obtained the corresponding analytical expression for MSS and have validated it experimentally. MSS is proposed as a quantitative measurement for evaluating the noise immunity of a RBFN configuration against multiplicative noise.

1 Introduction

Radial Basis Function Networks (RBFNs) [1] are neural paradigms that are currently receiving a great deal of interest and which can be considered universal approximators [2]. Nevertheless, the algorithms used to train them provide solutions that correspond to local optima in the space of parameter configurations [2]. In this way, for a fixed structure of the network, a modification in the constraints used during training would produce different configurations for the RBFN (different values of weights, centres and radii). These different configurations may present a similar response with respect to learning, i.e., they may present a similar Mean Squared Error (MSE). However, their performance with respect to noise immunity or generalization ability may present large differences. Thus, two different configurations of RBFNs that solve the same problem may present a similar Mean Squared Error (MSE) computed over the same set of input patterns, but if these patterns are altered by noise, the performance of the two configurations may be degraded to a different degree. The training algorithms are based on searching local minima of the mean squared error with respect to the parameters of the network, but there exist many possibilities (as many as there are local minima) and some of these minima are flatter than others [3].

In [4] a suitable approximation of the MSE degradation of a Multilayer Perceptron (MLP) subject to perturbations, called Mean Squared Sensitivity

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 237-244

(MSS), was obtained. MSS measures the MSE degradation of the MLP in the presence of deviations and its expression can be particularized to consider different types of weight or input perturbations. The same methodology was used in [5] to study the case of RBFNs subject to additive input deviations. In this work it is considered that the input deviations that affect a RBFN have a multiplicative nature. We obtain the expression of MSS for such kind of deviations and show the validity of the approximation. Therefore, we propose the use of MSS as a measure of performance degradation of a RBFN affected by multiplicative input noise, providing a useful criterion to select between different RBFN configurations, which is interesting, for example, when the inputs of the network come from analogue sensors.

The paper is organized as follows: in Section 2, the particular expression of statistical sensitivity to multiplicative perturbations is derived. The relationship between statistical sensitivity and MSE degradation is presented in Section 3, where MSS is defined. Section 4 shows the experimental results that enable us to demonstrate the validity of the expressions obtained and, finally, Section 5 draws some conclusions.

2 The statistical sensitivity of a RBF network to multiplicative deviations

Without loss of generality, let us consider a RBF network consisting of n inputs, a single output, and m neurons in the hidden layer. The output of this network is then computed as the averaged sum of the outputs of the m neurons, where each neuron is a radial function of the n inputs to the network:

$$y = \sum_{i=1}^{m} w_i \Phi_i = \sum_{i=1}^{m} w_i \exp\left(- \frac{\sum\limits_{k=1}^{n} (x_k - c_{ik})^2}{r_i^2} \right) \tag{1}$$

where x_k (k=1,...,n) are the inputs to the network, and c_{ik} and r_i are the centres and radius of the RBF associated with neuron i, respectively.

If the inputs presented to the RBF are perturbed by noise, then the output y of the network is changed with respect to its nominal output. The statistical sensitivity, S, enables us to estimate in a quantitative way the degradation of the expected output of the RBF network when the values of the inputs change by a given amount. Statistical sensitivity is defined in [7] by the following expression:

$$S = \lim_{\sigma \to 0} \frac{\sqrt{var(\Delta y)}}{\sigma} \tag{2}$$

where σ represents the standard deviation of the changes in the inputs, and $var(\Delta y)$ is the variance of the deviation in the output (with respect to the output in the absence of perturbations) due to these changes, which can be

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 237-244

computed as: $var(\Delta y) = E[(\Delta y)^2] - E[\Delta y]$, with $E[\cdot]$ being the expected value of $[\cdot]$.

If the deviations considered are small enough, then the corresponding deviation in the output of the network can be approximated as:

$$\Delta y \approx \sum_{k=1}^{n} \frac{\partial y}{\partial x_k} \Delta x_k \tag{3}$$

To compute expression (2), we assume a multiplicative model of input deviations that satisfies:

(a). $E[\Delta x_k] = 0$

(b). $E[(\Delta x_k \Delta x_l)] = \sigma^2 x_k x_l \delta_{kl}$

where δ_{kl} is the Kronecker delta. This perturbation model implies that each input x_k is randomly modified in a quantity proportional to its nominal value. It is assumed that these perturbations follow a normal distribution with average equal to zero and standard deviation equal to σ. Moreover, perturbations of different inputs are assumed not to be statistically correlated.

Proposition 1: if $E[\Delta x_k] = 0\ \forall k$ then $E[\Delta y] = 0$.

Proof 1:

$$\begin{aligned} E[\Delta y] &= E[\sum_{k=1}^{n} \frac{\partial y}{\partial x_k} \Delta x_k] = E[\sum_{k=1}^{n} \frac{\partial}{\partial x_k} \left(\sum_{i=1}^{m} w_i \Phi_i \right) \Delta x_k] \\ &= E[\sum_{k=1}^{n} \sum_{i=1}^{m} w_i \frac{\partial \Phi_i}{\partial x_k} \Delta x_k] = 2 \sum_{i=1}^{m} \frac{w_i \Phi_i}{r_i^2} \sum_{k=1}^{n} (c_{ik} - x_k)\, E[\Delta x_k] \\ &= 0 \end{aligned} \tag{4}$$

□

Proposition 2: the statistical sensitivity to multiplicative input perturbations of a RBF network can be expressed as:

$$S = 2 \sqrt{\sum_{k=1}^{n} \left(x_k \sum_{i=1}^{m} \frac{w_i \Phi_i}{r_i^2} (c_{ik} - x_k) \right)^2} \tag{5}$$

Proof 2:

$$\begin{aligned} E\left[(\Delta y)^2\right] = E &\left[\left(2 \sum_{i=1}^{m} \frac{w_i \Phi_i}{r_i^2} \sum_{k=1}^{n} (c_{ik} - x_k)\, \Delta x_k \right) \right. \\ &\left. \left(2 \sum_{j=1}^{m} \frac{w_j \Phi_j}{r_j^2} \sum_{l=1}^{n} (c_{jl} - x_l)\, \Delta x_l \right) \right] \end{aligned}$$

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 237-244

$$
\begin{aligned}
&= 4E\left[\sum_{i=1}^{m}\frac{w_i\Phi_i}{r_i^2}\sum_{j=1}^{m}\frac{w_j\Phi_j}{r_j^2}\sum_{k=1}^{n}(c_{ik}-x_k)\sum_{l=1}^{n}(c_{jl}-x_l)\Delta x_k\,\Delta x_l\right]\\
&= 4\sum_{i=1}^{m}\frac{w_i\Phi_i}{r_i^2}\sum_{j=1}^{m}\frac{w_j\Phi_j}{r_j^2}\sum_{k=1}^{n}(c_{ik}-x_k)\,(c_{jk}-x_k)\sigma^2x_k^2\\
&= 4\sigma^2\sum_{k=1}^{n}\left(\sum_{i=1}^{m}\frac{w_i\Phi_i}{r_i^2}(c_{ik}-x_k)\right)\left(\sum_{j=1}^{m}\frac{w_j\Phi_j}{r_j^2}\,(c_{jk}-x_k)\right)x_k^2\\
&= 4\sigma^2\sum_{k=1}^{n}\left(x_k\sum_{i=1}^{m}\frac{w_i\Phi_i}{r_i^2}(c_{ik}-x_k)\right)^2 \qquad (6)
\end{aligned}
$$

Substituting (6) in (2), Proposition 2 is proved.

□

3 The Mean Squared Sensitivity

It is usual to measure the learning performance of a RBF network using the Mean Squared Error (MSE). This error measurement is computed by the sum of a set of input patterns whose desired output is known, and its expression is the following:

$$MSE = \frac{1}{N_p}\sum_{p=1}^{N_p}\varepsilon(p) = \frac{1}{2N_p}\sum_{p=1}^{N_p}(d(p)-y(p))^2 \qquad (7)$$

where N_p is the number of input patterns considered, and $d(p)$ and $y(p)$ are the desired and obtained outputs for the input pattern p, respectively.

If the inputs of the network suffer any deviation, the nominal output is altered, as is the expected MSE. By developing expression (7) with a Taylor expansion near the nominal MSE found after training, MSE_0, it is obtained that:

$$MSE' = MSE_0 - \frac{1}{N_p}\sum_{p=1}^{N_p}(d(p)-y(p))\Delta y(p) + \frac{1}{2N_p}\sum_{p=1}^{N_p}(\Delta y(p))^2 + 0 \qquad (8)$$

Now, if the expected value of MSE' is computed, taking into account the perturbation model adopted that $E[\Delta y] = 0$, and that from expressions (2) and (3) we obtain $E[(\Delta y)^2] \simeq \sigma^2 S^2$, the following expression is deduced:

$$E[MSE'] = MSE_0 + \frac{\sigma^2}{2N_p}\sum_{p=1}^{N_p}(S(p))^2 \qquad (9)$$

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 237-244

Table 1: MSE_0 and MSS obtained after training

Problem	MSE_0	MSS
Mackey-Glass	$6.65\ 10^{-5}$	1.095
f_8 function	$8.41\ 10^{-4}$	5.126

By analogy with the definition of MSE, we define "Mean Squared Sensitivity "(MSS) as:

$$MSS = \frac{1}{2N_p}\sum_{p=1}^{N_p}(S(p))^2 \tag{10}$$

MSS is evaluated from the statistical sensitivities for a set of input patterns, as expression (10) shows and from the nominal values of the network parameters. By combining expressions (9) and (10), the expected degradation of the MSE, $E[MSE']$ can be expressed in a simplified way as:

$$E[MSE'] = MSE_0 + \sigma^2 MSS \tag{11}$$

Thus, as MSE_0 and MSS can be directly computed after training, it is possible to predict the degradation of MSE when the inputs of the network are deviated from their nominal values into a determinate range. Moreover, as can be deduced from (11), a lower value of MSS implies a lower degradation value of MSE; thus, we propose using MSS computed using (5) as a suitable measure of the noise immunity of RBF networks against multiplicative deviations.

4 Results

In order to validate the expressions presented, we compared the results experimentally obtained for $E[MSE']$ when the inputs of the network are affected by multiplicative deviations with the predicted value obtained by using (11).

Two RBFNs were considered: a predictor of the Mackey-Glass temporal series[8] and an approximator of the f_8 function proposed by Cherkassky et al. [9]. The Mackey-Glass predictor consisted of 4 inputs, 14 RBFs in the hidden layer and 1 output neuron, while the approximator consisted of 2 inputs, 16 RBFs and 1 output. Table 1 shows the values of MSE_0 and MSS obtained after training using the test patterns (different from those used for training).

The inputs of the networks were randomly deviated from their nominal values considering different values of σ and a multiplicative model of deviation. Each value of $E[MSE']$ was experimentally computed over 50 tests. In each test, all the inputs of the network x_k were perturbed taking a value equal to $x_k(1+\delta_k)$ with δ_k being a random variable that follows a gaussian distribution with average equal to zero and standard deviation equal to σ. Note that this kind of perturbation is standard when modelling the tolerance margin of analogue elements.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 237-244

Figures 1 and 2 represent the experimental and predicted values of $E[MSE']$. The experimental values are plotted with their respective confidence levels at 95%. It can be observed that the values predicted by expression (11) accurately fit those obtained experimentally until the degradation of MSE becomes large. For example, we deduce from Figure 2 that MSE for σ equal to 0.05 is 16 times larger than MSE_0. Note that we have assumed small perturbations, and thus the approximation is worse when σ increases.

Thus, the validity of the approximation is demonstrated and, according to (11), this means that the lower the MSS the lower the MSE degradation in the presence of multiplicative noise. Thus, among RBF network configurations that present a similar MSE_0, the one with the lowest MSS provides the most stable output when its inputs are perturbed.

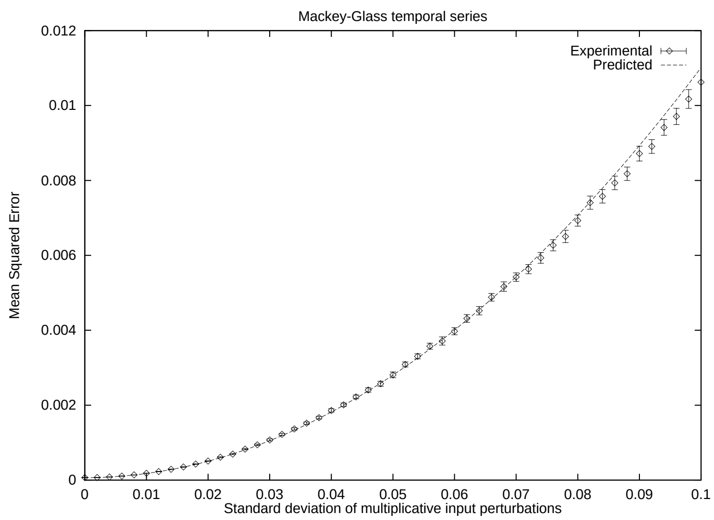

Figure 1: Experimental and predicted MSE for the Mackey-Glass temporal series

5 Conclusions

We have derived and validated a quantitative measure of noise immunity against multiplicative perturbations of RBF networks. This measure, which we term Mean Squared Sensitivity (MSS), is explicitly related to the MSE degradation in the presence of multiplicative input perturbations. This relationship shows that a lower value of MSS implies a lower degradation of MSE.

However, the analytical expression of MSS can be used as a regularizer during the training process in order to improve the final performance of the network with respect to noise immunity in the same way as was done for MLPs in [6].

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 237-244

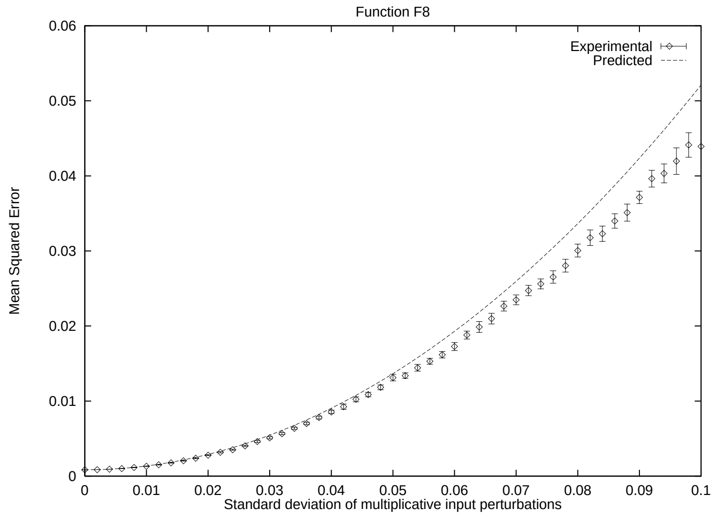

Figure 2: Experimental and predicted MSE for the F8 approximator

Acknowledgements

This work has been supported in part by projects CICYT TIC99-0550 and CICYT TIC2000-1348.

References

[1] D.S. Broomhead and D. Lowe, "Multivariate Functional Interpolation and Adaptive Networks", *Complex Systems* **2**, 321-355.

[2] S. Haykin, *Neural Networks. A Comprehensive Foundation*, (second edition, Prentice Hall, 1999).

[3] H. Hochreiter and J. Schmidhuber, "Flat Minima", *Neural Computation* **9**, 1-42 (1997).

[4] J.L. Bernier, J. Ortega, E. Ros, I. Rojas and A. Prieto, "A Quantitative Study of Fault Tolerance, Noise Immunity and Generalization Ability of MLPs", *Neural Computation* **12**, 2941-2964 (2000).

[5] J.L. Bernier, J. González, A. Cañas and J. Ortega, "Assessing the Noise Immunity of RBFNNs".*Lecture Notes in Computer Science* **2085**, 136-143 (2001).

[6] J.L. Bernier, J. Ortega, I. Rojas, E. Ros and A. Prieto, "Obtaining Fault Tolerant Multilayer Perceptrons Using an Explicit Regularization".*Neural Processing Letters* **12** (2), 107-113 (2000).

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 237-244

[7] J.Y. Choi and C. Choi, "Sensitivity Analysis of Multilayer Perceptron with Differentiable Activation Functions", *IEEE Trans. on Neural Networks* **3** (1), 101-107 (1992).

[8] M.C. Mackey and L. Glass, "Oscillation and Chaos in Physiological Control Systems", *Science* **197**, 287-289 (1977).

[9] V. Cherkassky, D. Gehring and F. Mulier, "Comparison of Adaptive Methods for Function Estimation from Samples", *IEEE Trans. Neural Networks* **7**(4), 969-984 (1996).