

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA  
COMPUTAÇÃO**

Giuliano Ferronato

**INTERVALOS DE PREDIÇÃO PARA REDES NEURAIS ARTIFICIAIS  
VIA REGRESSÃO NÃO LINEAR**

Dissertação submetida à Universidade  
Federal de Santa Catarina como parte dos  
requisitos para a obtenção do grau de  
Mestre em Ciência da Computação

Orientador: Prof. Dr. Dalton Francisco de Andrade

Florianópolis, Agosto de 2008



*"A mente que se abre a uma nova idéia  
jamais voltará ao seu tamanho original."*

*Albert Einstein*

## **Agradecimentos**

Agradeço aos meus professores, Dalton e Paulo, que orientaram o desenvolvimento deste trabalho, principalmente pela calma e paciência dispensada em inúmeras horas de reuniões, encontros e discussões, ensinando que para ser realmente um professor é preciso muito mais do que um título de doutor.

Agradeço também a todos os outros professores e profissionais que auxiliaram de forma direta ou indireta na conclusão do trabalho, quer seja em momentos burocráticos e de formalidades (Verinha), ou em momentos de descontração no tradicional cafezinho das 16h (professores Pedro, Adriano, etc...).

Ao CNPq pelo apóio financeiro. E aos amigos, Daniel, Káka, Cauê, Primo Léo enfim, a todos comigo conviveram durante todo o período de estudo.

Pai, mãe, tigão.... e nunca menos importante, pate:

Saibam que, serei sempre grato a vocês !

Não pela conclusão deste mestrado, nem pelos anos de graduação que passei longe de todos, tampouco pelos momentos difíceis, enfermidades, tristezas enfim, onde sempre pude contar com mãos amigas que me incentivam a continuar trilhando caminhos que parecem, por muitas vezes tempestuosos, na certeza de que sempre que for preciso retroceder alguns passos e mudar de direção irei encontrar pessoas que incondicionalmente me apoiarão e continuarão a jornada ao meu lado.

Agradeço, portanto, primeiramente a Deus por me proporcionar a oportunidade de viver ao lado de pessoas tão especiais e únicas, e a vocês que me ensinam cada vez mais o sentido de Viver e Amar.

Meu muito obrigado a todos!

Resumo da dissertação apresentada à UFSC como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências da Computação.

# **INTERVALOS DE PREDIÇÃO PARA REDES NEURAIS ARTIFICIAIS VIA REGRESSÃO NÃO LINEAR**

**Giuliano Ferronato**

Agosto/2008

Orientador: Prof. Dr. Dalton Francisco de Andrade

Área de concentração: Sistemas de Conhecimento

Palavras chave: Mínimos quadrados; Perceptron multicamadas; Predições intervalares;

RESUMO: Este trabalho descreve a aplicação de uma técnica de regressão não linear (mínimos quadrados) para obter predições intervalares em redes neurais artificiais (RNA's). Através de uma simulação de Monte Carlo é mostrada uma maneira de escolher um ajuste de parâmetros (pesos) para uma rede neural, de acordo com um critério de seleção que é baseado na magnitude dos intervalos de predição fornecidos pela rede. Com esta técnica foi possível obter as predições intervalares com amplitude desejada e com probabilidade de cobertura conhecida, de acordo com um grau de confiança escolhido. Os resultados e as discussões associadas indicam ser possível e factível a obtenção destes intervalos, fazendo com que a resposta das redes seja mais informativa e conseqüentemente aumentando sua aplicabilidade. A implementação computacional está disponível em [www.inf.ufsc.br/~dandrade](http://www.inf.ufsc.br/~dandrade).

Abstract of dissertation presented to UFSC as a partial fulfillment of the requirements of the degree of Master in Computation Science

## **PREDICTION INTERVALS FOR ARTIFICIAL NEURAL NETWORKS VIA NONLINEAR REGRESSION**

**Giuliano Ferronato**

August/2008

Advisor: Prof. Dr. Dalton Francisco de Andrade

Major area: Knowledge systems.

Key words: Least square; Multilayer perceptron; Prediction intervals;

Abstract: This work describes the application of a nonlinear regression technique (least squares) to create prediction intervals on artificial neural networks (ANN's). Through Monte Carlo's simulations it is shown a way of choosing the set of parameters (weights) to a neural network, according to a selection criteria based on the magnitude of the prediction intervals provided by the net. With this technique it is possible to obtain the prediction intervals with the desired amplitude and with known coverage probability, according to the chosen confidence level. The associated results and discussions indicate to be possible and feasible to obtain these intervals, thus making the network response more informative and consequently increasing its applicability. The computational implementation is available in [www.inf.ufsc.br/~dandrade](http://www.inf.ufsc.br/~dandrade).

# Sumário

<b>1. Introdução.....</b>	<b>8</b>
1.1. Problema e enfoque .....	9
1.2. Objetivo .....	9
1.2.1. Objetivo geral.....	10
1.2.2. Objetivos específicos.....	10
1.3. Justificativa .....	10
1.4. Limitações .....	10
1.5. Resultados esperados .....	11
1.6. Estrutura de apresentação.....	11
<b>2. Fundamentação teórica.....</b>	<b>13</b>
2.1. Redes neurais artificiais .....	13
2.1.1. Aproximador universal de funções .....	13
2.1.2. Perceptron multicamadas.....	14
2.2. Redes neurais artificiais via regressão não linear .....	15
2.2.1. Estimacão dos parâmetros do modelo de regressão.....	16
2.2.2. Predições intervalares .....	17
<b>3. Simulação de Monte Carlo .....</b>	<b>19</b>
3.1. Etapa A: Criação da base de dados.....	19
3.2. Etapa B: Definição da arquitetura neural utilizada .....	20
3.3. Etapa C: Treinamento de RNA's.....	21
3.4. Etapa D: Definição do critério de seleção composto utilizado .....	21
<b>4. Resultados e discussão.....</b>	<b>23</b>
<b>5. Considerações finais.....</b>	<b>27</b>
<b>6. Referências .....</b>	<b>29</b>
<b>Anexo 1 – Tabela de distribuição t-student .....</b>	<b>32</b>
<b>Anexo 2 – Funções de ativação utilizadas .....</b>	<b>33</b>
<b>Apêndice 1 – Tabela completa de análise da série CU_1 .....</b>	<b>34</b>

# 1. Introdução

O paradigma conexionista, dentro do contexto de inteligência artificial, originou o que se conhece por neurocomputação, disciplina que trata de dispositivos artificiais baseados em sistemas biológicos, em especial o sistema nervoso. Neste sentido acredita-se na possibilidade de construção de máquinas que imitem a estrutura e funcionamento do cérebro biológico, na esperança da emergência de um comportamento inteligente (KOEHLER, 1998); (BITTENCOURT, 2006). Considerada atualmente a principal técnica utilizada em neurocomputação, a teoria de redes neurais artificiais (RNA's), vem se consolidando mundialmente como uma eficiente ferramenta para se lidar com a ampla classe dos, assim chamados, problemas complexos, em que extensas massas de dados devem ser modeladas e analisadas em um contexto multidisciplinar, envolvendo simultaneamente, tanto aspectos estatísticos e computacionais como dinâmicos e de otimização. (BARBOSA, 2004) e (KOVACKS, 2002).

Esta ferramenta tem sido usada com grande eficiência em sistemas que precisam aprender com usuários e adaptar-se a eles, pois ao contrario de programas computacionais comuns, que executam apenas comandos de forma ordinária e fixa, esta possui flexibilidade, sofre modificações enquanto evolui e aprende através de exemplos. É utilizada em um grande número de aplicações, nas mais diversas áreas do conhecimento, tais como, financeira, industrial, médica, etc., com diferentes propósitos, como, por exemplo, *data mining*, classificação, reconhecimento de padrões, segmentação, estimação e predição. (TAFNER, et.al., 1996 apud SANTOS, M. 2006).

Segundo Haykin (1999), deve-se entender as RNA's como uma extensão das convencionais técnicas estatísticas, pela qual o conhecimento empírico sobre um fenômeno físico ou ambiente de interesse pode ser codificado através de treinamento. Neste sentido, uma das principais vantagens na utilização RNA's é a facilidade com que se pode alterar o modelo neural usado para a estimação. De forma que, alterando-se o número de neurônios, de camadas e/ou as funções de ativação da arquitetura, a função de regressão modelada é modificada. E isto, podendo ser feito empiricamente, sem que se faça necessário um amplo conhecimento sobre os modelos de regressão que estão sendo utilizados, simplifica a tarefa de obtenção de um modelo que satisfaça as necessidades específicas da aplicação. No caso, encontrar um modelo matemático capaz



de aprender a relação existente nos dados disponíveis, de forma a poder fazer predições acuradas. (BRAGA; CARVALHO, et.al., 1998), (BARRETO 2001) e (NUNES, 1998).

Este trabalho restringe a aplicação desta ferramenta à tarefa de predição, e interpreta seu funcionamento do ponto de vista de análise de regressão. O fato de interpretar o modelo neural sob o aspecto de um modelo de regressão, possibilita a aplicação de propriedades particulares, as quais não só permitem como também validam a criação de predições intervalares, desde que se respeitem alguns pressupostos e limitações deste tipo de estimação. Esta interpretação se torna possível devido à possibilidade de utilizar determinadas arquiteturas neurais como aproximadores universais de funções.

A tarefa de predição, dentro do contexto da teoria da aproximação, consiste em obter estimativas para a variável de interesse (resposta), a partir de um conjunto de exemplos (observações). (KOVACKS, 2002). De maneira que, o modelo utilizado para simular, ou ainda, reproduzir o comportamento presente nas observações, deve ser capaz de ‘aprender’ a relação existente entre as variáveis preditoras (entradas) e a variável resposta (saída), de forma aproximada. (HAIKYN, 1999).

## 1.1. Problema e enfoque

Considerando apenas a classe das RNA’s aproximadoras universais de funções utilizadas para predição, que é objeto desta pesquisa, sabe-se que grande parte das dúvidas e críticas, quanto a sua utilização e aplicação, são referentes a falta de uma medida de precisão para as predições, e a dificuldade de interpretação dos parâmetros da rede, ou seja, seu conhecimento não é explícito.

Para muitas aplicações a interpretação ou a extração do conhecimento que existe internamente nas RNA’s não é relevante e/ou não se faz necessário. Esta pesquisa enfoca o problema da falta de uma medida de precisão nas predições fornecidas, e propõe a utilização de uma das técnicas da teoria de regressão não linear (mínimos quadrados ordinários) para obtenção da mesma.

## 1.2. Objetivo

Desta forma, destacam-se como:

### 1.2.1. OBJETIVO GERAL

Implementar e disponibilizar uma forma de obtenção de uma medida de precisão para predições de RNA's.

### 1.2.2. OBJETIVOS ESPECÍFICOS

- Estudar a técnica de obtenção de intervalo de predição (mínimos quadrados), oriunda da teoria de análise de regressão, e avaliar sua aplicabilidade em RNA's;
- Aplicar a teoria de análise de regressão ao contexto de RNA's, obter predições intervalares, e indicadores de desempenho.
- Apresentar um critério de seleção de redes (escolha de um ajuste de parâmetros), que toma como base a amplitude dos intervalos de predição obtidos, e não somente suas estimativas.
- Analisar as mudanças, nos indicadores e nos intervalos obtidos, quanto a alterações nas características de nível de ruído inserido nos dados fornecidos para treinamento da rede.
- Disponibilizar uma forma de obtenção de predições intervalares para RNA's em forma de *package*/biblioteca para Matlab<sup>®</sup>.

### 1.3. Justificativa

Em geral as RNA's, utilizadas como aproximadoras universais de funções, fornecem predições pontuais baseadas em um melhor ajuste de seus pesos (parâmetros). Porém a falta de uma medida de qualidade (precisão) aliada às predições faz com que a aplicabilidade desta ferramenta possa ser discutível.

Acredita-se que ao obter uma medida de precisão baseada nas próprias predições fornecidas tradicionalmente pelas RNA's, haja um acréscimo qualitativo sob os resultados, vindo a agregar valor em termos de qualidade e informação contida nas predições. E desta forma auxiliar o utilizador da ferramenta na hora de escolher o ajuste de parâmetros que melhor adapta o modelo escolhido aos dados de seu problema.

### 1.4. Limitações

A pesquisa apresenta algumas limitações:

- Apenas RNA's que são aproximadoras de funções, utilizadas para fins de predição são consideradas, e apenas uma específica arquitetura é utilizada;

- Não são realizadas comparações entre diferentes topologias ou arquiteturas, e não se utiliza a técnica de *Weigth Decay* na estimação dos pesos das redes.
- A estimação dos parâmetros do modelo é realizada apenas por uma dentre as diversas técnicas estatísticas disponíveis;
- São utilizados apenas dados de simulação, não sendo considerados problemas com dados reais;
- Não são feitos testes sobre dados que não correspondam aos pressupostos dos modelos de regressão, afim de “testar a robustez” das predições.

### 1.5. Resultados esperados

Ao obter estimações intervalares, ao invés de simples predições, espera-se estar aumentando a qualidade das informações fornecidas pela rede, uma vez que, juntamente com uma estimativa de um valor ‘provável’ de resposta, correspondente a determinado estímulo apresentado, estar-se-á fornecendo indicadores referentes à margem de erro destas predições, construídas a partir de especificações de níveis de confiança desejados.

Mais do que isso espera-se poder verificar diferenças de magnitude na amplitude dos intervalos obtidos, mediante ao uso de conjuntos de dados com variâncias dos erros diferentes (porém constantes). Pretende-se ainda utilizar dados cuja variância dos erros não é constante, transgredindo desta forma um dos pressupostos dos modelos de regressão, porém sendo justificável em um contexto realista, onde não se conheça a variância dos erros. E desta forma fazer uma ‘suposição’ quanto à possibilidade de utilização destas técnicas perante tal característica dos dados.

### 1.6. Estrutura de apresentação

O segundo capítulo deste trabalho contém a fundamentação teórica que possibilita visualizar a ferramenta de RNA’s como uma aproximadora universal de funções para fins de predição, e explica como obter e interpretar os intervalos de predição. Primeiramente enfatizou-se a propriedade de aproximação universal de funções. Em seguida definiu-se uma arquitetura simples, enfatizando a compreensão do modelo neural sob o ponto de vista de regressão não linear. Por fim, uma vez atendidas as exigências da análise de regressão, mostrou-se como obter predições intervalares assintoticamente válidas para RNA’s.

O terceiro capítulo apresenta uma simulação de Monte Carlo, onde a partir de dados experimentais artificiais e uma arquitetura de RNA definida, é mostrada a aplicação da teoria para obtenção dos intervalos de predição.

No quarto capítulo, são apresentados e discutidos os resultados obtidos com a simulação, onde se enfatiza o fato de que a abordagem deste trabalho é apropriada apenas para dados que possuem erros com variância constante, porém quando utilizada experimentalmente em dados cuja variância não é constante – representando casos onde esta não é conhecida – se apresenta como promissora. São discutidos também outros indicadores obtidos, os quais possibilitam, fornecer uma maneira alternativa de ‘escolha’ da melhor rede treinada. Escolha esta, baseada nas amplitudes dos intervalos obtidos.

E o quinto capítulo resume as principais conclusões do trabalho, bem como recomendações, limitações de uso e aplicabilidade da teoria de regressão não linear às RNA's.

## 2. Fundamentação teórica

O objetivo deste capítulo é fornecer o embasamento teórico necessário, dos pontos de vista de redes neurais e regressão não linear, para permitir vislumbrar a ferramenta de RNA's como aproximadora universal de funções, mostrar como esta pode ser utilizada para fins de predição, e ainda, como obter predições intervalares assintoticamente válidas.

### 2.1. Redes neurais artificiais

O escopo deste trabalho foca o estudo no caso particular de RNA's do tipo *perceptron* multicamadas (MLP)<sup>1</sup>, projetadas e utilizadas para fins de predição. Para maiores detalhes da introdução e revisão histórica do assunto indica-se Kovacks (2002), Haykin (1999) e Widrow (1962).

Dentre as utilidades de uma rede MLP, a que é de principal relevância para o trabalho diz respeito ao fato desta poder ser utilizada como aproximadora universal de funções.

#### 2.1.1. APROXIMADOR UNIVERSAL DE FUNÇÕES

Um aproximador universal de funções é um mecanismo pelo qual pode-se representar qualquer função de uma determinada classe, com precisão pré-estabelecida. Foi provado (apud KOVACKS 2002 e HECHT-NIELSEN 1990) que uma rede neural com três camadas, é um aproximador universal de funções contínuas e limitadas sobre um domínio compacto. De acordo com Bishop (2005), esta propriedade das redes pode ser entendida como a habilidade de ajustar uma grande quantidade de funções a partir de um mesmo modelo. Esta habilidade está relacionada ao fato de que, considerando certas arquiteturas, o relacionamento entre as 'n' entradas e as 'm' saídas da rede definem um mapeamento de um espaço euclidiano  $n$ -dimensional para um  $m$ -dimensional, que é infinita e continuamente diferenciável.

O teorema de Kolmogorov-Nielsen é tido como um teorema de existência, no sentido que fornece uma justificativa matemática para a aplicação de uma RNA no mapeamento de funções (HAYKIN, 1999). Desta forma, as redes neurais podem ser vistas como uma ferramenta prática de uso geral para fazer mapeamentos (lineares ou não) do tipo entrada-saída, sendo ideal para tarefas de predição.

---

<sup>1</sup> do inglês *multilayer perceptron*.

### 2.1.2. PERCEPTRON MULTICAMADAS

Neste trabalho não são abordadas questões relativas à busca por arquiteturas mais apropriadas, sobre nenhum aspecto. O que se propõe é, dada uma arquitetura simples definida e um conjunto de dados, mostrar como é possível fazer previsões intervalares válidas do ponto de vista estatístico. A arquitetura escolhida, graficamente representada pela Figura 1, é uma rede *perceptron* multicamadas unidirecional totalmente conectada, com três camadas de neurônios – entrada, intermediária (oculta) e saída, – com 5, 5 e 1 neurônios respectivamente. Na camada oculta opta-se por utilizar funções de ativação não lineares do tipo tangente hiperbólica. As funções de ativação dos demais neurônios são funções identidade. A descrição matemática das funções utilizadas no modelo pode ser encontrada no Anexo 2.

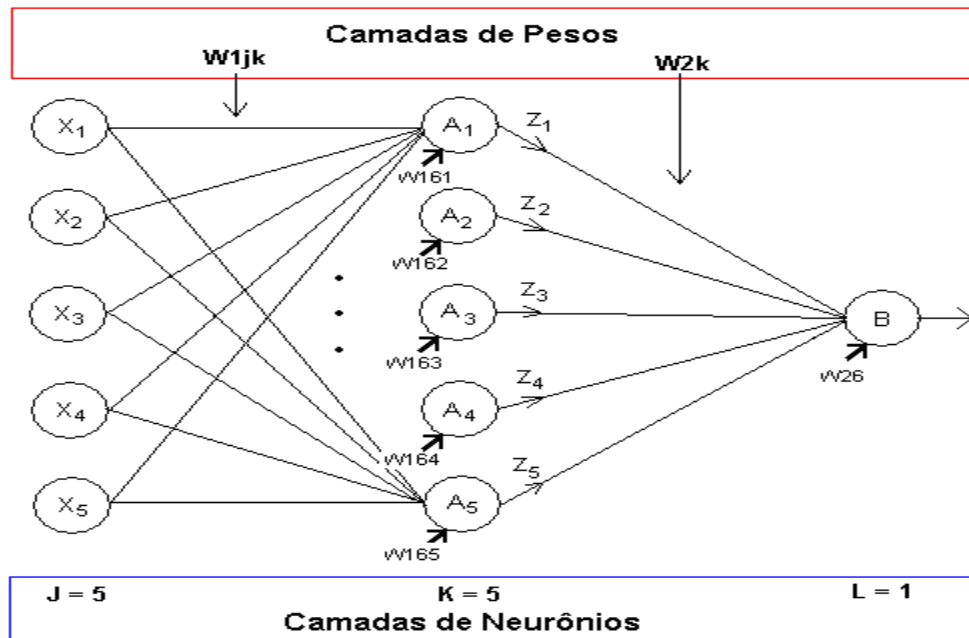


Figura 1 – representação gráfica da arquitetura (5:5:1) proposta para o trabalho.

As funções analíticas correspondentes à arquitetura da Figura 1 podem ser obtidas como segue: A função não linear correspondente ao modelo proposto (saída da rede), corresponde à expressão explícita:

$$f(X, \omega) = \text{lin}(B) = B = \sum_{k=1}^K [\omega_{2k} z_k] + \omega_{26}, \quad (2.1)$$

onde:  $f(X, \omega)$  é o valor de saída da rede (variável dependente), para a entrada  $X$ ,

$\omega$  são os pesos do modelo neural,

$X$  são os valores das entradas (variáveis independentes) e,

$Z_k$  são as saídas dos  $k$ -neurônios da camada intermediária dados por:

$$z_k = \tanh(A_k) = \frac{e^{A_k} - e^{-A_k}}{e^{A_k} + e^{-A_k}}, \quad (2.2)$$

com  $k = 1, 2, \dots, K$ , onde  $\tanh(A_k)$ , representa a função de ativação tangente hiperbólica aplicada a cada neurônio  $A_k$ , que por sua vez corresponde ao somatório ponderado, das entradas  $X$  e suas respectivas conexões (pesos) dados por :

$$A_k = \sum_{j=1}^J [\omega_{1jk} X_j] + \omega_{16k}, \quad (2.3)$$

com  $j = 1, 2, \dots, J$ . É esperado que, através do processo de treinamento (estimação dos pesos), este modelo seja capaz de aprender a relação existente entre as entradas  $X$  e as respostas  $Y$  a ele apresentadas, e desta forma se torne capaz de estimar e fazer previsões acuradas relativas a dados de entradas, conhecidos ou não durante a etapa de treinamento.

## 2.2. Redes neurais artificiais via regressão não linear

Supondo um conjunto de dados de tamanho  $n$ , que represente a relação existente entre uma (ou mais) variável(is) independente(s)  $X_i$ , e uma (ou mais) variável(is) dependente(s)  $Y_i$ , um modelo de regressão não linear que representa tal relação segue a seguinte forma básica:

$$Y_i = f(X_i, \omega) + \varepsilon_i, \quad (2.4)$$

com  $i=1, \dots, n$ , onde  $X_i$  e  $Y_i$  correspondem aos valores de  $X$  e  $Y$  para o  $i$ -ésimo elemento do conjunto de dados disponíveis,  $f(X_i, \omega)$  é a equação de regressão (parte determinística do modelo) que supostamente rege a relação entre as variáveis dependentes e independentes, e  $\varepsilon_i$  é o erro aleatório associado (existente por ser uma relação estatística);  $\omega$  é o vetor de parâmetros do modelo. Esta equação de regressão é tida como linear, ou não linear, dependendo da complexidade relativa à disposição dos parâmetros no modelo.

Em decorrência da característica de o modelo neural poder ser interpretado como um aproximador universal de funções, sua equação explícita (eq. 2.1) passa a representar a parte determinística do modelo de regressão (eq. 2.4). Permitindo desta forma, a utilização das propriedades e técnicas da análise de regressão em estimações, tanto na fase de estimação dos parâmetros do modelo (treinamento) quanto no fornecimento de previsões intervalares (generalização). Para que esta utilização seja assintoticamente validada, é preciso que se atendam os pressupostos deste tipo de

modelo: Aleatoriedade dos resíduos: Os resíduos do modelo devem se apresentar de forma aleatória, segundo os requisitos básicos de homocedasticidade (variância constante), independência e não auto-correlação; e Não multicolinearidade das variáveis preditoras: Significando que, não podem existir relações lineares entre as variáveis independentes do modelo de regressão. Outra característica que não é exigência para estimação dos parâmetros, porém, se faz necessária quando se deseja criar previsões intervalares para um modelo, é a Normalidade dos resíduos: ou seja, pressupõe-se que os resíduos, além de serem aleatórios e independentes sigam uma distribuição aproximadamente normal. Uma maneira de verificar se os dados atendem tais requisitos é recorrer a ferramentas como análise de resíduos e análise de correlação. Detalhes podem ser obtidos em Neter et al. (1996).

### 2.2.1. ESTIMAÇÃO DOS PARÂMETROS DO MODELO DE REGRESSÃO

Em redes do tipo MLP, para obtenção dos pesos do modelo neural freqüentemente utiliza-se o paradigma de aprendizado por correção de erro, sendo que o algoritmo de ajuste de parâmetros mais utilizado é o de retropropagação do erro. (BISHOP, 2005). Este algoritmo, tido como fortemente supervisionado segundo classificação proposta por Jacob (2003), utiliza observações na forma de pares entrada-saída  $[X_i, Y_i]$ , e se baseia no método de mínimos quadrados para estimação dos pesos. Segundo Neter et al. (1996), o teorema das amostras grandes afirma que, quando utilizando amostras suficientemente grandes, o método dos mínimos quadrados é capaz de fornecer estimadores  $\hat{\omega}$  aproximadamente sem viés e com variância quase nula. Tais estimadores podem ser encontrados analítica ou numericamente, minimizando-se a soma de quadrados da função do erro para todas as observações disponíveis, como segue:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (2.5)$$

onde  $\hat{y}_i$  é o valor obtido (estimado) pelo modelo neural, e  $y_i$  é o valor observado. Em redes neurais, por tipicamente serem tratados problemas complexos, muitas vezes não lineares, prefere-se a utilização de procedimentos numéricos, freqüentemente iterativos, para a solução das equações normais de regressão. Devido ao fato destes procedimentos utilizarem na solução, por exemplo, expansões em séries de *Taylor* ou derivadas de primeira e segunda ordem, demandam muito mais tempo e recursos computacionais. (KOVACKS, 2002) e (HAYKIN, 1999). Deve-se enfatizar que, a utilização de métodos



iterativos, por ser dependente dos valores iniciais atribuídos aos parâmetros, não garante que o algoritmo venha a convergir em um resultado mínimo, e mesmo que isso aconteça, não garante que este seja um mínimo global da função de regressão. Para maiores detalhes referentes ao método de mínimos quadrados, o algoritmo backpropagation, ou procedimentos numéricos recomenda-se De Azevedo (1997), Braga et al. (1998), e Neter et al. (1996).

### 2.2.2. PREDIÇÕES INTERVALARES

De acordo com Hwang e Ding (1997), a teoria estatística para criação de intervalos de confiança para os parâmetros destes modelos não é apropriada, pois estes não são identificáveis. Porém os autores provam que utilizando funções de ativação específicas (contínuas, diferenciáveis, e de preferência não descendentes em todo seu domínio) as predições preservam as características da teoria assintótica, permitindo a obtenção de predições intervalares, ao invés de apenas predições pontuais.

Uma predição intervalar  $\hat{Y}_{n+1|X_{n+1}}$ <sup>2</sup> é calculada através da estimativa da variância do erro do modelo (MSE) e da variância da resposta do modelo (VAR), dados respectivamente por 2.6 e 2.7:

$$MSE = \frac{SSE}{n-p} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-p}, \quad (2.6)$$

onde  $n$  é o numero de observações disponíveis para treinamento, e  $p$  é a quantidade de parâmetros do modelo,

$$VAR(\hat{Y}_{n+1}) \approx \sigma^2 \mathbf{g}_{n+1}^t (\mathbf{J}^t \mathbf{J})^{-1} \mathbf{g}_{n+1}, \quad (2.7)$$

onde  $\mathbf{J}$  é a matriz das derivadas de primeira ordem da estimação  $\hat{Y}_i|X_i$ , com  $i=1,2,\dots,n$ , e  $\mathbf{g}_{n+1}$  é um vetor de derivadas de primeira ordem da predição pontual  $\hat{Y}_{i+1}|X_{i+1}$ , ambos em relação a todos os parâmetros  $\hat{\omega}_j$ , com  $j=1,\dots,p$ . As  $i$ -ésimas entradas destes elementos são consistentemente calculadas por 2.8 e 2.9 respectivamente:

$$J_{ij} = \frac{\partial(\hat{Y}_i | \mathbf{X}_i, \hat{\omega})}{\partial \hat{\omega}_j}, \quad (2.8)$$

$$\mathbf{g}_{n+1} = \frac{\partial(\hat{Y}_{n+1} | \mathbf{X}_{n+1}, \hat{\omega})}{\partial \hat{\omega}_j}, \quad (2.9)$$

---

<sup>2</sup> onde 'n+1' é uma notação que representa um dado novo (presente ou não no conjunto de treinamento), e  $\hat{Y}|X$  representa uma estimação  $\hat{Y}$  dado que  $X$  é o conjunto de valores de entrada da rede.

E assim, baseado no teorema das amostras grandes, a seguinte equação apresenta uma estimativa intervalar assintoticamente válida para  $\hat{Y}_{n+1}|X_{n+1}$ .

$$\hat{Y}_{n+1} \pm t \cdot \sqrt{MSE \cdot \left[ 1 + \mathbf{g}_{n+1}^t (\mathbf{J}^t \mathbf{J})^{-1} \mathbf{g}_{n+1} \right]}, \quad (2.10)$$

onde  $t$  representa o valor crítico obtido pela distribuição t-student<sup>3</sup> com  $n-p$  graus de liberdade e grau de confiança  $\gamma$  desejado, escolhido arbitrariamente (geralmente entre 95% e 99%). A magnitude de  $t$ , e consequentemente a amplitude do intervalo obtido são diretamente proporcionais a este valor escolhido, de forma que, quanto maior o nível de acerto desejado para o intervalo, maior é o valor crítico a ser utilizado e consequentemente maior tende a ser o intervalo obtido (menos preciso).

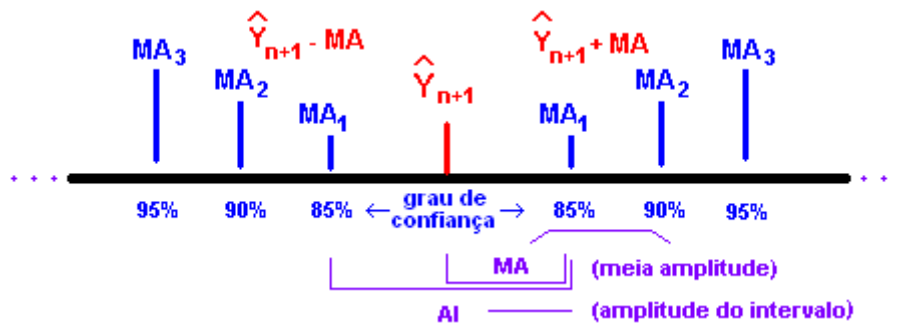


Figura 2: Representação gráfica das características de um intervalo de predição.

Um intervalo de predição (IP) é então obtido através da estimação pontual  $\hat{Y}_{n+1}|X_{n+1}$ , e da meia amplitude do intervalo (MA), que corresponde ao valor somado e/ou subtraído da predição, na forma  $\hat{Y}_{n+1} \pm MA$ .

De forma que, se os parâmetros do modelo são eficientemente estimados, a probabilidade de cobertura (PC) dos intervalos criados, ou seja, a probabilidade de que uma observação  $\hat{Y}_{n+1}|X_{n+1}$  realmente esteja compreendida no intervalo estimado, para todos os dados de teste, tende a corresponder ao valor escolhido para o nível de confiança  $\gamma$ . (PAPADOPOULUS, 2001).

<sup>3</sup> A tabela t-student pode ser consultada no Anexo 1. É considerada tal distribuição para as predições, uma vez que se utiliza uma estimação da variância do erro (MSE), e não seu real valor  $\sigma^2$ .

### 3. Simulação de Monte Carlo

Monte Carlo é um método estatístico utilizado em simulações estocásticas. Pode ser utilizado como forma de obter aproximações numéricas de funções complexas. Tipicamente envolve a geração de observações com alguma distribuição de probabilidades específica, e o uso de técnicas de amostragem, para então aproximar a função de interesse de acordo com a metodologia escolhida.

O que se propõe é utilizar esta metodologia, em conjunto com o modelo neural proposto na seção 2.1 e analisado na seção 2.2, para conduzir um experimento onde seja possível obter e analisar, intervalos de predição para RNA's quando utilizadas com objetivo de aproximar uma função complexa e fazer predições.

Para que isto se faça possível, propõe-se uma simulação em 4 etapas (A, B, C, D), onde resumidamente, na etapa A: são criadas três bases estocásticas de dados, com diferentes características em relação ao termo de erro inserido no modelo; na etapa B: é definida a arquitetura (única) utilizada pelas redes em todo o processo de simulação; a etapa C: apresenta a metodologia utilizada para o treinamento de RNA's, segundo arquitetura definida, para cada base de dados; e a etapa D: apresenta um critério composto para seleção da rede que melhor reproduz o conhecimento em cada base de dados.

#### 3.1. Etapa A: Criação da base de dados

Seguindo Papadopoulos (2001), para possibilitar possíveis comparações entre intervalos de predição obtidos a partir de dados com características distintas, propõe-se a criação de três bases de dados. Todas as bases são criadas a partir da equação (3.1) e um termo de erro aleatório, seguindo o modelo básico (2.4), transcrito abaixo:

$$Y_i = f(X_i, \omega) + \varepsilon_i,$$

sendo que,  $Y$  é a variável dependente, e  $f(X_i, \omega)$  é dado por:

$$f(X_i, \omega) = 10\text{sen}(\pi X_1 X_2) + 20(X_3 - 0,5)^2 + 10X_4 + 5X_5, \quad (3.1)$$

com  $i=1,2,\dots,n$ , onde os valores de  $X_1, X_2, X_3, X_4$  e  $X_5$  (variáveis independentes) são concebidos aleatoriamente a partir de uma distribuição uniforme no intervalo  $[0,1]$ . O termo do erro  $\varepsilon_i$  também é obtido de forma aleatória, a partir de uma distribuição normal com média zero e variância  $\sigma^2$  constante. Propõe-se que se utilize, por exemplo,

três níveis de variância  $\sigma^2$ , 1, 3 e 5. Neste trabalho, são utilizadas  $n = 1000$  observações. Para permitir uma comparação mais consistente, optou-se por utilizar exatamente as mesmas variáveis independentes  $\mathbf{X}_i$ , nas três bases de dados, porém cada uma com valores distintos de variância. E isto implica em diferentes valores gerados para os erros aleatórios e para a variável dependente.

Para utilização em redes neurais, a base de dados pode ser dividida em duas partes (considerando treinamento sem validação), base de treinamento e base de testes. Opta-se por utilizar como séries de treinamento as bases de treinamento inteiras ao invés de amostras. Desta forma a série de treinamento CU\_1, que possui  $\sigma^2 = 1$ , corresponde a própria base de treinamento CU\_1, e da mesma forma as séries CU\_3 ( $\sigma^2 = 3$ ) e CU\_5 ( $\sigma^2 = 5$ ). Em se tratando de um experimento artificial, onde os dados foram criados de uma só vez, a divisão pode ser feita simplesmente separando os dados disponíveis em duas partes iguais, sem necessitar aleatoriedade no processo de separação. Assim, são considerados os 500 primeiros casos de cada base de dados, como sendo pertencentes à base de treinamento, e os 500 restantes, pertencentes à base de testes.

Espera-se que esta variabilidade de características de dados utilizados para o treinamento das redes, interfira nas predições e conseqüentemente na magnitude dos intervalos de predição.

### 3.2. Etapa B: Definição da arquitetura neural utilizada

Segundo os propósitos do trabalho, todos os procedimentos relativos à utilização das RNA's são feitos utilizando-se sempre a mesma arquitetura neural. Por conveniência, a arquitetura escolhida é a mesma apresentada na seção 2.1 e analisada na seção 2.2. Esta arquitetura é graficamente representada pela figura 1, na seção 2.1.2, e sua equação explícita de saída corresponde à equação (2.1).

Para a simulação define-se que os pesos são estimados por treinamento sem validação, a partir das séries de treinamento definidas, utilizando o método dos mínimos quadrados (eq. 2.4) através do algoritmo de treinamento *backpropagation*, utilizado em composição com o algoritmo de *Marquardt*. As predições intervalares podem ser calculadas pela (eq. 2.10), através da estimativa da variância do erro do modelo (eq. 2.6) e da estimativa da variância da resposta do modelo (eq. 2.7):

### 3.3. Etapa C: Treinamento de RNA's

Seguindo Papadopoulos (2001), e De Veaux (1997), devido à natureza dos procedimentos numéricos utilizados para estimação dos pesos das redes serem dependentes das condições iniciais estipuladas para os mesmos ( $\omega_0$ ), opta-se por treinar 100 redes neurais para cada série de treinamento (CU1, CU3 e CU5), a fim de poder analisá-las sob os pontos de vista de seus resultados e diferenças, em termos de assertividade e precisão obtidas. Os valores iniciais dos pesos, utilizados no início de cada treinamento, são gerados aleatoriamente, dentro do intervalo [0,1].

### 3.4. Etapa D: Definição do critério de seleção composto utilizado

Considerando todas as redes que foram treinadas para certo conjunto de treinamento, interessa escolher a rede que melhor aprendeu a relação existente entre as variáveis de entrada e saída. Esta escolha é importante, pois cada conjunto de pesos encontrado por cada uma das redes corresponde a um mapeamento (aprendizado) diferente, mais ou menos eficiente. É essencial que se procure encontrar a rede que melhor se adapte ao ambiente de interesse, para permitir melhores generalizações, otimizando os resultados obtidos com a utilização da ferramenta.

Lembrando que, a probabilidade de cobertura (PC) diz respeito ao número de observações de teste que são acertadamente preditas, ou seja, observações que efetivamente estão compreendidas dentro do intervalo estimado. Lembrando também que o valor médio das meias amplitudes de uma rede ( $\overline{MA}$ ) corresponde à média aritmética dos valores individuais de cada meia amplitude (MA), estimada para cada observação (ilustrada pela figura 2, na seção 2.2). Ao utilizar apenas o indicador PC como critério de escolha da melhor rede, priorizam-se redes que acertem mais, não considerando a amplitude dos intervalos obtidos. Ao passo que, utilizando apenas o critério  $\overline{MA}$ , priorizam-se redes que apresentem predições mais 'precisas', porém não garantindo acerto. O que se propõe é a utilização de um critério de seleção (CS) composto, com o objetivo de otimizar ou equilibrar o grau de acerto *versus* a precisão das redes, possibilitando a escolha de uma rede que forneça os intervalos de predição com maior acerto possível desde que sua precisão não seja reduzida além de um limite 'aceitável'. Devido ao fato de o conceito de aceitável ser subjetivo e relativo à aplicação considerada, entende-se que este limite deve ser mantido como variável, porém sempre

calculado em relação ao menor valor das  $\overline{MA}$  ( $\text{Min}\overline{MA}$ ), encontrado entre todas as redes treinadas. Sendo assim, o critério de seleção prioriza redes que obtenham maior PC dentre as que possuem  $\overline{MA}$  aceitável, segundo o valor estabelecido para o critério. E assim, o limite aceitável para seleção das redes pode ser obtido segundo a equação:

$$LA = \text{Min}\overline{MA} + CS * \text{Min}\overline{MA} \quad (3.2)$$

onde, LA é o limite máximo para valores de  $\overline{MA}$ , que permite que uma rede possa ser classificada como aceitável, podendo assim vir a ser escolhida como a melhor dentre todas as redes treinadas para determinada série de dados.

Por exemplo: (acompanhar na tabela encontrada no Apêndice 1)

Supondo o treinamento de 100 redes válidas, a que apresenta valor mínimo encontrado para  $\overline{MA}$  foi a octagésima rede treinada (80ª) com valor de 2,129. Por outro lado, a 24ª e a 30ª rede treinada obtiveram valores máximos de PC, chegando a 96,20 %.

Segundo o critério de seleção composto apresentado, considerando  $CS = 5\%$ , dentre as 100 redes treinadas 18 delas poderiam ser escolhidas como melhores por apresentarem valor de  $\overline{MA}$  até 5% maior do que valor mínimo encontrado. Dentre estas, a que apresenta maior valor de PC considerando a ordem de treinamento é a 37ª com 95,20%, e então esta rede é escolhida como a melhor. O valor utilizado para o critério poderia ser maior ou menor, alterando a flexibilidade na escolha, por exemplo considerando-se redes com 10%, ou 50% maiores do que o mínimo.

## 4. Resultados e discussão

De acordo com o propósito e especificações da simulação de Monte Carlo utilizada, o experimento foi conduzido, sendo que os principais resultados são apresentados na tabela 1:

Tabela 1 – Relação das principais características e indicadores de desempenho das redes selecionadas, em cada série de dados, considerando o critério CS = 5%:

Serie Dados	Nº Redes Selecionadas	POSTO (RANK)					$\overline{MA}$	PC	Min PC	Max PC
		Treinamento	MSE	SSE	$\overline{MA}$	PC				
CU_1	18	37 <sup>a</sup>	10 <sup>a</sup>	6 <sup>a</sup>	5 <sup>a</sup>	20 <sup>a</sup>	2,13	95,20	2,07	96,2
CU_3	7	91 <sup>a</sup>	4 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	11 <sup>a</sup>	6,44	96,00	6,33	96,8
CU_5	8	64 <sup>a</sup>	1 <sup>a</sup>	1 <sup>a</sup>	1 <sup>a</sup>	19 <sup>a</sup>	9,59	94,00	9,59	95,6

Cada uma das linhas da tabela apresenta os principais indicadores de desempenho da melhor rede de cada série de treinamento, escolhidas segundo critério de seleção composto considerando CS=5%. Para obter tais valores, foram utilizados todos os dados disponíveis para teste, correspondentes a cada uma das três séries.

A segunda coluna indica que, para cada série, 18, 7 e 8 redes obtiveram valor de meia amplitude média ( $\overline{MA}$ ) aceitável, ou seja, até 5% maior do que o mínimo valor encontrado dentre todas as redes treinadas respectivamente em cada série. As colunas Ranking se referem à classificação da rede escolhida quanto a uma ordenação específica. Por exemplo, a melhor rede da série CU\_1 (37<sup>a</sup> na ordem de treinamento), seria classificada como sendo a 10<sup>a</sup> em termos de MSE, 6<sup>a</sup> em MSSE, 5<sup>a</sup> em  $\overline{MA}$  e 20<sup>a</sup> em PC. As colunas  $\overline{MA}$  e PC se referem aos valores obtidos pela rede escolhida, e as últimas duas colunas se referem respectivamente aos valores, mínimo e máximo obtidos para  $\overline{MA}$  e PC dentre todas as 100 redes treinadas em cada série de treinamento.

As tabelas 2, 3 e 4 resumem as principais características e indicadores quanto a escolha de redes utilizando diversos valores para o critério CS, nas séries CU\_1, CU\_3 e CU\_5 respectivamente.

Considerando a série CU\_1, dentre todas as redes treinadas, os valores mínimos e máximos encontrados para PC foram respectivamente 91,20% e 96,20%; e para  $\overline{MA}$ , 2,07 e 15,90. Na tabela 2 são resumidos os indicadores encontrados para a melhor rede considerando o critério composto quando especificado como 5, 10, 15, 20 e 50%.

De forma a exemplificar e simplificar a compreensão do processo de escolha da melhor rede segundo dados de uma série específica encontra-se no anexo 3 a tabela completa de valores de todos os indicadores em relação às 100 redes treinadas para a série CU\_1. Para as demais séries analisadas, CU\_3 e CU\_5, o procedimento é o mesmo.

Tabela 2 – Relação das principais características e indicadores de desempenho das redes selecionadas para série CU\_1 de acordo com diversos valores utilizados para o critério CS.

CS	Nº Redes Selecionadas	POSTO (RANK)					$\overline{MA}$	PC
		Treinamento	MSE	SSE	$\overline{MA}$	PC		
1%	1	80 <sup>a</sup>	1 <sup>a</sup>	1 <sup>a</sup>	1 <sup>a</sup>	53 <sup>a</sup>	2,07	94,40
5%	18	37 <sup>a</sup>	10 <sup>a</sup>	6 <sup>a</sup>	5 <sup>a</sup>	20 <sup>a</sup>	2,13	95,20
10%	28	37 <sup>a</sup>	10 <sup>a</sup>	6 <sup>a</sup>	5 <sup>a</sup>	20 <sup>a</sup>	2,13	95,20
15%	34	68 <sup>a</sup>	32 <sup>a</sup>	31 <sup>a</sup>	29 <sup>a</sup>	8 <sup>a</sup>	2,30	95,60
20%	34	68 <sup>a</sup>	32 <sup>a</sup>	31 <sup>a</sup>	29 <sup>a</sup>	8 <sup>a</sup>	2,30	95,60
50%	39	68 <sup>a</sup>	32 <sup>a</sup>	31 <sup>a</sup>	29 <sup>a</sup>	8 <sup>a</sup>	2,30	95,60

Considerando agora a série CU\_3, os valores mínimos e máximos encontrados para PC foram 92,20% e 96,80%, e para  $\overline{MA}$  6,33 e 37.686,63 respectivamente.

Este valor exorbitante encontrado reafirma o fato de necessitar-se de critérios de escolha de redes que considerem a precisão das predições, por exemplo em termos de magnitudes dos intervalos obtidos, e não somente a assertividade das mesmas. Esta exorbitância esta relacionada a algumas poucas observações que, por não terem sido bem assimiladas no treinamento, geram predições intervalares com amplitude extrema, fazendo com que a média dentre as predições se eleve bruscamente.



Tabela 3 – Relação das principais características e indicadores de desempenho das redes selecionadas para série CU\_3 de acordo com diversos valores utilizados para o critério CS.

CS	Nº Redes Selecionadas	POSTO (RANK)					MA	PC
		Treinamento	MSE	SSE	MA	PC		
1%	2	80 <sup>a</sup>	1 <sup>a</sup>	6 <sup>a</sup>	2 <sup>a</sup>	45 <sup>a</sup>	6,36	95,00
5%	7	91 <sup>a</sup>	4 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	11 <sup>a</sup>	6,44	96,00
10%	17	81 <sup>a</sup>	23 <sup>a</sup>	23 <sup>a</sup>	16 <sup>a</sup>	7 <sup>a</sup>	6,92	96,20
15%	35	48 <sup>a</sup>	26 <sup>a</sup>	11 <sup>a</sup>	20 <sup>a</sup>	4 <sup>a</sup>	6,98	96,40
20%	53	32 <sup>a</sup>	62 <sup>a</sup>	58 <sup>a</sup>	43 <sup>a</sup>	3 <sup>a</sup>	7,45	96,60
50%	87	9 <sup>a</sup>	60 <sup>a</sup>	24 <sup>a</sup>	68 <sup>a</sup>	2 <sup>a</sup>	7,94	96,80

E finalmente, dentre as 100 redes treinadas com dados cuja variância dos erros é igual a 5, observou-se como valores mínimos e máximos em relação a PC 90,20% e 95,60%; e em relação a MA 9,59 e 1.256,89.

Tabela 4 – Relação das principais características e indicadores de desempenho das redes selecionadas para série CU\_5 de acordo com diversos valores utilizados para o critério CS.

CS	Nº Redes Selecionadas	POSTO (RANK)					MA	PC
		Treinamento	MSE	SSE	MA	PC		
1%	1	64 <sup>a</sup>	1 <sup>a</sup>	1 <sup>a</sup>	1 <sup>a</sup>	19 <sup>a</sup>	9,59	94,00
5%	8	64 <sup>a</sup>	1 <sup>a</sup>	1 <sup>a</sup>	1 <sup>a</sup>	19 <sup>a</sup>	9,59	94,00
10%	25	53 <sup>a</sup>	48 <sup>a</sup>	20 <sup>a</sup>	20 <sup>a</sup>	2 <sup>a</sup>	10,46	95,20
15%	51	53 <sup>a</sup>	48 <sup>a</sup>	20 <sup>a</sup>	20 <sup>a</sup>	2 <sup>a</sup>	10,46	95,20
20%	69	53 <sup>a</sup>	48 <sup>a</sup>	20 <sup>a</sup>	20 <sup>a</sup>	2 <sup>a</sup>	10,46	95,20
50%	82	96 <sup>a</sup>	98 <sup>a</sup>	30 <sup>a</sup>	79 <sup>a</sup>	1 <sup>a</sup>	12,29	95,60

Pode-se perceber que a quantidade de redes selecionadas (que atendem ao critério de seleção), passíveis a serem eleitas as melhores em suas séries, é diretamente proporcional ao aumento do valor percentual do critério de seleção. Neste sentido quanto maior é o CS, mais flexível é a rede. Priorizando assim redes cuja probabilidade de cobertura é maior (mais assertividade), em detrimento do valor médio dos intervalos

de amplitude encontrados (menos precisão), e conseqüentemente, dos valores de MSE e SSE que são intimamente relacionados ao  $\overline{MA}$ .

Cabe ressaltar que, mesmo que o critério venha a ser definido com um valor alto de flexibilidade (ex.: 100%), a rede escolhida como melhor não necessariamente obterá valor elevado em relação a  $\overline{MA}$ , uma vez que este é calculado em relação ao menor valor encontrado para a magnitude dos intervalos. Fica a critério do utilizador da ferramenta, ajustar o critério para mais ou menos restritivo, dependendo dos resultados encontrados por seus modelos em relação aos dados disponíveis, bem como das características e exigências do ambiente de aplicação.

## 5. Considerações finais

Com este trabalho, foi possível demonstrar a factibilidade de se obter predições intervalares para RNA's utilizando a técnica de mínimos quadrados. Estes intervalos podem ser criados de acordo com características de desempenho específicas, aumentando ainda mais a flexibilidade desta ferramenta. Foi visto que, mesmo quando o critério de seleção utilizado foi estabelecido para ser bem restritivo, definindo os menores intervalos preditivos encontrados, a probabilidade de cobertura das redes escolhidas foi maior do que o nível de confiança  $\gamma = 95\%$  especificado para criação dos mesmos. Uma das características constatadas, é que existe uma relação diretamente proporcional entre a variabilidade dos erros e o tamanho dos intervalos obtidos.

Não somente as estatísticas de  $\overline{MA}$  e PC foram observadas, mas também outros indicadores de desempenho, tais como, as classificações das redes (*rankings*) em relação a quatro indicadores de desempenho diferentes, e ainda estatísticas como os mínimos e máximos obtidos para cada indicador. Tudo isso vem a auxiliar na escolha da rede que melhor se adapta ao ambiente da aplicação.

Bem ou nem tão bem ajustados, para a utilização da técnica de RNA's sempre um conjunto de parâmetros deve ser escolhido. Este deve representar a melhor rede encontrada pelo processo de treinamento. O critério composto de seleção de redes apresentado se mostra eficiente, uma vez que torna a resposta da rede mais informativa, ajudando o utilizador da ferramenta na tomada da decisão de qual conjunto de parâmetros deve ser utilizado, ou ainda, se nenhum dos conjuntos obtidos representa os dados disponíveis de forma satisfatória.

Alguns testes foram efetuados treinando e testando redes com dados cuja distribuição dos erros não é constante. Isto contraria um dos pressupostos dos modelos de regressão, mas é entendido como justificável em um contexto real, onde não se saiba qual é a distribuição associada aos dados disponíveis. Neste caso, embora existam, outras técnicas mais adequadas para estimação dos parâmetros, como, por exemplo, mínimos quadrados ponderados, os resultados preliminares seguem os padrões apresentados na simulação com dados cuja variância é constante, e as predições obtidas possuem amplitude compatível às apresentadas no capítulo anterior. Porém sugere-se planejar simulações mais criteriosas visando objetivamente o estudo deste comportamento, sendo indicado em um estudo futuro.

Neste sentido, indica-se ainda conduzir experimentos para avaliar as implicações de alterações na quantidade de dados utilizada para treinamento e teste. Outras propostas que são vislumbradas dizem respeito à arquitetura utilizada, variando números de neurônios, camadas, métodos de aprendizagem, algoritmos de treinamento, etc.

## 6. Referências

BARBOSA, P.A.. Mecanismo de Adaptação baseado em Redes Neurais Artificiais para Sistemas Hipermédia Adaptativos. Florianópolis. Tese (Doutorado em Engenharia Elétrica) – Instituto de Engenharia Biomédica, Universidade Federal de Santa Catarina, 2004.

BARRETO, J. M. Inteligência Artificial no Limiar do Século XXI. 3ª Edição. Florianópolis: Editora da UFSC, 2001.

BISHOP, C. M. Neural networks for pattern recognition. 9. ed. New York: Oxford University Press, 2005.

BITTENCOURT, Guilherme. Inteligência artificial: ferramentas e teorias. 3. ed. rev. Florianópolis: Ed. da UFSC, 2006. 371p. ISBN 8532801382.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDEMIR, T. B. Fundamentos de Redes Neurais Artificiais. Rio de Janeiro, 1998..

CAETANO, C. Redes Neurais Artificiais e Algoritmo Genético no estudo de Sistemas Quânticos. São José dos Campos. Dissertação (Mestrado em Física) – Instituto Tecnológico de Aeronáutica. 2005

De AZEVEDO, F. M.; Uma proposta de modelos formais de neurônios e redes neurais artificiais. In: Congresso Brasileiro de Redes Neurais, *Anais*, p. 503-514, Florianópolis, 1997.

HAYKIN, S. S. Neural networks: a comprehensive foundation. 2. ed. New York: Prentice Hall, 1999.

HECHT-NIELSEN, R. Neurocomputing. Reading, USA: Addison-wesley, 1990.

HWANG J. T.; DING A. A. Prediction intervals for artificial neural networks. Journal of the American Statistical Association, New York, v. 92 n. 438, p.748-757, jun. 1997.

JACOB, M. A. Classificação Supervisionada de imagens SAR via Redes Neurais Artificiais. São José dos Campos. Dissertação (Mestrado em Engenharia Eletrônica) – Instituto Tecnológico de Aeronáutica, 2003.

KOEHLER, C. Uma abordagem Probabilística para Sistemas Especialistas. Florianópolis. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Santa Catarina, 1998.

KOVACKS, Zsolt László. Redes neurais artificiais: fundamentos e aplicações. 2. ed. rev. e ampl. São Paulo: Acadêmica, 2002. 174p. ISBN 8586396028

MALINVERNI, M. S. Sistema hipermídia sobre câncer de colo de útero com interface adaptativa usando redes neurais artificiais MLP e sistema especialista. Florianópolis. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, 2006.

NETER, J.; Kutner, M.; H. Wasserman, W.; Nachtsheim, C. J. Applied Linear Regression Models. 3.ed. USA: Irwin, Inc, 1996.

NUNES, C. S. L. Análise e Síntese de Estratégias de Aprendizado para Redes Neurais Artificiais. Campinas. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Estadual de Campinas, 1998.

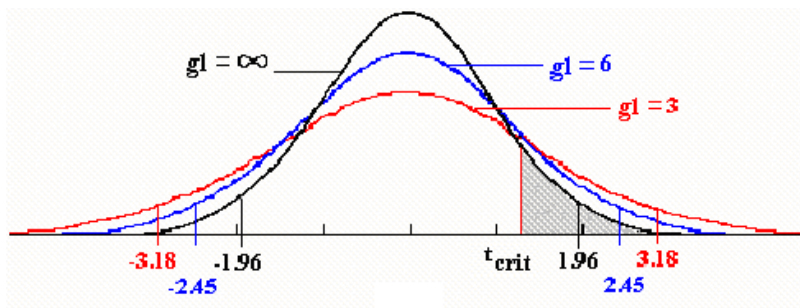
PAIM, R. L. Sistema hipermídia sobre câncer de mama com interface adaptativa usando redes neurais artificiais MLP e IAC. Florianópolis. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, 2006.

PAPADOPOULOS, Georgios; EDWARDS, Peter J.; MURRAY, Alan, F. Confidence estimation methods for neural networks: a practical comparison, IEEE Transactions on neural networks, v. 12, n. 6, p. 1278-1287, nov. 2001.

VEAUX, Richard D. de; UNGAR Lyle H.; SCHUMI Jennifer; SCHWEINSBERG, Jason. Prediction intervals for neural networks via nonlinear regression, *Technometrics*, New York, v. 40, n. 4, p.273-282, nov. 1998.

WIDROW, B., Generalization and Information Storage in Networks of Adaline Neurons, Em Yovitz, M., Jacobi, G., Goldstein, G., editors, *Self-Organizing Systems*, pp. 435-461, Spartan Books, Washington DC, 1962.

## Anexo 1 – Tabela de distribuição t-student

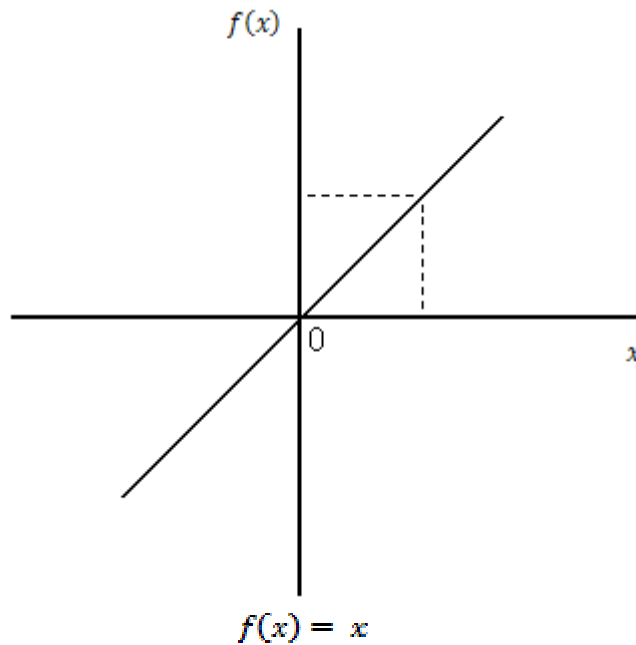


Gl (n-1)	0,2500	0,1000	0,0500	0,0250	0,0100	0,0083	0,0050
1,0000	1,0000	3,0780	6,3140	12,7060	31,8210	38,3430	63,6560
2,0000	0,8160	1,8860	2,9200	4,3030	6,9650	7,6640	9,9250
3,0000	0,7650	1,6380	2,3530	3,1820	4,5410	4,8640	5,8410
4,0000	0,7410	1,5330	2,1320	2,7760	3,7470	3,9660	4,6040
5,0000	0,7270	1,4760	2,0150	2,5710	3,3650	3,5380	4,0320
6,0000	0,7180	1,4400	1,9430	2,4470	3,1430	3,2910	3,7070
7,0000	0,7110	1,4150	1,8950	2,3650	2,9980	3,1300	3,4990
8,0000	0,7060	1,3970	1,8600	2,3060	2,8960	3,0180	3,3550
9,0000	0,7030	1,3830	1,8330	2,2620	2,8210	2,9360	3,2500
10,0000	0,7000	1,3720	1,8120	2,2280	2,7640	2,8720	3,1690
11,0000	0,6970	1,3630	1,7960	2,2010	2,7180	2,8220	3,1060
12,0000	0,6950	1,3560	1,7820	2,1790	2,6810	2,7820	3,0550
13,0000	0,6940	1,3500	1,7710	2,1600	2,6500	2,7480	3,0120
14,0000	0,6920	1,3450	1,7610	2,1450	2,6240	2,7200	2,9770
15,0000	0,6910	1,3410	1,7530	2,1310	2,6020	2,6960	2,9470
16,0000	0,6900	1,3370	1,7460	2,1200	2,5830	2,6750	2,9210
17,0000	0,6890	1,3330	1,7400	2,1100	2,5670	2,6570	2,8980
18,0000	0,6880	1,3300	1,7340	2,1010	2,5520	2,6410	2,8780
19,0000	0,6880	1,3280	1,7290	2,0930	2,5390	2,6270	2,8610
20,0000	0,6870	1,3250	1,7250	2,0860	2,5280	2,6140	2,8450
21,0000	0,6860	1,3230	1,7210	2,0800	2,5180	2,6030	2,8310
22,0000	0,6860	1,3210	1,7170	2,0740	2,5080	2,5930	2,8190
23,0000	0,6850	1,3190	1,7140	2,0690	2,5000	2,5840	2,8070
24,0000	0,6850	1,3180	1,7110	2,0640	2,4920	2,5750	2,7970
25,0000	0,6840	1,3160	1,7080	2,0600	2,4850	2,5680	2,7870
26,0000	0,6840	1,3150	1,7060	2,0560	2,4790	2,5610	2,7790
27,0000	0,6840	1,3140	1,7030	2,0520	2,4730	2,5540	2,7710
28,0000	0,6830	1,3130	1,7010	2,0480	2,4670	2,5480	2,7630
29,0000	0,6830	1,3110	1,6990	2,0450	2,4620	2,5430	2,7560
30,0000	0,6830	1,3100	1,6970	2,0420	2,4570	2,5370	2,7500
35,0000	0,6820	1,3060	1,6900	2,0300	2,4380	2,5160	2,7240
40,0000	0,6810	1,3030	1,6840	2,0210	2,4230	2,5010	2,7040
50,0000	0,6790	1,2990	1,6760	2,0090	2,4030	2,4790	2,6780
100,0000	0,6770	1,2900	1,6600	1,9840	2,3640	2,4360	2,6260
130,0000	0,6760	1,2880	1,6570	1,9780	2,3550	2,4270	2,6140



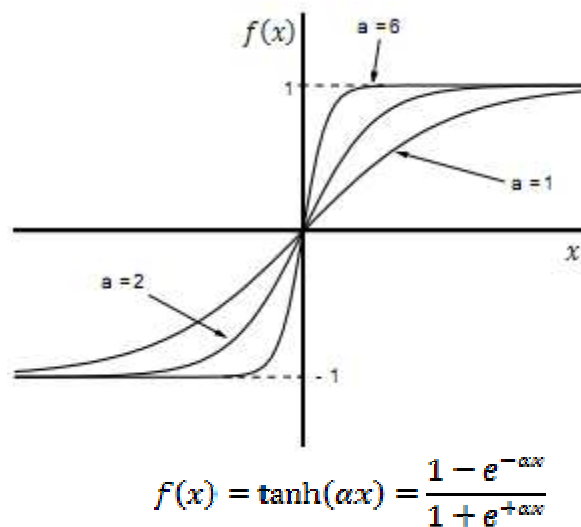
## Anexo 2 – Funções de ativação utilizadas

**Função Identidade:** Sua função matemática pode ser representada pela figura abaixo.



*Representação gráfica da função identidade.*

**Função Tangente Hiperbólica:** É contínua e diferenciável em todo o seu domínio, e permite valores de saída negativos. Sua função matemática pode ser expressa pela figura abaixo:



*Representação gráfica da função tangente hiperbólica para três valores diferentes da constante  $a$ .*

## Apêndice 1 – Tabela completa de análise da série CU\_1

**Tabela 01:** Valores referentes aos indicadores MSE, MSSE, SSE,  $\overline{MA}$  e PC obtidos mediante a utilização de todos os dados de teste da série CU\_1, em cada uma das 100 redes treinadas.

Legenda:

	Rede que possui menor valor de $\overline{MA}$
	Redes que possuem maiores valores de PC
	Redes que possuem valor aceitável de $\overline{MA}$ , segundo CS = 5%
	Rede escolhida como melhor, com maior PC dentre as redes com valor $\overline{MA}$ aceitável.

REDE	MSE	MSSE	SSE	$\overline{MA}$	PC
1	2,662	2,827	1413,703	3,321	93,400
2	2,112	2,294	1146,984	2,966	94,400
3	2,747	2,889	1444,442	3,991	95,600
4	1,286	1,384	691,816	2,576	95,000
5	2,906	2,883	1441,608	3,488	95,600
6	4,213	4,347	2173,599	4,169	95,800
7	4,259	5,008	2504,248	4,217	92,800
8	2,859	2,871	1435,719	3,497	95,000
9	1,096	1,269	634,255	2,147	93,600
10	4,647	5,114	2556,997	4,498	94,000
11	1,118	1,216	608,051	2,156	94,400
12	1,081	1,213	606,392	2,131	95,200
13	3,886	3,976	1988,101	4,097	95,400
14	3,026	3,553	1776,585	3,536	94,600
15	1,136	1,240	619,775	2,200	94,400
16	3,076	3,582	1790,768	3,572	94,800
17	3,826	4,393	2196,501	4,050	94,000
18	1,084	1,234	617,176	2,131	94,000
19	1,085	1,159	579,671	2,120	94,800
20	2,900	2,777	1388,594	3,731	95,400
21	1,100	1,259	629,435	2,129	94,200
22	1,086	1,214	607,097	2,118	94,000
23	3,443	4,437	2218,552	3,801	92,000
24	2,801	2,761	1380,260	3,539	96,200
25	4,771	4,801	2400,737	4,409	95,000
26	2,937	3,675	1837,704	3,480	93,800

27	1,104	1,221	610,418	2,144	94,200
28	6,358	6,804	3402,169	5,466	94,200
29	1,120	1,232	616,213	2,181	94,800
30	2,936	2,872	1436,054	3,469	96,200
31	3,797	3,859	1929,561	4,009	95,000
32	1,098	1,286	642,970	2,132	93,800
33	5,664	6,448	3224,128	15,902	92,800
34	1,140	1,217	608,421	2,181	94,400
35	1,309	1,393	696,706	2,336	95,200
36	3,245	3,938	1968,835	3,676	93,200
37	1,102	1,209	604,447	2,129	95,200
38	1,130	1,207	603,438	2,170	95,000
39	3,080	3,522	1760,939	3,740	95,600
40	3,174	3,552	1775,847	3,742	94,800
41	1,130	1,232	616,043	2,176	94,000
42	2,743	3,212	1606,103	3,441	94,400
43	5,070	5,019	2509,571	4,656	96,000
44	2,754	2,839	1419,686	3,372	94,800
45	2,836	2,880	1440,127	3,524	94,600
46	1,117	1,255	627,487	2,611	94,000
47	5,972	6,728	3363,859	9,021	94,600
48	3,918	4,114	2056,972	4,031	94,600
49	3,717	4,247	2123,435	4,144	93,000
50	3,125	3,698	1849,211	3,716	94,400
51	3,639	3,783	1891,426	4,058	95,600
52	2,730	2,666	1333,179	3,380	96,000
53	6,147	6,755	3377,321	5,162	93,600
54	7,051	7,155	3577,421	9,583	95,600
55	3,003	3,637	1818,274	3,749	93,600
56	4,208	4,757	2378,416	4,307	94,600
57	3,717	4,079	2039,292	3,927	94,000
58	1,142	1,253	626,722	2,192	94,200
59	3,094	3,929	1964,430	3,609	93,800
60	5,227	7,181	3590,412	4,663	91,200
61	1,113	1,243	621,579	2,181	94,400
62	1,099	1,161	580,499	2,136	95,000
63	1,143	1,258	628,812	2,184	94,600
64	3,834	5,107	2553,253	4,132	92,200
65	1,113	1,210	604,763	2,155	94,200
66	3,678	4,158	2079,210	4,006	94,400
67	2,845	2,791	1395,484	3,900	95,400
68	1,286	1,329	664,716	2,300	95,600
69	2,798	2,739	1369,296	3,423	95,000
70	2,965	3,542	1771,046	3,593	93,600
71	1,125	1,232	615,751	2,161	94,600

72	1,119	1,271	635,652	2,157	94,400
73	1,166	1,296	647,793	2,203	94,200
74	5,769	6,881	3440,726	5,271	93,600
75	1,307	1,373	686,589	2,370	94,800
76	2,843	2,854	1427,062	3,650	95,800
77	3,130	3,640	1819,993	3,890	94,600
78	4,392	5,195	2597,667	4,701	94,200
79	2,883	2,824	1411,763	3,511	95,600
80	1,034	1,125	562,643	2,073	94,400
81	2,987	3,641	1820,475	3,548	93,400
82	1,113	1,217	608,517	2,181	93,800
83	2,163	2,746	1372,893	3,026	92,400
84	5,532	6,148	3073,821	5,655	93,800
85	1,307	1,387	693,370	2,344	94,800
86	1,135	1,241	620,504	2,163	94,600
87	1,145	1,249	624,376	2,177	94,400
88	2,876	2,847	1423,446	3,480	95,200
89	1,134	1,238	618,948	2,311	94,800
90	4,482	6,003	3001,551	4,360	92,200
91	2,900	2,802	1400,894	3,746	95,000
92	5,850	6,735	3367,321	4,966	94,200
93	2,162	2,605	1302,604	3,102	93,800
94	3,293	3,683	1841,497	3,992	95,200
95	2,355	2,379	1189,352	3,117	95,400
96	2,517	2,526	1262,922	3,318	94,400
97	2,648	2,638	1319,065	3,350	95,000
98	1,124	1,197	598,381	2,195	94,800
99	1,306	1,363	681,482	2,352	94,600
100	4,259	5,057	2528,638	4,434	94,200