

Seeking Qualitative and Quantitative Robustness in Artificial Neural Networks

Héctor Allende^a, Claudio Moraga^{b,c} and Rodrigo Salas^a

a) Universidad Técnica Federico Santa María; Dept. de Informática;
Casilla 110-V; Valparaíso-Chile; {hallende, rsalas}@inf.utfsm.cl

b) Technical University of Madrid; Dept. Artificial Intelligence.
E-28660 Boadilla del Monte Madrid; Spain.

c) University of Dortmund; Department of Computer Science;
D-44221 Dortmund; Germany; moraga@cs.uni-dortmund.de

Abstract. In this paper, we discuss the local robustness of the most popular and general feedforward artificial neural networks (FANN). Robustness of ANNs is an important characteristics for many applications, ranging from simple function approximation to more complicated nonlinear control problems. The robustness of the system can be define as the soften of the effect of errors in the input space on the output system. In particular, the accuracy of the estimates in the presence of outlying observations is an important goal so we investigate the qualitative and quantitative robustness properties of the parameter estimation known as Learning process of the FANN. The robust learning process is based on the generalized maximum likelihood (GM) type estimators. The sensitivity to outliers is demonstrated based on synthetic data set.

Keywords: Artificial Neural Models; Robustness; Efficiency bounds; Adaptive Estimator

This paper should be consideres in **Paper Track**

Conference Topics

1. Evolutionary Computation, Genetic Algorithms and Neural Networks
2. Machine Learning, Knowledge Discovery and Data Mining

Seeking Qualitative and Quantitative Robustness in Artificial Neural Networks¹

Héctor Allende^a, Claudio Moraga^{b,c} and Rodrigo Salas^a

a) Universidad Técnica Federico Santa María; Dept. de Informática;
Casilla 110-V; Valparaíso-Chile; {hallende,rsalas}@inf.utfsm.cl

b) Technical University of Madrid; Dept. Artificial Intelligence.
E-28660 Boadilla del Monte Madrid; Spain.

c) University of Dortmund; Department of Computer Science;
D-44221 Dortmund; Germany; moraga@cs.uni-dortmund.de

Abstract. In this paper, we discuss the local robustness of the most popular and general feedforward artificial neural networks (FANN). Robustness of ANNs is an important characteristics for many applications, ranging from simple function approximation to more complicated nonlinear control problems. The robustness of the system can be define as the soften of the effect of errors in the input space on the output system. In particular, the accuracy of the estimates in the presence of outlying observations is an important goal so we investigate the qualitative and quantitative robustness properties of the parameter estimation known as Learning process of the FANN. The robust learning process is based on the generalized maximum likelihood (GM) type estimators. The sensitivity to outliers is demonstrated based on synthetic data set.

Keywords: Artificial neural Models; Robustness; Efficiency bounds; Adaptive Estimator

1 Introduction

Much of the recent work on robust ANN brings to mind research in nonparametric statistical inference. The ANN are seen by researchers as either highly parametrized model or nonparametric structures. In particular we can characterize the ANN as “semiparametric” whenever we relax some of the assumptions about the activation functions or the noise distribution corrupting the data measurements. With the term semiparametric NN (SNN) (See [Cap97]) it is thus recognized that both parametric and nonparametric model components are simultaneously present in the net architecture. Therefore, there are components that belong to finite-dimensional parameter spaces, e.g. weights, and components that range in function spaces, i.e. the same activation functions.

¹ This work was supported in part by Research Grant Fondecyt 1010101, in part by Research Grant BMBF CHL-99/023 (Germany) and in part by Research Grant DGIP-UTFSM
Work of C. Moraga was supported by Grant SAB2000-0048 of the Ministry of Education,
Culture and Sport (Spain) and the Social Fund of the European Union.

A typical nonparametric inference problem is the learning (or estimating in statistical language) of arbitrary decision boundaries for classification task, based on a collection of labeled (pre-classified) training samples. The boundaries are arbitrary in sense that no particular structure, or class of boundaries, is assumed a priori. In particular, there is no parametric model, as there would be with the presumption of, say, linear or quadratic decision surface. This point of view is implicit in many networks formulations suggesting a close analogy to nonparametric inference.

Nonparametric robust inference is an important topic in statistics and also artificial neural networks (ANN). The latter are seen by researchers as either highly parametrized models or nonparametric structures. In particular we characterize the networks as semiparametric models. With the term semiparametric neural networks it is thus recognized that both parametric and nonparametric model components are simultaneously present in the net architecture. We consider one crucial aspect for semiparametric inference that is the possible degradations that is brought into the estimates by the presence of outliers in the sample. This aspect has been widely investigated in statistics, but not in the ANN literature.

2 Artificial Neural Networks

Mathematically, an ANN consists of elementary processing elements (neurons), organized in Layers. The Layers between the input and the output layers are called "hidden". The number of input units is determined by the application. The architecture or topology of a network refers to the topological arrangement of the network connections. We define a class of neural models according to [Vap95]. A class of neural models is specified by

$$S_\lambda = \{g_\lambda(\underline{x}, \underline{w}), \underline{x} \in R^m, \underline{w} \in W\}, \quad W \subseteq \Re^p \quad (1)$$

where $g_\lambda(\underline{x}, \underline{w})$ is a non-linear function of \underline{x} with \underline{w} being its parameter vector, and p is the number of free parameters determined by A_λ , i.e., $p = p(A_\lambda)$. A_λ denotes the architecture of the neural network and λ the number of hidden neurons.

A class (or family) of neural models is a set of ANN models which share the same architecture and whose individual members are continuously parameterized by the vector $\underline{w} = (w_1, w_2, \dots, w_p)^T$. The elements of this vector are usually referred to as weights. For a single-hidden-layer architecture, the number of hidden units λ indexes the different classes of ANN models (S_λ) since it is an unambiguous descriptor of the dimensionally p of the parameter vector ($p = (m + 2)\lambda + 1$).

Given the sample of observations, the task of neural learning is to construct an estimator $g(\underline{x}, \underline{w})$ of the unknown function $\varphi(\underline{x})$

$$g(\underline{x}, \underline{w}) = \gamma_2 \left(\sum_{j=1}^{\lambda} w_j^{[2]} \gamma_1 \left(\sum_{i=1}^m w_{ij}^{[1]} x_i + w_{m+1,j}^{[1]} \right) + w_{\lambda+1}^{[2]} \right) \quad (2)$$

where $\underline{w} = (w_1, w_2, \dots, w_p)^T$ is a parameter vector to be estimated, γ'_s are linearity or non-linearity and λ is a control parameter (number of hidden units). An important factor in the specification of neural models is the choice of base function γ . Otherwise known as 'activation' or 'squashing' functions, these can be any non-linearity as long as they are continuous, bounded and differentiable. Typically γ_1 is a sigmoidal or the hyperbolic tangent. All these functions belong to the family $\Gamma \equiv \{\gamma_1 = \gamma_1(z, k, T, c) : z, k \in R, T, c \in R - \{0\}\}$, where γ_1 is defined as $\gamma_1(z, k, T, c) = k + c[1 + \exp\{Tz\}]^{-1}$, when $c = 1, k = 0$ and $T = -1$ the classical asymmetric activation function is obtained, which is the most commonly used.

The estimated parameter $\hat{\underline{w}}$ is obtained by minimizing iteratively a cost functional $L(\underline{w})$, i.e., $\hat{\underline{w}} = \arg \min\{L_n(\underline{w}) : \underline{w} \in W\}$, $W \subseteq \Re^p$, where $L_n(\underline{w})$ is for example the ordinary least squares function i.e.

$$L_n(\underline{w}) = \frac{1}{2n} \sum_{i=1}^n (y_i - g(\underline{x}_i, \underline{w}))^2 \quad (3)$$

The loss function in equation (3) gives us a measure of accuracy with which an estimator A_λ , fits the observed data but it does not account for the estimator's (model) complexity. Given a sufficient large number of free parameter, $p = p(A_\lambda)$, a neural estimator A_λ , can fit the data with arbitrary accuracy.

3 Robust Statistical Theory

A qualitative robustness is a robustness concept that measure the global reliability of the estimator, which describes up to what distance from the model distribution the estimator still gives some relevant information.

For given F , the asymptotic distribution of $T(F)$ is G_F , so $L_G(\sqrt{n}[T_n - T(G)]) \xrightarrow{d} N(0, V(T, G))$, where T_n is the empirical distribution. The estimator T is said to be qualitatively robust at F if for every $\varepsilon > 0$ there exists a $\delta > 0$, such that $d(F, \hat{F}) < \delta$ implies $d(G_F, G_{\hat{F}}) < \varepsilon$. This mean that if the distribution of the y_t 's (F) is perturbed only slightly, then the corresponding change in the asymptotic distribution of the estimator (G_F) should also be small.

Another important robustness concept is the quantitative robustness, because it allows us to assess the influence of individual observations toward the value of an estimate. This is a local concept.

A basic quantitative robustness information is the influence function. The influence function(IF) of T at F is given by:

$$IF(x, T, F) = \lim_{t \rightarrow 0} \frac{T((1-t)F + t\delta_x) - T(F)}{t} \quad (4)$$

which gives the influence of the effect of an infenitesimal contamination of an statistic T of one additional observation, for $n \rightarrow \infty$. It measures the asymptotic bias caused by contamination in the observations.

Under mild regularity conditions, $\int IF(x; T, F) dF(x) = 0$ where 0 is the p -dimensional null vector. The covariance matrix is:

$$V(T, F) := \int IF(x; T, F) IF(x; T, F)^T dF(x) \quad (5)$$

Now let $\{F_\theta\}_\Theta$ be a parametric model with densities $f_\theta(x)$, and denote by:

$$s(x, \theta) := \frac{\partial}{\partial \theta} \ln f_\theta(x) := \left(\frac{\partial}{\partial \theta^{(1)}} \ln f_\theta(x), \dots, \frac{\partial}{\partial \theta^{(p)}} \ln f_\theta(x) \right)^T \quad (6)$$

the vector of likelihood scores, and by $J(\theta) := \int s(x, \theta) s(x, \theta)^T dF_\theta(x)$ the Fisher information matrix.

3.1 M-Estimators

Definition: An M-estimator (functional) is defined through a function $\rho : \chi \times \Theta \rightarrow \Re$ as the value $T(F) \in \Re^p$ minimizing $\int \rho(x, t) dF(x)$ over t , or through a function $\psi : \chi \times \Theta \rightarrow \Re^p$ as the solution for t of the vector equation

$$\int \psi(x, t) dF(x) = 0 \quad (7)$$

where $\psi(x, \theta) = \partial(x, \theta) / \partial \theta$. In this case the influence function is

$$IF(x; T, F) = M(\psi, F)^{-1} \psi(x, T(F)) \quad (8)$$

with the $p \times p$ matrix M given by $M(\psi, F) := - \int [\frac{\partial}{\partial \theta} \psi(x, \theta)]_{T(F)} dF(x)$, and the asymptotic covariance matrix is

$$V(T, F) = M(\psi, F)^{-1} Q(\psi, F) M(\psi, F)^{-T} \quad (9)$$

with $Q(\psi, F) := \int \psi(x, T(F)) \psi(x, T(F))^T dF(x)$

4 Robust/Efficient ANN Learning

In highly nonlinear models we could choose work under robust conditions that guarantee a small asymptotic biased but possibly inconsistent estimator, due to model misspecification, or with a consistent and probably asymptotically more efficient semiparametric M-estimator.

As a starting point for this analysis, we consider that an M-estimator is asymptotically efficient if and only if

$$IF = \frac{1}{I(F_\theta)} \frac{\partial \log f_\theta}{\partial \theta} \quad (10)$$

where we indicate with F_θ and f_θ the distribution and the density functions of a parametric model, and with

$$I(F_\theta) = \int \left(\frac{\partial \log f_\theta}{\partial \theta} \right)^2 dF_\theta \quad (11)$$

the Fisher information. Therefore IF must be proportional to the score function and for the asymptotic variance (AV) the following holds:

$$AV = \int IF(x, \theta, F)^2 dF_\theta \geq \frac{1}{I(F_\theta)} \quad (12)$$

In order to find a satisfactory compromise between efficiency and robustness in a general semiparametric set-up, we should regularize an infinite-dimensional problem reducing it to a constrained optimization objective function. This would lead to a so-called B-robust estimator, given its characterizing bounded IF. Like before, an optimality criterion is that of fixing bound through a constant $c > 0$ such that a bounded IF would result in the solution of the constrained optimization problem. In order not to sacrifice much efficiency, we could use an optimal B-robust estimator, i.e., an estimator that is consistent and asymptotically efficient within a class of estimators endowed with a certain degree of robustness. The class is characterized by

$$\min \int \psi^T \psi dP_{\theta,g} \quad \text{for all } \psi \text{ subject to } \sup_x \|\psi(x)\| \leq c \quad (13)$$

In semiparametric models, on the other hand, it is difficult to provide an explicit optimal IF; it can be found an approximate solution such that the lowest bound can be identified from the optimization problem. In the case of $c = \infty$ the problem is to find an efficient estimate, while with $c < \infty$ we need to find an optimal IF for an optimal estimator. The solution depends on the choice of the estimator and, in general, one-step versions of M-estimators should be considered because they result asymptotically efficient after one iteration from a consistent, robust and possibly endowed with an high BP initial estimate of the parameters of interest and with the functionals of the likelihood consistently estimated.

With iterative algorithms at hand, an important issue is to consolidate the network learning. The backpropagation algorithm is a possible recursive estimator that locally solves optimization problems, but it results asymptotically inefficient relatively to a NR-type estimator. Thus, one-step versions of m-estimators should be considered. They are based on the IF calculated above and solve moment equations like $\sum_i m(z_i, \phi, \hat{\eta})/n = 0$, given a general interest parameter ϕ . Then, considering $\hat{\phi} = \operatorname{argmax}_\phi \sum_i \ln f(z_i | \phi, \eta_\phi)$ as the GP(Max)L-estimator, where the estimation of η does not affect the AV, and given $M = (\frac{\partial E(m(z, \phi, \eta_0))}{\partial \phi})|_{\phi=\phi_0}$ nonsingular, we have $\psi(z) = M^{-1}m(z, \phi, \eta_0)$ and $\phi = \hat{\phi} + \sum_i \hat{\psi}(z_i, \hat{\phi})/n$, which is the one-step version of the GP(Max)L-estimator.

These Newton's M-estimates are usually only approximate; just in special cases we are able to find an explicit form of the IF that does not involve adjustment terms, while in the more general cases we should consider an estimate some bias correction terms, and this is indeed the case whenever we have a plugged-in estimator in an optimization function.

5 Application of ANN in the Regression Analysis

A multivariate model-free regression problem consists in n pairs of vectors

$$(\underline{y}_l, \underline{x}_l) = (y_{l1}, y_{l2}, \dots, y_{lq}; x_{l1}, x_{l2}, \dots, x_{lp}) \quad l = 1, 2, \dots, n \quad (14)$$

that have been generated from unknown models

$$y_{li} = h_i(\underline{x}_l) + \varepsilon_{li} \quad l = 1, 2, \dots, n \quad i = 1, 2, \dots, q \quad (15)$$

where h_i are unknown smooth non-parametric (model-free) functions, $h_i : \Re^p \rightarrow \Re$ $i = 1, \dots, q$, and ε_{li} are random variables i.i.d with zero mean, i.e. $E[\varepsilon_{li}] = 0$, and independent of \underline{x}_l .

The goal of regression is to construct estimators $\hat{h}_1, \hat{h}_2, \dots, \hat{h}_q$, which are functions of the data $(\underline{y}_l, \underline{x}_l)$, $l = 1, \dots, n$, to best approximate the unknown functions h_1, \dots, h_q , and use these estimates to predict a new \underline{y} given a new \underline{x} , $\hat{y}_i = \hat{h}_i(\underline{x})$ $i = 1, \dots, q$.

The regression problem using ANN as non-linear estimator of the unknown underlying function, have been longly study (See [HLM⁺94], [Spe91]).

A FANN provides a nonlinear approximation to h given by

$$\hat{y} = g(\underline{x}, \underline{w}) = \sum_{j=1}^{\lambda} w_j^{[2]} \gamma_1 \left(\sum_{i=1}^p w_{ij}^{[1]} x_i + w_{p+1,j}^{[1]} \right) \quad (16)$$

where the function $\gamma_1(\cdot)$ is a smooth bounded monotonic function.

The estimated parameter $\hat{\underline{w}}$ is obtained by minimizing iteratively a cost functional $L_n(\underline{w})$ i.e., $\hat{\underline{w}} = \arg \min \{L_n(\underline{w}) : \underline{w} \in \Theta\}$, $\Theta \subset \Re^p$, where $L_n(\underline{w})$ is for example the ordinary least squares function i.e.

$$L_n(\underline{w}) = \frac{1}{2n} \sum_{i=1}^n (y - g(\underline{x}_i, \underline{w}))^2 \quad (17)$$

6 Robust Neural Network Learning

The loss function presented in (17) is very sensitive in the presence of outliers, causing that the error produced by the outlier gives a large value compared to the other estimated errors caused by the training algorithm, so that the network will trend to approximate the outlier point over all other points.

With highly nonlinear models we could therefore choose of working under robust conditions that guarantee a small asymptotic biased but possibly inconsistent estimator, due to model misspecification, or with a consistent and probably more efficient semiparametric estimators.

So a robust learning algorithm it is necesary to overcome this difficulties. Huber and Hampel proposed a Robust theory in statistics and lately some investigations have been

done to introduce this theory in the field of Neural Networks (See [CM94], [Cap00], [AMS01]). Here we present some results.

Consider the non-linear model

$$y_i = h(\underline{x}_i^T, \theta) + \varepsilon_i \quad 1 \leq i \leq n \quad (18)$$

with i.i.d errors ε_i , $E[\varepsilon_i] = 0$, $V[\varepsilon_i] = \sigma_\varepsilon$ and θ the unknown parameter. But the observed data z_i , $1 \leq i \leq n$ is contaminated by an outlier generating process v_i , so $z_i = x_i + u_i v_i$, where u_i is a zero-one process with $P[u_t \neq 0] = \alpha$ and u_t has distribution G_u , $0 < \alpha << 1$, with heavy tail.

Due to the universal approximator porperty of the FANN, we use it to model the underlying function of the data,

$$z_i = h(\underline{x}_i, \underline{w}) + e_i \quad 1 \leq i \leq n \quad (19)$$

with i.i.d e_i , $E[e_i] = 0$, $V[e_i] = \sigma_e$, and $h(\underline{x}, \underline{w})$ has the form described in 16. \underline{w} are the unknown parameter of dimension p to be estimated.

$$\hat{e}_i = z_i - h(\underline{x}_i, \underline{w}) \quad 1 \leq i \leq n \quad (20)$$

The class of Generalize M (GM) estimators is used to obtain $\hat{\underline{w}}$ as an approximation of \underline{w} . This method consists in a measure to estimate the fitting of the ANN to the data other than the traditional measure. The class of generalized M (GM) estimators is defined implicity by the first order condition

$$\frac{1}{n} \sum_{t=1}^n \psi(x_t, r/\hat{\sigma}) D_{w_j} h(\underline{x}, \hat{\underline{w}}) = 0 \quad 1 \leq j \leq p \quad (21)$$

where the parameter $\hat{\sigma}$ and $r = x_t - h(\underline{x}, \hat{\underline{w}})$ and $D_{w_j} h(\underline{x}, \hat{\underline{w}}) = \frac{\partial}{\partial w_j} h(\underline{x}, \hat{\underline{w}})$.

The conditions that $\psi(\cdot, \cdot) : \mathbb{R}^p \rightarrow \mathbb{R}^p$ must satisfy to have nice asymptotic properties can be found in [HRRS86]

Instead of defining GM estimators as the solution to a first order condition, they can be defined as the minimand of the objective function

$$RL_n(\underline{w}) = \sum_{t=1}^n \rho\left(\frac{r}{s}\right) \quad (22)$$

with $\partial \rho(x, r)/\partial r = \psi(x, r)$. ρ is a robustifying function that introduces a bound influence of the outlier on the loss function, and s is a data-dependent robust scale estimate whose objective is to make the parameters invariant to scale transformation.

The estimated parameter $\hat{\underline{w}}$ is obtained by minimizing iteratively a cost functional $RL_n(\underline{w})$ i.e.

$$\hat{\underline{w}} = \arg \min \{RL_n(\underline{w}) : \underline{w} \in W\}, \quad W \subseteq \mathbb{R}^p \quad (23)$$

There are different classes of GM estimators. For example Mallows' GM estimators given by

$$\zeta(x, r, c) = v_x(x) \psi(r, c) \quad (24)$$

where $\tau(x, r, c) = \rho(r, c)$, $d\rho(r, c)/dr = \psi(r, c)$, and $v_x(x)$ is a weight function, $v_x : \Re^p \rightarrow [0, 1]$. A popular choice of the scale parameter is done by $\hat{\sigma} = 1.483\text{med}[|\hat{U}_t(\lambda) - \text{med}[\hat{U}_t(\lambda)]|]$.

When c is chosen fix the learning process of the ANN, i.e., estimation process of the model parameters, has some problems: an initial model of the parameters is needed to find the final parameters, and the efficiency of the algorithm is reduced. We propose a dynamic GM estimator by letting the learning process change the parameter c . The objective of changing c is to start from a GM estimator close to the LS estimator, so a much better initialized ANN is no longer needed, and the efficiency is improved in the early stage of the algorithm accelerating the convergence process.

There are several ways of changing the c parameter through time, some proposal can be found in [Sch91], as a stochastic technique for global optimization.

7 Simulation Results to the Synthetic Data

In this section the performance of the Robust Neural Network apply to parameter estimation in regression problem, introduced in the preceding sections, is investigated for an example of nonlinear function.

Suppose we have a complicated interaction function as proposed by [HLM⁺94],

$$h(x_1, x_2) = 1.9(1.35 + e^{x_1} \sin(13(x_1 - 0.6)^2) e^{-x_2} \sin(7x_2)) \quad (25)$$

where (x_1, x_2) are two independent variables generated from the uniform distribution $U([0, 1]^2)$. The output data is generated using equation (25) with an addition of an independent and identically distributed (iid) Gaussian noises, $y_i = h(x_1^i, x_2^i) + a_i$, where $a \sim N(0, \sigma_a^2)$.

The observed data is obtained by $z_t = x_t + u_t v_t$, where v_t is a zero-one process with $P[u_t \neq 0] = \alpha$ and u_t has distribution $F_u = N(0; \sigma_u^2)$. With $0 < \alpha << 1$.

The experiment was performed for a fixed sample size of $n = 10000$. To avoid initialization effects the first 300 observations were discarded, 300 observations were used for training and 1000 for testing. A FANN with 2 inputs, 1 output and one hidden layer with ten hidden neurons was used to fit the data.

The experiment was performed using $\sigma_a^2 = 0.1$ with Additive outliers and $\alpha = 0.05$ and $\sigma_u^2 = 10$.

To train the FANN, backpropagation with momentum enhancement was used to find the parameters of the network. The learning process was performed using LSE as described in equation (17), and for the robust learning a combination of the η function of the Mallows type as described in equation (24) with the Huber's ψ function and Tukey's bi-square function,

$$\psi_H(r, c) = \text{sgn}(r) \min\{|r|, c\} \quad (26)$$

$$\psi_B(r, c) = \begin{cases} r(1 - (r/c)^2)^2 & r \in [-c, c] \\ 0 & r < -c \text{ or } r > c \end{cases} \quad (27)$$

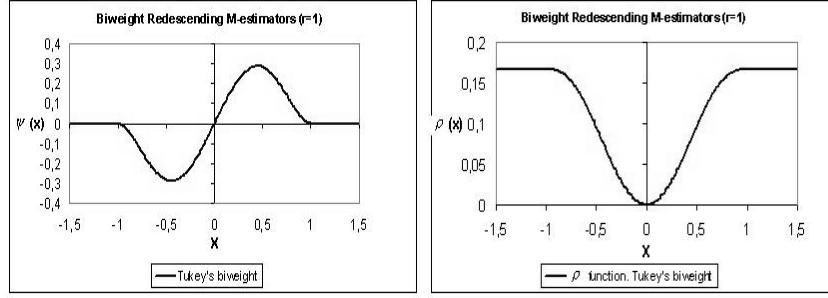


Fig. 1. a)(left)Biweight Redescending M-estimator b)(right) Biweight ρ function

called GMH and GMB respectively, was used. (The graph of this functions are presented in figure 1. The assessment of performance was done by comparing the fitted model with the “true” function counterparts on a large independent test set. The results obtained are shown in table 1 and in figure ?? is shown the unknown surface, and the surface generated by the ANN when different estimate it is used.

Table 1. MSE results for the Robust Learning process of the FANN

Estimates	Additive	
	$\alpha = 0.05$ and $\sigma_u^2 = 10$	
	Train	Pred
LSE	1.1881	4.8985
GMB	1.1735	4.5007
GMH	1.0901	4.5119

8 Concluding Remarks

We show instead that statistical robustness and efficiency pay a key role and suggest an initial approach to merge the advantages that they bring into the statistical analysis of the so-called semiparametric neural networks. We illustrate that many real problems could benefit from robustness concepts. As special case of non-linear function to model data from an unknown function. ANN fit by classical least squares estimate were shown to be sensitive to the presence of outliers. We introduced a robust learning procedure based on the GM-estimate to improve the qualitative and quantitative robustness of the neural network. The robust learning algorithm proposed shows good performance for the predictor models on synthetic data set.

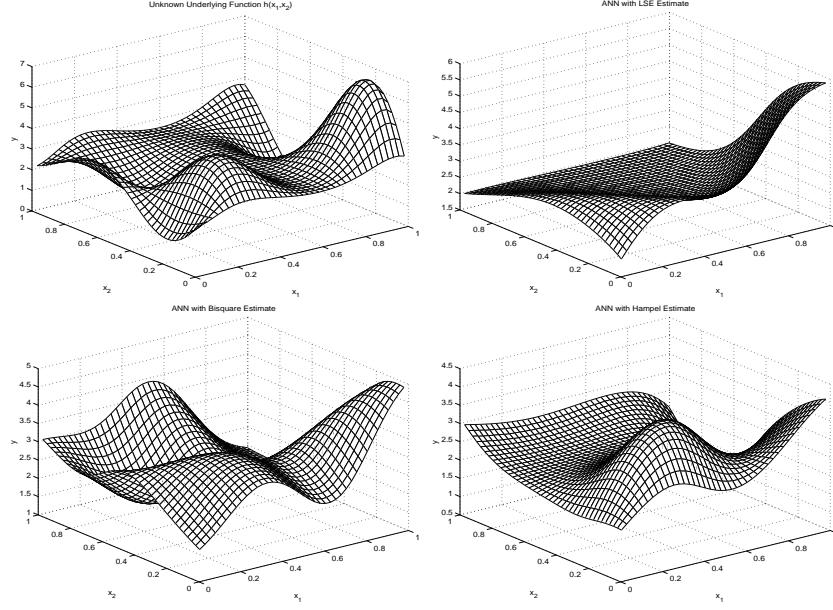


Fig. 2. a)(upperleft)Unknown Underlying Function b)(upperrighth) ANN with LSE Estimate
c)(downleft)ANN with GMB Estimate d)(downrigh) ANN with GMH Estimate

References

- [AMS01] H. Allende, C. Moraga, and R. Salas. Neural model identification using local robustness analysis. *Lecture Notes in Computer Science*, 2206:162–173, Nov 2001.
- [Cap97] E. Capobianco. A semiparametric approach for artificial neural networks modelling and learning theory. *Operations Research and Computer Science Interfaces Series, Mathematics of Neural Networks: Models, Algorithms and Applications, Kluwer, Dordrecht*, pages 140–145, 1997.
- [Cap00] E. Capobianco. Neural networks and statistical inference. seeking robust and efficient learning. *Comp. Statistics & Data Analysis*, (32):443–454, 2000.
- [CM94] J.T. Connor and R.D. Martin. Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, 2(5):240–253, 1994.
- [HLM⁺94] J. Hwang, S. Lay, M. Maechler, D. Martin, and J. Schmert. Regression modeling in back-propagation and projection pursuit learning. *IEEE Transactions on Neural Networks*, 5(3):342–353, May 1994.
- [HRRS86] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics, 1986.
- [Sch91] F. Schoen. Stochastic techniques for global optimization: a survey of recent advances. *J. Global Optim. 1*, 3:207–228, 1991.
- [Spe91] Donald Specht. A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6):568–576, November 1991.
- [Vap95] V. Vapnik. *The nature of Statistical Learning Theory*. Ed. Springer Verlag, 1995.