
UMA NOVA METODOLOGIA PARA TREINAMENTO EM REDES NEURAS MULTI CAMADAS

Luiz Carlos C. Pedroza

Pedroza@cefet-rj.br
CEFET-RJ
Av. Maracanã, 229
Rio de Janeiro, CEP 20271-110

Carlos E. Pedreira

pedreira@ele.puc-rio.br
DEE PUC-RIO
CP.: 38063
Rio de Janeiro, CEP 22452-970

RESUMO : Neste artigo apresenta-se uma metodologia de aprendizado supervisionado não usual baseada em um operador, batizado de *Operador de Extensão*. As redes neurais multicamadas são abordadas sob a ótica deste operador, produzindo algumas vantagens em comparação aos algoritmos de treinamento tradicionais, além de uma visão sistematizada de algumas questões de grande importância como generalização e incorporação de informação *a priori*. No algoritmo proposto o número de parâmetros a ser estimado é inferior ao algoritmo de Retropropagação, propiciando assim, entre outras vantagens, uma maior probabilidade de generalização. É apresentado um conjunto de resultados numéricos.

A New Methodology for Training Multilayer Neural Networks

Abstract: In this paper a non-conventional methodology for supervised learning is proposed. Multilayer neural networks are approached by using a new operator, the extension operator, producing some advantages. Furthermore, it introduces a systematic view of some very important subjects, like generalization and *a priori* knowledge incorporation. In the proposed algorithm a smaller number of estimated parameters is needed, increasing the probability of generalization. Some numerical results are presented.

1 INTRODUÇÃO

Os esquemas de aprendizado através de exemplos, que fundamentam a aquisição de conhecimento em redes neurais, podem ser entendidos como uma busca sistemática de um mapeamento que reconstrua satisfatoriamente o ambiente

gerador. O que se procura, do ponto de vista funcional, é um mapeamento, dentro de um conjunto de ‘mapeamentos hipótese’, que minimize uma medida de distância entre uma função geradora desconhecida e esta hipótese.

A eficiência de um esquema de aprendizado pode ser associada a três questões fundamentais: a uma escolha de um conjunto de hipóteses G apropriado ao problema em questão; à disponibilidade de um conjunto de exemplos que contenha informação suficiente sobre o mapeamento que se deseja reconstruir; e a um esquema eficiente de busca da melhor hipótese g dentro de um conjunto G .

Em geral, a utilização adequada de conhecimento prévio de algumas características do mapeamento que se deseja reconstruir, permite obter uma solução melhor. Esta possibilidade de incorporar conhecimento *a priori* foi originalmente desenvolvida em Abu-Mostafa (1994) e Abu-Mostafa (1995) através da indução de ‘Sugestões’ (Hints). Através desta metodologia, é possível usar o conhecimento *a priori* para ampliar artificialmente o conjunto de exemplos que compõe o conjunto de treinamento e também para direcionar a busca no conjunto de hipóteses G . Por exemplo, sabendo-se de antemão que uma função que se deseja reconstruir é uma função par, pode-se adicionar uma espécie de penalidade toda vez que a função hipótese for ímpar, ou gerar um conjunto adicional de exemplos artificiais (Abu-Mostafa, 1994; Abu-Mostafa, 1995). Em ambos os casos, se está sugerindo o direcionamento no espaço de busca.

A metodologia proposta neste artigo baseia-se em um operador originalmente introduzido em Pedroza (1997), o *Operador de Extensão*. A partir deste, pretende-se, entre outros objetivos, fomentar a discussão de alguns aspectos relativos à incorporação de informação *a priori* em redes neurais.

O Operador de Extensão, peça central deste artigo, será formalmente definido na seção 2. A idéia fundamental é definir um operador $T: [B \times M \times P] \rightarrow P$ que mapeie um ponto (b, I, f) em uma função f_{aum} , onde:

¹ Este trabalho foi parcialmente financiado pela CAPES através de bolsa de doutorado, e pelo CNPq, através de bolsa de produtividade em pesquisa, concedidas aos autores.

- (i) O conjunto de funções (no Espaço Produto Interno (Naylor e Sell, 1982) P) $b \in B$ é batizada como 'Base de Extensão' do Conjunto de Bases B .
- (ii) Qualquer combinação linear das funções de uma Base de Extensão b pertence ao espaço de produto interno P .
- (iii) O intervalo I , do conjunto de intervalos (Bartle, 1966) M , é o conjunto de pontos onde a função f é previamente conhecida.

Como ver-se-a na seqüência, quando o conjunto de exemplos (f,I) contém uma quantidade suficiente de informação e o conjunto de funções $b \in B$ é adequadamente escolhido, a função f_{aum} gerada pelo Operador de Extensão mostra-se uma boa reconstrução da função f original. O problema que se apresenta neste ponto, é o de como fazer uma escolha adequada da base de extensão b . Este problema é abordado na seção 3. Verificar-se-a que a partir de um conhecimento *a priori* do conjunto de hipóteses G , é possível efetuar uma escolha apropriada do conjunto de bases B . Será mostrado que se pode buscar a Base de Extensão b que faça uma reconstrução otimizada da função f .

Na seção 4, apresenta-se uma abordagem para redes neurais multicamadas (Rumelhart *et alii*, 1986; Haykin S., 1999) sob a ótica do Operador de Extensão. A idéia central parte da verificação de que uma rede neural em camadas pode ser vista como uma representação em grafo de funções no conjunto imagem de um operador. Este operador, derivado do Operador de Extensão e intitulado *Operador de Extensão Projeção*, será definido na próxima seção. Esta analogia entre *Operador de Extensão Projeção* e redes neurais permitirá que se escolha conjuntos de hipóteses G apropriados.

Embora redes neurais tenham alcançado bastante sucesso em aplicações nas mais diversas áreas do conhecimento, relativamente pouco se pode encontrar na literatura com respeito a desenvolvimentos teóricos. Este artigo pretende apresentar resultados com algum grau de rigor com a intenção de contribuir neste aspecto.

Na seção 5, apresenta-se alguns resultados numéricos e a seção 6 é reservada às conclusões e observações finais.

2 O OPERADOR DE EXTENSÃO

Seja f uma função definida sobre o domínio M , $D(f) = M$. Seja h uma função definida sobre o subconjunto $I \subset M$, $D(h) = I$. Se $h(x) = f(x)$ para todo $x \in I$, então h é dita uma restrição de f em I ; denominada $f|_I$. Por outro lado, seja $f|_I$ uma função definida sobre o domínio I . Uma função g tal que $I \subset D(g)$ e $g(x) = f(x)$ para todo $x \in I$ é dita uma extensão de $f|_I$; $g = f_{\text{aum}}$ (Naylor e Sell, 1982). Note-se que não necessariamente $f_{\text{aum}} = f$.

A partir destas definições, é possível estabelecer um operador, o *Operador de Extensão*, que realizará a extensão de uma restrição $f|_I$ em uma função f_{aum} , baseando-se para isto em algumas *sugestões* que, se forem suficientemente consistentes, farão que a f_{aum} seja uma reconstrução satisfatória da função f original.

Seja P um Espaço Produto Interno de funções definidas sobre o suporte M . Seja M uma álgebra de intervalos (Bartle, 1966) gerada por subintervalos de M . Seja $B = \{b_w; w \in W\}$ um conjunto de bases de extensão $b_w = \{e_i(w; \bullet); i \in A\}$ onde cada base é um conjunto de funções em P . Sejam W e A

conjuntos de índices apropriados para indexar os elementos de B e as funções das bases de extensão.

Define-se uma matriz G e um vetor H tal que seus elementos são dados por:

$$\langle e_i(w; \bullet), e_j(w; \bullet) \rangle_I \text{ e } \langle f, e_i(w; \bullet) \rangle_I$$

respectivamente.

Definição D-1:

Se o sistema

$$G(b_w)\theta(b_w) = H(b_w), \quad (2.1)$$

tiver solução única, o *Operador de Extensão*

$T: [B \times M \times P] \rightarrow P$ pode ser definido no ponto (b_w, I, f) como:

$$T(b_w, I, f)(x) \equiv \begin{cases} f(x) & \text{se } x \in I \\ \sum_{k \in A} \theta_k(w) e_k(w; x) & \text{se } x \in M - I \end{cases}$$

Definição D-2:

Sob as mesmas hipóteses pode-se definir, no ponto em questão, o *Operador de Extensão Projeção* $T^*: [B \times M \times P] \rightarrow P$ como:

$$T^*(b_w, I, f) \equiv \sum_{k \in A} \theta_k(w) e_k(w; \bullet). \quad (2.2)$$

Observe-se que os operadores da definição D-1 e D-2 operam a restrição de f definida sobre o intervalo I (denominada $f|_I$), já que o produto interno $\langle \bullet; \bullet \rangle_I$ é calculado considerando as funções com valor nulo fora deste intervalo. Note-se também, que *sugestões* de como proceder à extensão são passadas para os operadores através de um conjunto de funções, as Bases de Extensão. Caso a Base de Extensão seja bem escolhida a função f_{aum} será uma boa reconstrução da função f .

O Operador de Extensão e o Operador de Extensão Projeção foram desenvolvidos (Pedroza, 1997; Pedroza e Pedreira, 1998) de forma que caso as funções que compõem a Base de Extensão b_w sejam ortogonais entre si sobre um intervalo M , então, para qualquer conjunto de escalares α_n ,

$$\|T(b_w, I, f) - T^*(b_w, I, f)\|_M \leq \left\| T(b_w, I, f) - \sum_{k \in A} \alpha_k e_k(w; \bullet) \right\|_M \quad (2.3)$$

A partir de 2.3, é possível verificar que $T^*(b_w, I, f)$ é a projeção ortogonal de $T(b_w, I, f)$ no espaço expandido pelas funções da Base de Extensão b_w . Esta propriedade justifica a denominação do Operador de Extensão Projeção.

No caso das funções da Base de Extensão não serem ortogonais entre si só é possível garantir que

$$\left\| f - \sum_{i \in A} \theta_i(w) e_i(w; \bullet) \right\|_I \leq \left\| f - \sum_{i \in A} \alpha_i e_i(w; \bullet) \right\|_I \quad \forall \alpha_n \in C \quad (2.4)$$

onde, de acordo com 2.2,

$$\sum_{i \in A} \theta_i(w) e_i(w; \bullet) = T^*(b_w, I, f).$$

A inequação 2.4 implica que a solução do sistema 2.1 fornece a combinação linear das funções que compõem a base de extensão b_w que melhor aproxima a função f no intervalo I , onde ela é previamente conhecida. Pode-se então concluir que, indiferente ao uso de uma Base de Extensão b_w ou de uma Base de Extensão composta da normalização das funções da base b_w , a função mapeada pelo Operador de Extensão é a mesma. De um ponto de vista informal, corresponderia a afirmar que os operadores realizam uma normalização prévia das funções que compõem a base de extensão.

Para ilustrar a capacidade de reconstrução do Operador de Extensão e a relevância da escolha apropriada da Base de Extensão, suponha-se um exemplo de uma função

$$f = \sum \alpha_n e_n, \text{ que se sabe ser definida sobre o suporte } M.$$

Conhece-se apenas sua restrição sobre o intervalo I . Ao escolher-se uma Base de Extensão b que contenha todas as funções e_i que compõem o somatório descritivo de f , se esta base garantir a unicidade de solução do sistema 2.1, então, pela inequação 2.3, $f_{\text{aum}}(b)=f$. Por este exemplo, percebe-se que a capacidade de reconstrução do Operador de Extensão está intimamente relacionada com a escolha apropriada da Base de Extensão. Para fazer-se uma escolha apropriada da base de extensão, foi necessária uma boa 'sugestão' sobre a função f ; no caso o conhecimento prévio da função. Contudo, geralmente a 'sugestão' que se dispõe constitui-se no conhecimento prévio do conjunto G ao qual a função f pertence. Baseando-se neste conhecimento, em alguns casos, é possível fazer a escolha de um conjunto de Base de Extensão B que faça com que o conjunto imagem de $T(B, I, \bullet)$ iguale-se a G ; $\mathfrak{R}(T(B, I, \bullet)) = G$. Ao fazer-se esta escolha certamente uma das bases de extensão de B reconstruirá a função f satisfatoriamente.

Entretanto, devido à existência de um compromisso entre a cardinalidade das Base de Extensão de B e o intervalo I , a fim de satisfazer unicidade de solução do sistema 2.1, geralmente só é permitido escolher-se um B tal que $\mathfrak{R}(T(B, I, \bullet)) \subset G$.

Com a finalidade de ilustrar este conceito, suponha que se saiba de antemão que a função f , que se deseja reconstruir, é um sinal com espectro de frequência limitado em w_{max} . Escolhendo-se um conjunto B de todas as Bases de Extensão possíveis compostas por conjuntos de Z funções senoidais com frequência máxima w_{max} , ainda não se pode garantir que $f \in \mathfrak{R}(T(B, I, \bullet))$. Entretanto, pode-se perceber que algumas Base de Extensão irão realizar uma boa reconstrução de f , pelo menos dentro de um intervalo Q onde I está contido. Aumentando-se Z , aumenta-se a probabilidade de f pertencer a imagem do operador, contudo, aumenta-se também a probabilidade do sistema 2.1 não ter solução única.

É esperado que uma Base de Extensão produza uma boa reconstrução da função f dentro do intervalo I , e também para os pontos externos a este intervalo, situados próximos a sua fronteira. As simulações apresentadas ao final deste artigo apresentam este comportamento. O tamanho dos intervalos externos ao intervalo I , para os quais é possível obter-se um bom resultado, está diretamente ligado a escolha de B . É importante ressaltar que esta escolha reflete a existência e qualidade de incorporação de conhecimento a priori. Como era esperado, a adição adequada de conhecimento a priori propicia uma melhora na capacidade de generalização.

3 ALGORITMO DE ESCOLHA DA BASE DE EXTENSÃO ÓTIMA

O problema da escolha da Base de Extensão em B que possibilita uma melhor reconstrução da função f pode ser formulado do seguinte modo:

Problema 1

$$\min \|f - T^*(b_w, I, f)\|_I$$

$$b_w \in B.$$

onde B é o conjunto de Bases de Extensão pertencentes à mesma família e obtidas pela variação de w em W .

É possível solucionar este problema recursivamente, variando-se w em W , de forma a buscar-se uma nova Base de Extensão $b_{w(k+1)}$ em B , tal que

$$\begin{aligned} & \left\| f - \sum_{i \in A} \theta_i(w(k)) e_i(w(k+1); \bullet) \right\|_I \\ & \leq \left\| f - \sum_{i \in A} \theta_i(w(k)) e_i(w(k); \bullet) \right\|_I \end{aligned} \quad (3.1)$$

Se as funções que compõem as Bases de Extensão forem diferenciáveis no argumento, pode-se utilizar o algoritmo do gradiente descendente para realizar esta busca em B . Neste caso, o cálculo do novo $w(k+1)$ é efetuado da seguinte forma:

$$\begin{aligned} w_n(k+1) &= w_n(k) \\ & - \eta \frac{\partial \left(\left\| f - \sum_{i \in A} \theta_i(w_i(k)) e_i(w_i(k), \bullet) \right\|_I \right)}{\partial w_n(k)} \end{aligned} \quad (3.2)$$

onde $0 < \eta < 1$.

De posse da nova base $b_{w(k+1)}$, utiliza-se o sistema 2.1 para calcular o novo espectro $\theta(b_{w(k+1)})$. A partir da inequação 2.3 pode-se garantir (Pedroza, 1997) que:

$$\begin{aligned} & \left\| f - \sum_{i \in A} \theta_i(w(k+1)) e_i(w(k+1); \bullet) \right\|_I \\ & \leq \left\| f - \sum_{i \in A} \theta_i(w(k)) e_i(w(k+1); \bullet) \right\|_I \end{aligned} \quad (3.3)$$

Percebe-se que, em cada iteração k , primeiramente busca-se, variando-se apropriadamente w em W , um novo conjunto de funções no qual a combinação linear, dada pelo conjunto de pesos $\theta(b_{w(k)})$, reduz a função de custo. Em seguida calcula-se, através do sistema 2.1, uma nova combinação linear para estas funções que reduza ainda mais a função de custo. Este algoritmo é descrito a seguir de forma esquemática.

Algoritmo proposto

- 1) Faça $k=0$ e inicialize o vetor de parâmetros $w(k)$.
- 2) Calcule o espectro $\theta(b_{w(k)})$ resolvendo o sistema 2.1.

3) Calcule as derivadas

$$\delta_n = \frac{\partial \left(\left\| f - \sum_{i \in A} \theta_i(w_i(k)) e_i(w_i(k), \bullet) \right\|_I \right)}{\partial w_n(k)}$$

considerando o espectro $\theta(b_{w(k)})$ constante.

4) Calcule o novo vetor de parâmetros $w(k+1)$ pela fórmula

$$w_n(k+1) = w_n(k) - \eta \delta_n$$

5) Calcule o novo espectro $\theta(b_{w(k+1)})$ a partir do sistema 2.1.

$$6) \text{ Calcule } \left\| f - \sum_{i \in A} \theta_n(w(k+1)) e_i(w_i(k+1), \bullet) \right\|_I .$$

$$7) \text{ Se } \left\| f - \sum_{i \in A} \theta_n(w(k+1)) e_i(w_i(k+1), \bullet) \right\|_I \geq \left\| f - \sum_{i \in A} \theta_n(w(k)) e_i(w_i(k), \bullet) \right\|_I \text{ então pare;}$$

caso contrário faça $k=k+1$, e vá para 3.

A convergência do algoritmo 1 está provada, quando η é suficientemente pequeno, em Polak (1971). Vale notar que o algoritmo gradiente decrescente, e conseqüentemente o algoritmo proposto, está sujeito a convergir para mínimos locais. Também, porque o conjunto B não é um espaço métrico completo, a função de custo a ser minimizada pode possuir mais de um mínimo global. Estas situações podem ser contornadas escolhendo-se um intervalo I suficientemente grande.

Com o propósito de viabilizar a aplicação do algoritmo proposto em casos reais, onde o intervalo I constitui-se de um conjunto de N pontos e as Bases de Extensão tem cardinalidade Z, descreve-se o problema 1 como :

Problema 2

$$\min \sum_{i=1}^N \left(f(x_i) - \sum_{n \in A} \theta_n(w) e_n(w_n; x_i) \right)^2 \quad (3.4)$$

onde

$$\begin{bmatrix} \theta_1(w) \\ \theta_2(w) \\ \vdots \\ \theta_z(w) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N e_1(w_1, x_i) e_2(w_2, x_i) & \cdots & \cdots & \cdots \\ \sum_{i=1}^N e_2(w_1, x_i) e_1(w_1, x_i) & \cdots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \cdots & \sum_{i=1}^N e_z(w_1, x_i) e_z(w_2, x_i) \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_{i=1}^N f(x_i) e_1(w_1, x_i) \\ \sum_{i=1}^N f(x_i) e_2(w_2, x_i) \\ \vdots \\ \sum_{i=1}^N f(x_i) e_z(w_z, x_i) \end{bmatrix} \quad (3.5)$$

4 REDE NEURAIS E O OPERADOR DE EXTENSÃO

Considere-se a rede neural representada na Figura 1, onde o neurônio na camada de saída tem função de ativação linear. Não há perda de generalidade pela utilização desta arquitetura simplificada, a extensão para arranjos mais complexos é imediata.

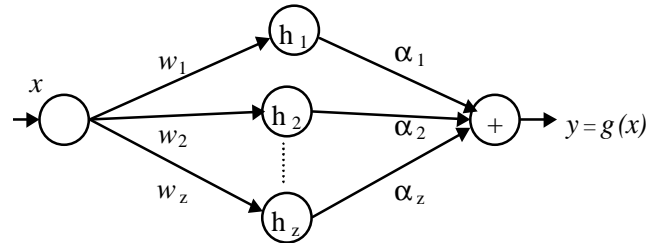


Figura 1

Os neurônios da camada escondida tem uma função de ativação $h(\bullet)$ que necessita apenas ser diferenciável no argumento e não monotonamente crescente, como nos casos clássicos do algoritmo de Retropropagação. Com a estrutura proposta, um ponto x na entrada da rede mapeia um ponto $g(x)$ na sua saída, descrito como :

$$g(x) = \sum_n \alpha_n h_n(w_n x) \quad (4.1)$$

Observando-se a equação 2.2, percebe-se que o somatório descritivo do Operador de Extensão Projeção, definido na seção 2, pode ser realizado pela rede neural representada na Figura 1 desde que se escolha $\theta_k(w) = \alpha_k$ e $h_k(w_k x) = e_k(w; x)$.

Conclui-se então que, para toda função f_{aum}^* pertencente à imagem Operador de Extensão Projeção, existe uma representação em grafo dada por uma rede neural como a da Figura 1.

Como por hipótese na definição 2.1 todas as Bases de Extensão de B pertencem à mesma família de funções, tem-se que a variação dos pesos w anteriores a camada oculta corresponde, sob a ótica do Operador de Extensão, à busca uma nova Base de Extensão em B.

Durante a fase de treinamento varia-se os pesos da rede de forma a buscar uma nova função g que minimize a norma $\|f - g\|_I$. Sob a ótica do Operador de Extensão, isto corresponde a solucionar o problema 1 introduzido na seção 3.

Note-se que a função de custo (3.4) corresponde à função de energia do algoritmo de Retropropagação. Deste modo, pode-se afirmar que a diferença central entre o algoritmo de Retropropagação e o algoritmo proposto reside que, no primeiro, o gradiente decrescente é utilizado para alterar tanto os pesos das conexões anteriores à camada oculta w_i , quanto os pesos ligados à saída α_i . No esquema proposto, o algoritmo de gradiente decrescente é utilizado apenas para otimizar os pesos das conexões anteriores à camada oculta enquanto os pesos conectados à saída são calculados de forma fechada resolvendo-se o sistema 3.5. Como conseqüência, além da diminuição do esforço computacional, obtém-se uma redução da quantidade de variáveis livres no espaço de estado da função de energia. Com um menor número de parâmetros a estimar,

diminui-se os riscos de super treinamento e, portanto, melhora-se a capacidade de generalização. Também como consequência da diminuição do número de parâmetros a estimar reduz-se o volume de dados necessário a um bom desempenho.

5 RESULTADOS NUMÉRICOS

Nesta seção ilustra-se a metodologia proposta através da apresentação de exemplos numéricos. O propósito destas ilustrações não é o de proporcionar comparações sistemáticas com técnicas tradicionais.

No primeiro exemplo procura-se demonstrar como a presença de uma 'sugestão' pode auxiliar na escolha de um conjunto de hipóteses G apropriado para o algoritmo operar.

Exemplo 5.1

Neste exemplo é abordado um problema de extrapolação e apresentada uma comparação com algoritmo clássico de Retropropagação. O problema de extrapolação é aquele no qual se deseja estimar valores para uma função, fora do intervalo aonde são colhidos os pontos usados para treinamento. O que se procura é, a partir de um treinamento com pontos amostrados em um intervalo $[a, b]$, obter uma estimativa para a função f em x_k , $\hat{f}(x_k)$, $x_k > b$. Em um problema de interpolação, a solução seria, em geral, mais fácil, buscar-se-ia uma estimativa para a função f em x_k , $\hat{f}(x_k)$, $x_k \in [a, b]$.

Suponhamos que a função $f(x)=0.7+\text{sen}(2\pi x)+0.5\text{cos}(\pi x)$ seja conhecida a priori nos pontos que vão de 0 até 3 com um passo de 0.01; $I=\{0, 0.01, 0.02, \dots, 2.99, 3\}$. Suponha-se que se saiba, de antemão, que este sinal é obtido na saída de um filtro passa-baixa ideal com frequência de corte em $2,5\pi$ rad/s. A partir desta 'sugestão' e da limitação imposta pela cardinalidade de I , escolhe-se conjunto de base de extensão

$B = \{b_w = [1, \text{sen}(w_1x), \text{cos}(w_2x), \text{sen}(w_3x), \text{cos}(w_4x)] : w_i \in \mathbb{R} \text{ e } |w_i| \leq 2,5\pi \text{ rad/s}\}$.

A partir da analogia discutida na seção 4, esta escolha de B corresponde a escolher uma rede neural, como a da Figura 1, com cinco neurônios na camada escondida; um com função de ativação linear e peso unitário, dois com funções de ativação senoidal e pesos w_1 e w_3 , e dois com funções de ativação cosenoidais e pesos w_2 e w_4 .

Inicializado com a Base de Extensão

$$b_{in}=[1, \text{sen}(2x), \text{cos}(3x), \text{sen}(4x), \text{cos}(5x)]$$

e a constante de aprendizado $\eta=0.2$, em 1000 interações o algoritmo proposto converge para a função

$$g(x) = 0.7 + 1.00\text{sen}(6.283x) - 0.6 * 10^{-12} \text{cos}(0.25x)$$

$$-0.83 * 10^{-10} \text{sen}(6.1536x) + 0.4986\text{cos}(3.1415x).$$

obtendo-se um erro médio quadrático, no intervalo de treinamento, de 10^{-13} . Como pode-se observar, através da figura 2, a reconstrução da função é bastante satisfatória na vizinhança dos pontos do intervalo I , e permanece com boa performance mesmo para pontos bastante afastados do horizonte de treinamento.

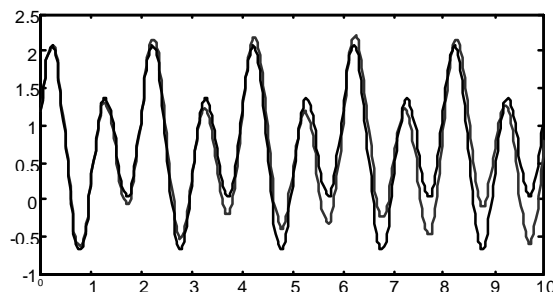


Figura 2 - Funções verdadeira e estimada pela metodologia proposta para horizonte de treinamento entre 0 e 3 e de generalização (a partir de 3)

O mesmo experimento foi implementado utilizando-se uma rede, com uma camada oculta de 10 neurônios e funções de ativação sigmoidais, treinada através do algoritmo tradicional de Retropropagação. A figura 3 mostra a reconstrução operada pela rede. Pode-se observar que, apesar de uma boa reconstrução dentro do intervalo $[0, 3]$, a performance deteriora-se rapidamente fora desta região.

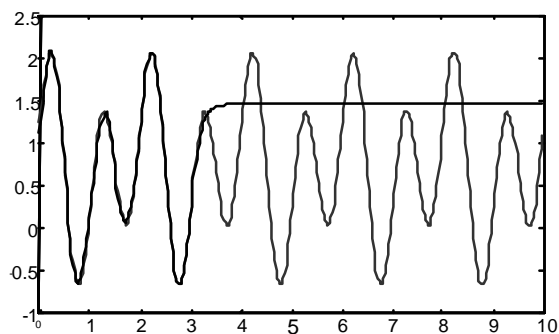


Figura 3 - Funções verdadeira e estimada por Retropropagação para horizonte de treinamento entre 0 e 3 e de generalização (a partir de 3)

O exemplo a seguir trata a simulação de um problema simples de identificação paramétrica a partir do uso do Operador de extensão.

Exemplo 5.2

Suponha-se que um sistema real é descrito pela equação:

$$y(t) = 0.7x(t) + 0.5u(t) + 0.2 + \varepsilon$$

Este sistema linear está sujeito a uma perturbação ε que corresponde a um ruído Gaussiano com média nula e variância 0.1.

O sistema será modelado através de :

$$\hat{y}(t) = \mathbf{a}\hat{x}(t) + \mathbf{b}\hat{u}(t) + \mathbf{c}.$$

onde $\hat{x}(t) = x(t) + \mu$, $\hat{u}(t) = u(t) + v$ e $\hat{y}(t) = y(t) + \pi$ representam as observação de x , u , y num instante t adicionadas a ruídos com distribuição normal de média nula e variância 0.01.

Escolhendo-se $I_k=\{0, h, 2h, \dots, (k-1)h\}$, onde h é o intervalo de amostragem e a Base de Extensão $b_w = [\hat{x}, \hat{u}, 1]$ pode-se resolver o sistema 2.1 com

$$G(b_w) = \begin{bmatrix} \langle \hat{x}, \hat{x} \rangle_{I_k} & \langle \hat{x}, \hat{u} \rangle_{I_k} & \langle \hat{x}, 1 \rangle_{I_k} \\ \langle \hat{u}, \hat{x} \rangle_{I_k} & \langle \hat{u}, \hat{u} \rangle_{I_k} & \langle \hat{u}, 1 \rangle_{I_k} \\ \langle 1, \hat{x} \rangle_{I_k} & \langle 1, \hat{u} \rangle_{I_k} & \langle 1, 1 \rangle_{I_k} \end{bmatrix}$$

$$H(b_w) = \begin{bmatrix} \langle \hat{y}, \hat{x} \rangle_{I_k} \\ \langle \hat{y}, \hat{u} \rangle_{I_k} \\ \langle \hat{y}, 1 \rangle_{I_k} \end{bmatrix},$$

para se calcular o vetor $\theta(b_w)$ cujos elementos correspondem as estimativas de \mathbf{a} , \mathbf{b} e \mathbf{c} que minimizam a norma $\|y - \hat{y}\|_{I_k}$.

A figura 4, a seguir mostra a evolução das estimativas no tempo.

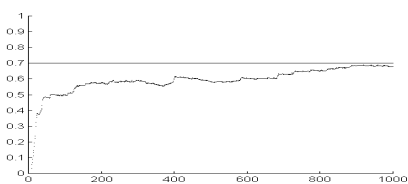


Figura 4.a - estimativa de \mathbf{a}

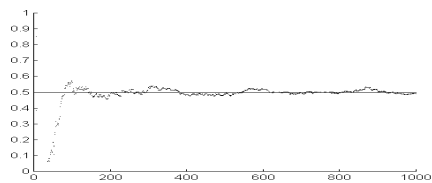


Figura 4.b - estimativa de \mathbf{b}

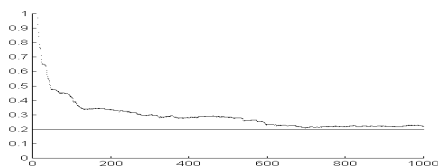


Figura 4.c - estimativa de \mathbf{c}

6 OBSERVAÇÕES FINAIS

Neste artigo apresentou-se uma abordagem para treinamento de redes neurais em camadas diferenciada da metodologia clássica baseada em Retropropagação do Erro. O algoritmo proposto apresenta características interessantes: permite o uso de outras funções de ativação sem a restrição de monotonicidade crescente; reduz a probabilidade de atingir-se mínimos locais; e acelera o processo de convergência. Estas são algumas conseqüências de tratar-se as redes neurais como representações em grafo de pontos na imagem do Operador de Extensão.

Uma linha interessante de continuação seria a incorporação da metodologia de sugestão proposta em Abu-Mostafa (1994) e Abu-Mostafa (1995) com a finalidade de buscar-se uma Base de Extensão otimizada.

7 REFERÊNCIAS BIBLIOGRÁFICAS

Abu-Mostafa. Y. S. , 1994, "Learning from Hints", *Jornal of Complexity*, V.10, PP.165-178.

Abu-Mostafa. Y. S., 1995, "Hints", *Neural Computation*, Vol7, PP 639-671.

Bartle, R.G., 1966, *The elements of integration*. New York, Jonh Wiley & Sons Inc.

Haykin S., 1999, *Neural Networks: A Comprehensive Foundation*, second edition, Prentice Hall.

Naylor, A.W., Sell, G.R., 1982, *Linear Operator Theory in engineering and science*. 2nd ed. New York, Springer-Verlag.

Pedroza, L.C.C. e Pedreira C.E., 1998, "O Operador de extensão e Redes Neurais", *Anais do 5o Simpósio Brasileiro de Redes Neurais*, PP 55-60, Belo Horizonte.

Pedroza, L.C.C., 1997, "O Operador de extensão e Redes Neurais", Tese de Doutorado, Pontifícia Universidade Católica, Rio de Janeiro, RJ.

Polak, E., 1971, *Computational Methods in Optimization: A unified approach*. New York and London, Academic Press.

Rumelhart, D.E., McClelland, J.L. and PDP Reserch Group, 1986, "Parallel Distributed Processing : Exploration in the microstruture of cognition", The MIT Press, Volume 1:Foundation.