**XVII COBREAP – CONGRESSO BRASILEIRO DE ENGENHARIA DE AVALIAÇÕES E PERÍCIAS – IBAPE/SC – 2013**

# Estimativa do intervalo de confiança em avaliações de bens por redes neurais artificiais

**Agnaldo Calvi Benvenho**
Eng. Mecânico, Especialista em Engenharia de Avaliações e Perícias

## RESUMO

As redes neurais artificiais são uma metodologia emergente na avaliação de bens, em geral, e de imóveis, em particular. No entanto, o emprego desta técnica fornece estimativas pontuais de valor. É importante na maioria das aplicações, especialmente na avaliação, uma estimativa da confiabilidade do valor calculado, o intervalo de confiança. Neste artigo, será apresentada uma metodologia para o cálculo do intervalo de confiança de um valor pontual estimado, considerando a rede neural inferida como uma regressão não linear múltipla.

**Palavras-chave:** intervalo de confiança, redes neurais artificiais.

## 1. INTRODUÇÃO

Com o advento da NBR 14.653, as avaliações de imóveis passaram a ser classificadas quanto ao seu grau de fundamentação e de precisão (ABNT, 2004).

Para o enquadramento de qualquer avaliação quanto à precisão, é necessário calcular o intervalo de confiança associado ao estimador central de tendência resultante da modelagem estatística utilizada.

Em avaliações onde é utilizada a estatística descritiva ou a regressão linear via método dos mínimos quadrados ordinários, existem expressões analíticas e o cálculo do intervalo de confiança é direto.

Redes neurais artificiais usualmente fornecem apenas estimativas pontuais de valor, não sendo possível obter uma medida da confiabilidade da avaliação obtida. Desta forma, não é possível o enquadramento de uma avaliação de imóveis quanto ao grau de precisão.

Vários estudos tem mostrado que as redes neurais artificiais são boas ferramentas para a predição de valores imobiliários, sendo que em alguns deles, são superiores à regressão linear múltipla. Nguyen e Cripps (2001) concluem a superioridade preditiva das Redes Neurais Artificiais em relação à Regressão Linear Múltipla. Guedes (1995) efetuou o estudo de uma amostragem de 102 escritórios, comparando aquelas com a técnica de Regressão Linear Múltipla, concluindo que as Redes Neurais apresentam performance preditiva bastante superior. Melo (2002) compara as Redes Neurais Artificiais com dois diferentes modelos obtidos por Regressão Linear. Utilizando uma amostra de 128 elementos, ele divide a comparação em fases de treinamento e predição. Na fase de treinamento as Redes Neurais Artificiais tem um desempenho levemente superior. Na fase de predição, os modelos via Regressão Linear revertem o quadro, apresentando desempenho superior, também de forma tímida.

Apesar de não se tratar de uma metodologia consagrada para avaliações de imóveis, as redes neurais artificiais poderiam ser uma boa alternativa, especialmente em situações onde as regressões lineares não apresentassem bons resultados. Entretanto, a impossibilidade do cálculo de intervalo de confiança e, consequentemente, do enquadramento da avaliação quanto à precisão tornariam a mesma incompleta diante da norma.

A proposta deste artigo é desenvolver uma metodologia para estimar intervalos de confiança para predições oriundas em redes neurais artificiais, treinadas através do método de retropropagação, possibilitando o enquadramento da predição quanto ao grau de precisão, validando a mesma para uso em conformidade com a NBR 14.653-2/02.

Também é feita uma discussão sobre o uso dos resultados calculados para a determinação do intervalo de confiança na verificação da adequação da arquitetura e dos parâmetros da rede treinada.

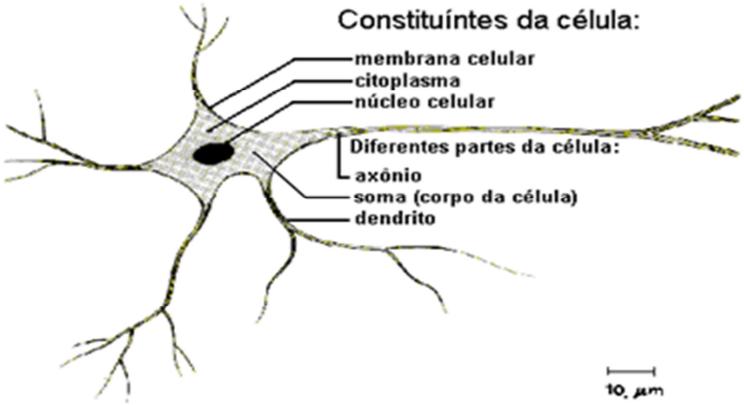
## 2. REDES NEURAIS ARTIFICIAIS

Segundo Gonzáles e Formoso (2000), redes neurais artificiais são modelos computacionais que buscam emular o cérebro humano em sua forma de processar informações. Elas são estruturalmente compostas por um conjunto de múltiplos

processadores (unidades), cada qual tendo uma pequena quantidade de memória, ligados entre si por canais de comunicação (conectores), que tem a capacidade de transmitir dados numéricos, codificados em vários formatos (sinais), à semelhança do neurônio humano. As unidades operam apenas seus dados locais e suas entradas ocorrem pelas conexões.

No modelo humano, os neurônios se comunicam através de sinapses, que são regiões onde dois neurônios fazem contato. Este contato não é físico, mas sim químico, através de neurotransmissores. O sinal de entrada é captado pelo dendrito, que o envia ao núcleo. O mesmo processa o sinal, e caso ele atinja um certo limiar, o neurônio se excita, gerando uma nova sinapse, na qual o sinal processado é passado pelo axônio para o neurônio subsequente. Caso o sinal processado não atinja o limiar, o neurônio fica inibido, não gerando uma nova sinapse. Esse processo depende do sinal de entrada, da geometria da sinapse e dos neurotransmissores envolvidos. A Fig. 1 exibe uma representação esquemática de um neurônio humano.

*Fig. 1: Neurônio humano*

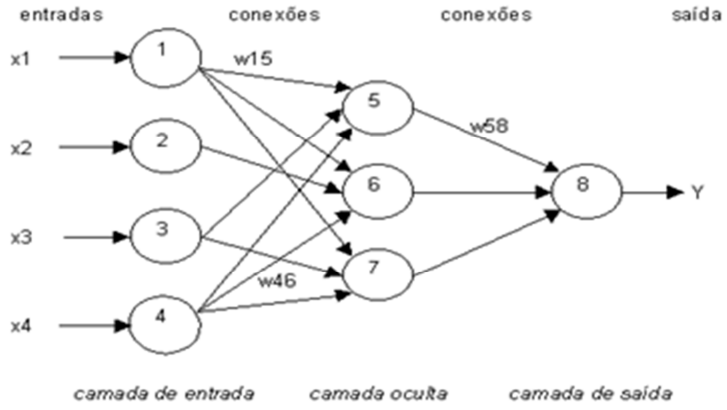



O modelo básico de uma rede neural artificial trata-se de uma série de neurônios de entrada (camada de entrada), os quais são responsáveis pelo primeiro processamento, utilizando funções matemáticas e pesos e por um ou mais neurônios de saída (camada de saída), onde são passadas as variáveis resposta.

O modelo de rede neural artificial mais difundido é o *perceptron* de múltiplas camadas, também conhecido como *multilayer perceptron*. Este modelo consiste na utilização de camadas de neurônios artificiais, sendo uma de entrada, uma de saída e uma série de intermediárias, também conhecidas como *hidden layers*.

A Fig. 2 mostra uma configuração básica de rede neural *multilayer perceptron* com uma camada intermediária

*Fig. 2: Rede neural multilayer perceptron com 1 camada oculta*




Dependendo do tipo de problema, pode-se usar diversas camadas intermediárias, que são função da complexidade do problema representado. Entretanto, de acordo com Kovács (2006), um teorema proposto originalmente por Andrei Kolmogorov em 1957 para funções contínuas foi revisto posteriormente por Robert Hecht-Nielsen (Hecht-Nielsen, 1989), tendo aplicação direta para redes neurais artificiais, no qual qualquer função contínua e diferenciável, ainda que não linear, pode ser aproximada por um rede neural do tipo *multilayer perceptron* com uma camada escondida[1]. O resultado foi, posteriormente, demonstrado de forma rigorosa por Hornik, Stinchcombe e White (1989).

Para as ligações entre dois neurônios i e j, teremos um peso específico associado a conexão, $w_{ij}$. Para cada neurônio, existe uma função de ativação. No caso das *multilayer perceptron*, as mais usadas são a linear, y = x e a sigmóide logística, representada pela Eq. 1.

$$y = \frac{1}{1 + e^{-x}} \tag{1}$$

Segundo Guedes (2001), é necessário treinar a rede neural, ou seja, apresentam-se pares de dados de entrada e saída, permitindo a ela fazer associações descobrindo assim a existência de algum padrão de comportamento.

O treinamento de uma rede *multilayer perceptron* é feito pelo algoritmo de retropropagação, também conhecido com regra delta generalizada. Ela consiste em ajustar os pesos $w_{ij}$, conforme as entradas $x_p$ ($x_1$, $x_2$,...,$x_p$) e saídas $y_k$ ($y_1$, $y_2$,...,$y_k$), através da minimização de uma função de erro quadrado, conforme a Eq. 2 a seguir.

$$\text{Erro} = \frac{1}{2}\sum_j \left(\hat{y}_j - y_j\right)^2 \tag{2}$$

---

[1] **Teorema de Kolmogorov-Nielsen**: Dada uma função contínua arbitrária $f:[0,1]^n \rightarrow R^m$, f(x) = y, *existem sempre* para f uma implementação exata com uma rede neural de três camadas, sendo a camada de entrada um vetor de dimensão n, a camada oculta composta por (2n + 1) neurônios, e a camada de saída com m neurônios representando as m componentes do vetor y.

Uma descrição detalhada pode ser encontrada em Guedes (2001), Azevedo, Brasil e Oliveira (2000) e Braga, Carvalho e Ludemir (2003).

## 3. INTERVALO DE CONFIANÇA

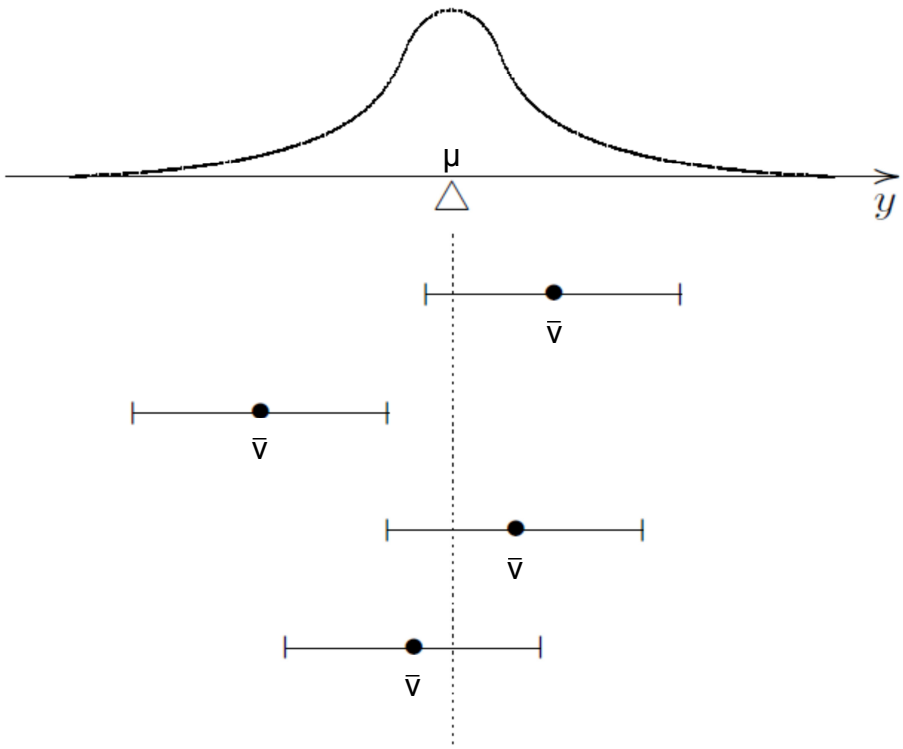
### 3.1 Intervalo de Confiança Clássico

O intervalo de confiança é uma medida estatística de confiabilidade de um determinado parâmetro populacional, inferido através de uma amostra.

Suponha que μ seja a média de uma determinada população de valores v. A média $\bar{v}$ de uma amostra aleatória S da população de interesse é um estimador pontual para μ, mas considerando que seja improvável que $\mu = \bar{v}$, quão confiável é a estimativa do padrão populacional pela média amostral?

A resposta está em construir um limite inferior $\lambda_i(S)$ e um limite superior $\lambda_s(S)$ da amostra S tal que exista uma possibilidade de 95% do intervalo $[\lambda_i(S), \lambda_s(S)]$ contenha μ[2]. Em outras palavras, se forem extraídas infinitas amostras $S_1$, $S_2$,..., aleatórias e de mesmo tamanho da população de interesse, 95% dos intervalos $[\lambda_i(S_1), \lambda_s(S_1)]$,$[\lambda_i(S_2), \lambda_s(S_2)]$,..., associados a essas amostras conterá o valor fixo μ. Tal é conhecido como *intervalo de confiança clássico de 95% em torno de μ* (Dybowski e Roberts, 2001). Esta caracterização pode ser vista na Fig. 3 a seguir.

*Fig. 3: Intervalo de confiança clássico*



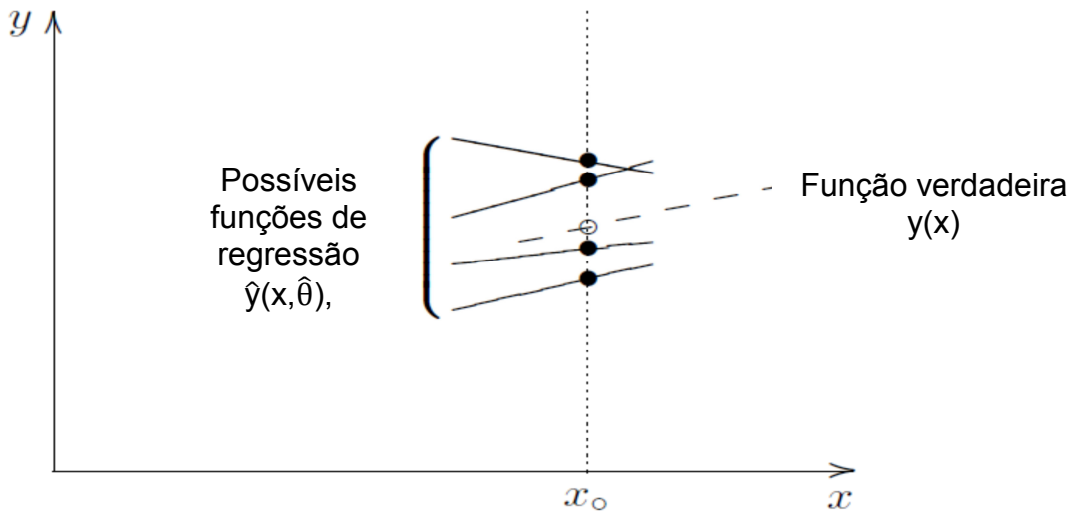

[2] Neste caso, está sendo considerado o intervalo de confiança de 95%.

### 3.2 Intervalo de Confiança para Regressão Linear Simples

Para uma equação de regressão qualquer $\hat{y}(x,\hat{\theta})$, que busca aproximar a verdadeira função y(x), pode haver significativa variação nos estimadores $\hat{\theta}$ em função da amostra aleatória S extraída para sua inferência.

Neste caso, busca-se estabelecer a confiabilidade de que a equação inferida $\hat{y}(x,\hat{\theta})$ aproxime a verdadeira função y(x) nas proximidades de um ponto $x_0$ qualquer, conforme pode ser visto na Fig. 4.

*Fig. 4: Intervalo de confiança para regressão linear*




De maneira análoga ao intervalo de confiança clássico, pode-se definir um intervalo $[\lambda_i(S,x_0),\lambda_s(S, x_0)]$ no qual haja a probabilidade de 95% do verdadeiro valor $y(x_0)$ ocorrer.

### 3.3 Intervalo de Confiança para Regressão Linear Múltipla

Na seção 2.2 apresentamos a derivação do intervalo de confiança considerando uma regressão linear simples, do tipo y = f(x). Esta opção se deu pois a visualização gráfica de uma função de uma variável é bastante intuitiva. No entanto, os conceitos são diretamente extensíveis para funções a várias variáveis.

Considerando uma regressão linear múltipla a n variáveis nos termos da Eq. 3

$$y_i = w_1x_{i1} + w_2xi2 + w_3x_{i3} + ... + w_kx_{ik} + e_i, \quad i = 1, 2,...,n \qquad (3)$$

sendo $y_i$ a variável dependente, $w_i$ os regressores a serem determinados e $x_{ik}$ as variáveis independentes ou explicativas. As técnicas de regressão linear múltipla são descritas em Dantas (1999).

Considerando-se um valor estimado $\hat{y}_h = w_1x_{1h} + w_2x_{2h} +... + w_ix_{ih}$, sendo $\hat{y}_h$ o valor estimado para $y_h$, o vetor $(x_{1h},...,x_{ih})$ dos valores para cada uma das variáveis independentes e considerando $s(\hat{y}_h)$ o desvio padrão relativo a estimativa de $\hat{y}_h$, $S_1$ o desvio padrão dos resíduos do modelo, n o número de amostras utilizado na

determinação do modelo, $s(w_i)$ o desvio padrão do regressor i, $\overline{x}_{hi}$ as médias para cada variável independente e $cov(w_i,w_l)$, sendo (l, i = 1, 2, ..., k, l < i), teremos as Eq. 4 e 5:

$$S(\hat{y}_h) = \sqrt{\frac{S_l^2}{n} + \sum_{j=i}^{k}\left(x_{hi} - \overline{x}_{hj}\right)^2 s(w_j) + 2\sum_{l<j}^{k}\left(x_{hl} - \overline{x}_{hl}\right)\left(x_{hj} - \overline{x}_{hj}\right)cov(w_j, w_l)} \quad (4)$$

Em notação matricial, temos:

$$S(\hat{y}_h) = S_e\sqrt{\frac{1}{n} + (X - \overline{X})^T (X^T X)^{-1}(X_h - \overline{X})} \quad (5)$$

### 3.4 Intervalos de Confiança e Intervalos de Predição

Uma outra medida estatística de confiabilidade, também apresentada em forma de intervalo, e que se costuma fazer confusão com o intervalo de confiança, é o denominado intervalo de predição.

Na definição de intervalo de predição clássico, analogamente ao apresentado na seção 2.1, segundo Dybowski e Roberts (2001), se considerarmos uma amostra aleatória S com n valores $v_1, v_2, ..., v_n$, com média $\overline{v}$, é possível construir um intervalo $[\psi_i(S), \psi_s(S)]$ de tal forma que um novo elemento $v_{n+1}$ coletado aleatoriamente na mesma população tenha 95% de chance de estar contido no mencionado intervalo.

Para uma regressão linear[3], supondo que exista um conjunto de observações Sn = (xi, yi), para 1 ≤ i ≤ n, segundo Zapranis e Livanis (2005), existe uma expressão (Eq. 6) que modela a relação Sn:

$$y_i = f(x_i, \theta) + e_i, i = 1,2,...,n \quad (6)$$

sendo que $\theta$ é o vetor de parâmetros "verdadeiros" da função desconhecida. Considerando ainda que $f(x_i)$ é a função a ser estimada, pode-se considerar como verdadeira a relação da Eq. 7.

$$f(x_i,\theta) \approx g(x_i) \equiv E[y_i|x_i] \quad (7)$$

Considerando a minimização de uma função de erro quadrático similar a (2), infere-se um modelo que aproxima a verdadeira função $g(x_i)$ em termos de $x_i$ e $\hat{\theta}$, que são os estimadores dos verdadeiros parâmetros $\theta$, nos termos da Eq. 8.

$$\hat{y}_i = f(x_i,\hat{\theta}) \quad (8)$$

Neste âmbito, conforme já comentado na seção 2.2, o intervalo de confiança é uma representação da acurácia da estimativa da verdadeira, mas desconhecida,

[3] Os resultados podem ser adotados diretamente para uma regressão não linear.

função de regressão $g(x_i)$, portanto, o valor de interesse é a distribuição da quantidade (Eq. 9).

$$g(x_i) - f(x_i,\hat{\theta}) \equiv g(x_i) - \hat{y}_i \tag{9}$$

Por outro lado, o intervalo de predição representa a acurácia do valor estimado pela regressão, ou seja a quantidade expressa na Eq. 10.

$$y_i - f(x_i,\hat{\theta}) \equiv y_i - \hat{y}_i \tag{10}$$

De (9) e (10) resulta a relação da Eq. 11:

$$(y_i - \hat{y}_i) = (g(x_i) - \hat{y}_i) + e_i \tag{11}$$

Decorrendo daí que o intervalo de confiança está contido no intervalo de predição.

### 3.5 Intervalo de Confiança para Regressão Não Linear

Uma função de regressão do tipo $y = f(x, \theta)$ quando a função que apresenta os melhores estimadores para y, tal que: $y = f(x_i,\hat{\theta}) + e_i$, sendo i = 1, 2,...,n, $e_i$ é um termo estocástico de erro e $\hat{\theta}$ é não linear, ou seja, a função y não pode ser representada pela Eq. 12.

$$y = f(x_i, \hat{\theta}) = \sum_{j=1}^{p} x_{ij}\hat{\theta}_j \text{, com } j = 1, 2, ...p. \tag{12}$$

De acordo com Donaldson e Schnabel (1985), existem três métodos para se determinar o intervalo de confiança para uma regressão não linear:

- Linearização dos regressores;
- Máxima verossimilhança;
- Bootstrap.

Ainda de acordo com os autores, o método da linearização dos regressores é o mais freqüentemente utilizado, pois requer menor poder computacional e produz resultados de mais fácil interpretação. Todavia, produz subestimativas dos intervalos de confiança. Os outros métodos, apesar de apresentar estimativas melhores, requerem grande poder de computação e produzem resultados que muitas vezes são difíceis de interpretar.

Neste âmbito, estudaremos o método da linearização para a determinação de intervalos de confiança m regressões lineares. Ainda, na linha dos autores, quando a determinação dos parâmetros $\hat{\theta}$ $\left(\hat{\theta} = \left(X^T X\right)^{-1} X^T Y\right)$, sendo X a matriz das variáveis independentes, Y a matriz das variáveis dependentes e as notações T indicando

matriz transposta e -1 matriz inversa, é feita pelo método dos mínimos quadrados não lineares

De acordo com Chryssolouris, Lee e Ramsey (1996), considerando um modelo para um sistema qualquer $f(x,\theta^*)$, sendo x as entradas do modelo (variáveis independentes) e $\theta^*$ representa os verdadeiros valores dos estimadores $\theta$. Assume-se que o erro associado ao modelo é independente, com variância $\sigma^2$ e possui distribuição normal tal que $N(0, \sigma^2)$. Para n observações, tais que i = 1, 2,...,n, o sistema é representado pela Eq. 13:

$$y_i = f(x_i, \theta^*) + e_i, i = 1,2,...,n \tag{13}$$

A estimativa por mínimos quadrados de $\theta^*$ é $\hat{\theta}$, obtida através da minimização da função de erro, expressa na Eq. 14.

$$S(\theta) = \sum_{i=1}^{n} [y_i - f(x_i, \theta)]^2 \tag{14}$$

onde $\hat{y}_i = f(x_i, \hat{\theta})$

Sendo o modelo um bom estimador do sistema, então teremos $\hat{\theta} \approx \theta^*$, logo, segundo Rivals e Personnaz (1998), é possível linearizar o modelo aplicando a série de Taylor nas imediações de $\theta^*$. De acordo com Leite e Singer (1990), a expansão de funções em séries de Taylor é um das ferramentas mais utilizadas no estudo de métodos assintóticos. Aplicando a expansão, temos as Eqs. 15 e 16.

$$f(x_i, \hat{\theta}) \approx f(x_i, \theta^*) + f_o^T(\hat{\theta} - \theta^*) \tag{15}$$

$$f_o^T = \left( \frac{\partial f(x_i, \theta^*)}{\partial \theta_1^*}, \frac{\partial f(x_i, \theta^*)}{\partial \theta_2^*}, \ldots, \frac{\partial f(x_i, \theta^*)}{\partial \theta_n^*} \right) \tag{16}$$

Derivado das expressões acima, o erro pontual entre o valor do y observado e do $\hat{y}$ estimado é dado pelas Eqs. 17.

$$y_o - \hat{y}_o \approx y_o - f(x_o, \theta^*) - f_o^T(\hat{\theta} - \theta^*) = e_o - f_o^T(\hat{\theta} - \theta^*) \tag{17}$$

A quantidade expressa por (17) é similar a distribuição de (10), que se trata do intervalo de predição. A Eq. 18 apresenta a esperança matemática da distribuição em análise.

$$E[y_o - \hat{y}_o] \approx E[e_o] - f_o^T E[(\hat{\theta} - \theta^*)] \approx 0 \tag{18}$$

Considerando a independência estatística entre $\hat{\theta}$ e o termo de erro, a Eq. 19 exprime variância esperada entre y e $\hat{y}$.

$$\mathrm{var}\left[y_o - \hat{y}_o\right] = \mathrm{var}\left[e_o\right] - \mathrm{var}\left[f_o^T\,(\hat{\theta} - \theta^*)\right] \tag{19}$$

Antes de prosseguir, a linearização em séries de Taylor depende do emprego da Matriz Jacobiana (F), a qual, segundo Lima (1969), trata-se da matriz das derivadas parciais de primeira ordem de uma função vetorial $f(x_1, x_2, \ldots, x_n)$ no ponto $x_0$. Define-se também (embora não utilizada neste artigo) a Matriz Hessiana (H) como a matriz das derivadas parciais de segunda ordem da mesma função vetorial $f(x_1, x_2, \ldots, x_n)$ no ponto $x_0$.

A matriz Jacobiana para os parâmetros da regressão, denominada $F\cdot(\hat{\theta})$, de acordo com Ungar, De Veaux e Rosengarten (1994), é representada pela Eq. 20

$$F\cdot(\hat{\theta}) = \left[\frac{\partial f_i(x_i,\hat{\theta})}{\partial \hat{\theta}_j}\right] \tag{20}$$

ou sua expansão, de acordo com Chryssolouris, Lee e Ramsey (1996), acarreta na Eq. 21

$$F\cdot(\hat{\theta}) = \begin{bmatrix} \frac{\partial f_1(x_1,\hat{\theta})}{\partial \hat{\theta}_1} & \frac{\partial f_1(x_1,\hat{\theta})}{\partial \hat{\theta}_2} & \Lambda & \frac{\partial f_1(x_1,\hat{\theta})}{\partial \hat{\theta}_p} \\ \frac{\partial f_2(x_2,\hat{\theta})}{\partial \hat{\theta}_1} & \frac{\partial f_2(x_2,\hat{\theta})}{\partial \hat{\theta}_2} & \Lambda & \frac{\partial f_2(x_2,\hat{\theta})}{\partial \hat{\theta}_p} \\ & & \mathrm{M} & \\ \frac{\partial f_n(x_n,\hat{\theta})}{\partial \hat{\theta}_1} & \frac{\partial f_n(x_n,\hat{\theta})}{\partial \hat{\theta}_2} & \Lambda & \frac{\partial f_n(x_n,\hat{\theta})}{\partial \hat{\theta}_p} \end{bmatrix} \tag{21}$$

Segundo Chryssolouris, Lee e Ramsey (1996), o erro $e_o$ pode ser aproximado por uma distribuição normal tal que $e_o \sim N(0,\sigma^2 I)$ onde I é a matriz identidade e o termo $\hat{\theta}$ pode ser aproximado por uma distribuição normal tal que $(\hat{\theta} - \theta^*) \sim N(0,\sigma^2[F\cdot(\hat{\theta})^T F(\hat{\theta})]^{-1})$, então $\mathrm{var}(y_o - \hat{y}_o) \approx \sigma^2 + \sigma^2\ f_o^T\,(F^T\cdot F\cdot)^{-1} f_o$. Também, segundo os mesmos, temos um estimador $s^2$ para a variância $\sigma^2$ que segue uma distribuição t de Student é calculado pela Eq. 22

$$s^2 = \frac{\left\|y - f(x,\hat{\theta})\right\|^2}{n - p} \tag{22}$$

sendo n é o número de amostras e p o número de parâmetros $\hat{\theta}$.

Assim, teremos um intervalo de predição IP, considerando uma distribuição t de Student com n – p graus de liberdade e grau de confiança $\alpha$ é calculado pela Eq. 23.

$$IP = t_{n-p}^{\alpha/2}\, s\,(1 + f_o^T (F^T F)^{-1} f_o)^{\frac{1}{2}} \tag{23}$$

Conjugando (23) com (11), decorre que o intervalo de confiança IC para os mesmos parâmetros apresentados é dado pela Eq. 24.

$$IC = t_{n-p}^{\alpha/2}\, s\,(f_o^T (F^T F)^{-1} f_o)^{\frac{1}{2}} \tag{24}$$

Segundo Chryssolouris, Lee e Ramsey (1996), existem três variantes para a matriz variância-covariância na linearização de modelos não lineares, um utilizando o operador Jacobiano (Eq. 25), um utilizando o operador Hessiano (Eq. 26) e outro usando uma combinação dos dois (Eq.27).

$$V_\alpha = s^2 \left(F^T F\right)^{-1} \tag{25}$$

$$V_\beta = s^2 H^{-1} \tag{26}$$

$$V_\gamma = s^2 H^{-1} \left(F^T F\right) H^{-1} \tag{27}$$

Donaldson e Schnabel (1985) concluem que o método que utiliza o operador Jacobiano é preferível, pois é mais simples, demanda menor poder computacional, é numericamente mais estável e tão preciso quanto os outros métodos. De fato, o emprego de expressões com a Matriz Hessiana demanda o cálculo de derivadas de segunda ordem, nem sempre existentes e muitas vezes instáveis.

## 4. INTERVALO DE CONFIANÇA PARA REDES NEURAIS ARTIFICIAIS

Segundo Oliveira, Montini e Bergman (2007), métodos que fornecem intervalos de predição assintóticos para regressão não linear podem, em teoria, ser aplicados diretamente a redes neurais.

Uma vez que redes neurais artificiais podem ser modeladas como regressões não lineares, a aplicação dos resultados derivados no capítulo 2 é direta. Cabe, entretanto, verificar que os erros do modelo sigam uma distribuição normal (ou ao menos próxima) e que tenham variância constante (homocedasticidade).

Por outro lado, algumas observações na adoção destes resultados, para questões de aplicações práticas, especialmente após a modelagem, são interessantes e podem melhorar a aplicabilidade dos resultados.

### 4.1 Cálculo dos Elementos da Matriz Jacobiana

A equação 21 mostra a forma de calcular os elementos da matriz jacobiana de variância-covariância do modelo de regressão não linear.

Para a utilização em redes neurais artificiais, faz-se necessário o cálculo de derivadas pontuais de funções de ativação dos neurônios da rede, as quais podem ser complexas como a sigmoide logística e a tangente hiperbólica.

Em seu trabalho, Chryssolouris, Lee e Ramsey propõe uma metodologia que busca efetuar os cálculos dos elementos em função de pesos sinápticos e saídas dos neurônios. Para isso são definidos os termos $net_j^{[\beta]}$ (Eq. 28) e $layer^{[\beta]}$ (Eq. 29).

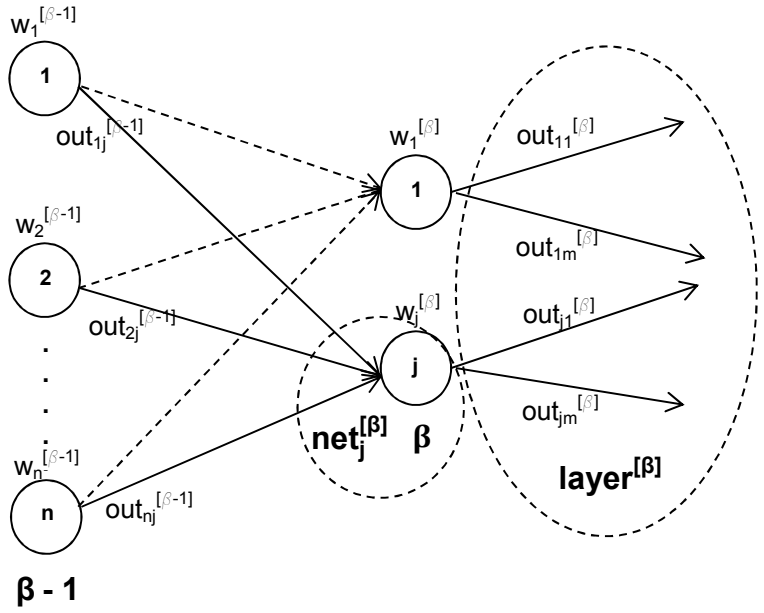
$$net_j^{[\beta]} = \left(\sum_{i=1}^{n} w_i^{[\beta-1]} out_i^{[\beta-1]}\right) - b_i \tag{28}$$

$$layer^{[\beta]} = \sum_{i=1}^{m} net_j^{[\beta]} \tag{29}$$

O item $net_j^{[\beta]}$ é o somatório das saídas dos nós da camada β – 1 entrando no nó j na camada β. Na eq. 27, $w_i^{[\beta-1]}$ é o peso do i-ésimo neurônio da camada β – 1, $out_i^{[\beta-1]}$ é a saída do i-ésimo neurônio da camada β – 1, após a função de ativação e $b_i$ é o termo de bias. O item $layer^{[\beta]}$ é a resposta da camada β para um dado conjunto de entradas, o qual é o somatório dos termos $net_j^{[\beta]}$, sendo m o número de neurônios na camada β.

O esquema a seguir (Fig. 5) facilita a visualização dos termos apresentados.

*Fig. 5: Esquema de representação de $net_j^{[\beta]}$ e $layer^{[\beta]}$*




Segundo os autores, é possível escrever os termos $\frac{\partial y}{\partial w_\gamma^{[\alpha]}}$ onde y é a resposta e $w_\gamma^{[\alpha]}$ é o peso do neurônio γ na camada α pela Eq. 30.

$$\frac{\partial y}{\partial w_\gamma^{[\alpha]}} = \frac{\partial y}{\partial net^{[m]}} \frac{\partial net^{[m]}}{\partial layer^{[m-1]}} \frac{\partial layer^{[m-1]}}{\partial layer^{[m-2]}} \cdots \frac{\partial layer^{[\alpha+3]}}{\partial layer^{[\alpha+2]}} \frac{\partial layer^{[\alpha+2]}}{\partial net^{[\alpha+1]}} \frac{\partial net^{[\alpha+1]}}{\partial w_\gamma^{[\alpha]}} \tag{30}$$

Sendo que os termos podem ser calculados pelas seguintes equações:

$$\frac{\partial y}{\partial net^{[m]}} = \begin{cases} 1 \text{ , se o neurônio de saída tiver função de ativação linear} \\ y(1-y), \text{ se o neurônio de saída tiver função de ativação sigmóide} \end{cases} \quad (31)$$

$$\frac{\partial net^{[m]}}{\partial layer^{[m-1]}} = \sum_{i=1}^{n} \frac{\partial net^{[m]}}{\partial net_i^{[m-1]}} \quad (32)$$

$$\frac{\partial layer^{[m-1]}}{\partial layer^{[m-2]}} = \sum_{j=1}^{p} \left( \sum_{i=1}^{n} \frac{\partial net_j^{[m-1]}}{\partial net_i^{[m-2]}} \right) \quad (33)$$

sendo p o número de neurônios na camada m-1.

$$\frac{\partial layer^{[\alpha+2]}}{\partial net_\gamma^{[\alpha+1]}} = \sum_{i=1}^{p} \frac{\partial net_j^{[\alpha+2]}}{\partial net_\gamma^{[\alpha+1]}} \quad (34)$$

$$\frac{\partial net_{\cdot}^{[\alpha+1]}}{\partial w_\gamma^{[\alpha]}} = out_\gamma^{[\alpha]} \quad (35)$$

A Eq. 35 apresenta uma forma geral de $\frac{\partial net_\lambda^{[\phi+1]}}{\partial net_\xi^{[\phi]}}$, onde $\phi$ é uma camada arbitrária, $\lambda$ é um neurônio arbitrário na camada $\phi + 1$, $\xi$ é um neurônio arbitrário na camada $\phi$ e A é o número de neurônios na camada $\phi$.

$$\frac{\partial net_\lambda^{[\phi+1]}}{\partial net_\xi^{[\phi]}} = \sum_{a=1}^{A} \left\{ w_a^{[\phi]} out_a^{[\phi]} \left(1 - out_a^{[\phi]}\right) \right\} \quad (36)$$

O desenvolvimento apresentado acima mostra a tentativa dos autores de simplificar o cálculo dos componentes da matriz jacobiana. Em princípio, as relações parecem complexas, mas a idéia é desenvolver expressões que não dependam do cálculo de derivadas parciais de forma analítica.

Não obstante, seria necessário o programa utilizado para o treinamento da rede fornecer os pesos sinápticos, bem como as saídas de cada neurônio para cada um dos neurônios da camada seguinte. Com estes dados, escaparia-se de calcular derivadas parciais de funções complexas. No entanto, desconhecemos qualquer programa de treinamento de redes neurais artificiais que “abra” a este nível o processamento interno.

Por outro lado, futuras aplicações de redes neurais, sobretudo dedicadas a avaliação imobiliária, poderiam incorporar sub-rotinas para estes cálculos, utilizando algoritmos numéricos que usualmente fornecem boas aproximações.

### 4.2 Influência do Tamanho da Amostra de Treinamento no Cálculo do Intervalo de Confiança

Um dos pontos comuns nos trabalhos sobre intervalos de confiança em redes neurais por meio de modelos de regressão não linear (Ungar, De Veaux e

Rosengarten, 1994; Chryssolouris, Lee e Ramsey, 1996; Rivals e Personnaz, 1998 e Dybowski e Roberts, 2001) é a hipótese de que a amostra utilizada no treinamento é muito grande, tendendo ao infinito, ou seja, os resultados são válidos assintoticamente.

No entanto, para a grande maioria das aplicações práticas, trabalha-se com um número limitado de elementos na amostra. Para o caso das avaliações imobiliárias, dependendo da situação, tal ocorrência é praxe.

Segundo De Veaux, Schumi, Schweinsberg e Ungar (1994), quando o conjunto amostral é pequeno e a rede é treinada até a convergência, a matriz $F^TF$ pode se tornar singular, impedindo o cálculo de sua inversa e, consequentemente, do intervalo de confiança.

De modo a prevenir esta ocorrência, Yang, Kavli, Carlin, Clausen e De Groot, (2000) recomendam a utilização do método de treinamento conhecido como *weight decay*[4], no qual a função de erro a ser minimizada, ao invés de (14) passa a ser a Eq. 37.

$$S(\theta) = \sum_{i=1}^{n} \left[y_i - f(x_i, \theta)\right]^2 + c\sum_{i=1}^{p} \theta_i \tag{37}$$

onde c > 0 é o parâmetro de decaimento de peso (*weight decay*).

Consequentemente, a nova expressão para o cálculo do intervalo de confiança do modelo é a Eq. 38.

$$IC = t_{n-p}^{\alpha/2}\, s\, (f_o^T (F^TF + cI)^{-1} (F^TF(F^TF + cI)^{-1} f_o)^{\frac{1}{2}} \tag{38}$$

onde I é a matriz identidade.

Como pode ser visto, a expressão torna-se ligeiramente mais complexa, envolvendo um maior número de cálculos matriciais.

Uma alternativa recomendada por alguns autores autores (De Veaux, Schumi, Schweinsberg e Ungar, 1994 e Yang, Kavli, Carlin, Clausen e De Groot, 2000) é a parada do treinamento da rede anteriormente a convergência. No entanto, Zapranis e Livanis (2005) alertam que este procedimento por levar a intervalos de confiança muito amplos, acarretando em impossibilidade de inferências em relação aos mesmos. Como alternativa, recomendam a eliminação de conexões irrelevantes, com o objetivo de reduzir os parâmetros da rede. Este procedimento é denominado *Irrelevant Connection Elimination* (Eliminação de Conexões Irrelavantes), e já é implementado por uma série de softwares para treinamento de redes neurais.

### 4.3 Análise da Matriz Jacobiana, Existência do Intervalo de Confiança e Adequação do Modelo

O uso do decaimento de pesos ou a parada antes da convergência de uma rede podem resolver o problema da singularidade da matriz $F^TF$ ou outro problema, conhecido como *overfitting*, ou excesso de ajuste, com mais parâmetros do que os necessários no modelo.

---

[4] Decaimento de pesos

Entretanto, segundo Rivals e Personnaz (1998), o uso destas técnicas usualmente acarreta na introdução de viés no modelo, gerando intervalos de confiança dos quais não se pode fazer inferências, ou seja, inúteis.

Rivals e Personnaz (2004) recomendam efetuar uma decomposição singular de valor na matriz Jacobiana F (ou da matriz produto $F^TF$) nos termos propostos por Golub e Reinsch (1970), analisando seu número de condição, que é a razão entre seu maior e menor valores. Como estamos lidando com aspectos computacionais, o número de condição deve ser inferior ao inverso da precisão do computador, ou seja $10^{16}$ ($10^8$, caso se lide com a matriz produto $F^TF$).

Para valores acima destes, a matriz se torna singular e o intervalo de confiança não pode ser calculado. Logo, a rede treinada é problemática, seja por incoerência, seja por excesso de ajuste, sendo necessário um novo treinamento.

Rivals e Personnaz (1998) também recomendam efetuar outros dois testes estatísticos, para verificar a consistência do intervalo de confiança.

Inicialmente, deve ser calculada a projeção ortogonal O da matriz $F^TF$, de acordo com a Eq. 39.

$$O = F(F^TF)^{-1}F^T \tag{39}$$

Neste momento, cabe definir dois termos: (i) posto de uma matriz e (ii) traço de uma matriz.

i. *Posto de uma matriz*: é o número de linhas de uma matriz linearmente independente das outras, ou seja, cujos elementos não podem ser escritos como combinação linear das demais;

ii. *Traço de uma matriz*: é a soma dos elementos principais da diagonal de uma matriz quadrada.

Logo, se F tem posto completo, ou seja, todas as linhas são linearmente independentes, tem-se as Eq. 40 e 41.

$$\text{traço}(O) = \sum_{k=1}^{N} o_{kk} = \text{posto}(O) = p \tag{40}$$

$$\frac{1}{N} \leq o_{kk} \leq 1 \tag{41}$$

Caso estas condições sejam satisfeitas, verifica-se que não há excesso de ajuste na rede, e o intervalo de confiança para os valores preditos dos elementos de treinamento da mesma, ao invés de (24), passa a ser calculado pela Eq. 42.

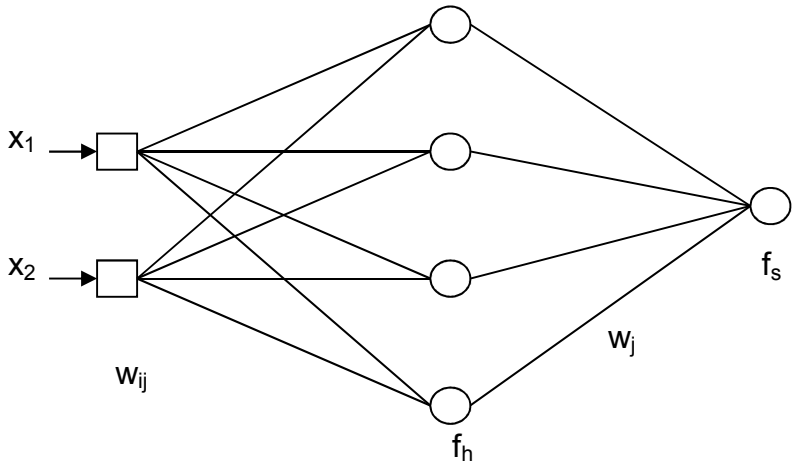
$$IC = t_{n-p}^{\alpha/2}\, s\,(o_{kk})^{\frac{1}{2}} \tag{42}$$

Para predições da rede, deve-se continuar usando (24) para determinar o intervalo de confiança. Mas (42) permite averiguar a adequação do intervalo de confiança perante os dados de treinamento.

## 5. APLICAÇÃO PRÁTICA

Apresentamos uma rede neural artificial do tipo mutilayer perceptron constituída por três camadas (entrada, intermediária e saída), com dois neurônios de entrada ($x_1$, $x_2$), quatro neurônios na camada intermediária ($h_1$, $h_2$ e $h_3$ e $h_4$) e um neurônio de saída (y). Na Fig. 5, incluímos a representação dos pesos sinápticos ($w_{ij}$) e as funções de ativação dos neurônios (f).

*Figura 6: Rede neural multilayer perceptron com uma camada intermediária*




A rede neural esquematizada pode ser representada pela Eq. 24.

$$y = f_s\left( \sum_{j=1}^{4} w_j f_h\left( \sum_{i=1}^{2} w_{ij} x_i \right) \right) \tag{43}$$

Como podemos observar, a função y é dependente das entradas $x_i$ e dos pesos, $w_i$ e $w_j$. Estes, por sua vez, são não lineares, uma vez que variam conforme as equações $f_h$ e $f_s$. Logo, os pesos w são equivalentes aos parâmetros $\hat{\theta}$ da regressão não linear apresentada na seção anterior. Desta forma, podemos calcular o intervalo de confiança para uma estimativa feita pela função y = f(x, w) através da expressão apresentada previamente.

A rede pode apresentar várias camadas ocultas. Escolhemos uma rede com três camadas pela simplicidade do tratamento matemático, bem como pela conseqüência do Teorema de Kolmogorov-Nielsen, onde a mesma pode aproximar qualquer função não linear contínua e diferençável, ao menos localmente, o que decorre que a rede tem capacidade de modelar o mercado imobiliário

Considerando as que as funções de ativação sejam a linear e a sigmóide logística, dadas pelas Eqs. 25 e 26

$$fs(x) = x \tag{44}$$

$$f_h(x) = \frac{1}{1 + e^{-x}} \tag{45}$$

obtemos a função resposta para Y, em função dos pesos e das entradas, acordo com Eq. 27:

$$Y = \frac{w_1}{1+e^{-(w_{11}X_1 + w_{21}X_2)}} + \frac{w_2}{1+e^{-(w_{12}X_1 + w_{22}X)}} + \frac{w_3}{1+e^{-(w_{13}X_1 + w_{23}X_2)}} + \frac{w_4}{1+e^{-(w_{14}X_1 + w_{24}X_2)}} \quad (46)$$

O modelo acima pode ser utilizado para calcular as saídas y dadas as variáveis de entrada, $x_1$ e $x_2$. Busca-se então os pesos sinápticos $w_1$, $w_2$, $w_3$, $w_4$, $w_{11}$, $w_{12}$, $w_{13}$, $w_{14}$, $w_{21}$, $w_{22}$, $w_{23}$, $w_{24}$ que forneçam a melhor aproximação para a variável de saída.

Consideremos o conjunto de dados apresentado na Tab. 1.

*Tabela 1: Dados para treinamento de rede neural*

| Vu (y) | Local ($x_1$) | Área ($x_2$) |
|---|---|---|
| 473,68 | 2,00 | 665,00 |
| 600,00 | 3,00 | 600,00 |
| 675,00 | 3,00 | 600,00 |
| 675,00 | 3,00 | 600,00 |
| 670,53 | 3,00 | 604,00 |
| 670,53 | 3,00 | 604,00 |
| 691,75 | 3,00 | 618,00 |
| 750,00 | 3,00 | 630,00 |
| 261,90 | 2,00 | 1.890,00 |
| 951,34 | 3,00 | 596,00 |
| 983,19 | 3,00 | 595,00 |
| 898,62 | 3,00 | 651,00 |
| 504,40 | 3,00 | 1.249,00 |
| 1.270,59 | 4,00 | 595,00 |
| 85,96 | 1,00 | 8.900,00 |
| 1.554,28 | 4,00 | 608,00 |
| 1.813,95 | 4,00 | 645,00 |
| 3.450,00 | 5,00 | 600,00 |
| 225,00 | 1,00 | 10.000,00 |
| 288,00 | 1,00 | 10.000,00 |
| 1.598,89 | 3,00 | 2.533,00 |
| 1.928,57 | 4,00 | 2.100,00 |
| 2.571,43 | 4,00 | 2.100,00 |
| 424,53 | 2,00 | 742,00 |
| 680,67 | 3,00 | 595,00 |
| 639,20 | 3,00 | 704,00 |
| 1.111,76 | 3,00 | 425,00 |
| 780,99 | 3,00 | 605,00 |
| 832,46 | 3,00 | 573,00 |
| 831,93 | 3,00 | 595,00 |

Faz-se o treinamento, obtendo os pesos sinápticos apresentados na Tab. 2 a seguir.

*Tab. 2: Pesos sinápticos da rede*

| Peso | Valor |
|---|---|
| w1 | -0,355948 |
| w2 | 0,238536 |
| w3 | 0,441313 |
| w4 | -0,416180 |
| w11 | 0,455222 |
| w12 | -0,266306 |
| w13 | -0,443213 |
| w14 | -0,350105 |
| w21 | 0,424081 |
| w22 | 0,583436 |
| w23 | 0,168560 |
| w24 | -0,173391 |

Tomando como ponto de partida a Eq. 22, calculando-se as derivadas necessárias para as Eq. 10 e 16, obtemos a forma analítica da expressão do Intervalo de Confiança. Como trata-se de um intervalo para uma estimativa pontual, adotamos $x_1$ = 3 e $x_2$ = 1.000,00, resultando em um y = 936,64.

Tendo-se os dados de treinamento é possível calcular s pela Eq. 17, sendo s = 232,42. Tomando estes, bem como os valores Local e Área da estimativa pontual, é possível calcular o termo $f_o^T(F^TF))^{-1} f_o$ cujo valor é 0,57735.

Considerando-se um nível de confiança de 95%, determina-se o termo $t_{n-p}^{\alpha/2}$ pela distribuição t de Student, sendo o valor 0,461968.

Com os termos determinados, através da Eq. 18 chega-se ao intervalo de confiança para a estimativa pontual para as variáveis de entrada Local = 3 e Área = 1.000,00 e resultado Vu = 936,64.

IC = 61,99
Vu = 936,64 ± 61,99

## 6. CONCLUSÕES

A determinação de intervalos de confiança em redes neurais artificiais é fundamental para analisar a confiabilidade de predições advindas das mesmas. Para o caso específico da engenharia de avaliações, as normas que regem os procedimentos avaliatórios tomam como base o intervalo de confiança para determinar a precisão da avaliação.

O método de cálculo exposto neste artigo demanda relativo esforço matemático, sobretudo no emprego de derivadas parciais de equações complexas e cálculo matricial. Também exige o conhecimento detalhado da arquitetura da rede

treinada, inclusos aí número de camadas intermediárias, número de neurônios nas camadas, pesos sinápticos e funções de ativação.

No entanto, tais esforços são compensados, a medida que é possível inferir a confiabilidade da rede treinada com finalidade de avaliação, permitindo ao avaliador determinar se a mesma apresenta grau de precisão adequado.

Além disso, é possível implementar programas de computador que calculem os dados necessários a estimativa do intervalo de confiança, bem como a sua análise, assim como já é feito nos softwares de regressão linear. Com efeito, existem algoritmos bem definidos para cálculo numérico de derivadas parciais, matrizes Jacobianas, inversões e transposições de matrizes e decomposições singulares.

O intervalo de confiança também se presta a analisar se a arquitetura da rede treinada, a qual é definida *a priori*, é adequada à resolução do problema e obtenção do valor perseguido.

Naturalmente, as hipóteses adotados no método apresentado podem trazer alguma distorção no cálculo dos intervalos de confiança, sobretudo na adoção da premissa de erros com distribuição normal e no cálculo de matrizes, onde pode haver necessidade de hipóteses adicionais. Entretanto, acreditamos que tais inconvenientes não o desmereçam, e, considerando que as redes neurais artificiais são uma metodologia emergente no campo da engenharia de avaliações, o cálculo do intervalo de confiança para valores preditos pelas mesmas é primordial para seu bom uso.

Enfim, existem dois problemas no cálculo do intervalo de confiança em redes neurais. O primeiro é inerente à formulação matemática das mesmas. Segundo Geman, Bienenstock e Doursat (1992), redes neurais com grande poder de generalização usualmente apresentam elevada variância, acarretando que, para boas estimativas pontuais, paga-se o preço de maior intervalo de confiança. Outro ponto, de acordo com Chryssolouris, Lee e Ramsey (1996), é de que ocorre uma subestimativa dos graus de liberdade em redes neurais, uma vez que as mesmas são treinadas com mais parâmetros do que o necessário. No caso deste último, as análises do número de condição da matriz decomposta singularmente ou o cômputo do matriz ortogonal e as verificações de (40) e (41) podem trazer informações valiosas para concluir sobre a ocorrência de excesso de ajuste, escolha correta da arquitetura e eventual necessidade de novos treinamentos.

Fica como recomendação para futuros trabalhos, o cálculo de intervalos de confiança em redes neurais utilizando os métodos bootstrap, de máxima verossimilhança e inferência Bayesiana, e sua comparação com o método da regressão não linear.

## 7. BIBLIOGRAFIA


ABNT – Associação Brasileira de Normas Técnicas. *NBR 14653-2/2004. Avaliação de Bens. Parte 2: Imóveis urbanos*. Rio de Janeiro: ABNT, 2004;

AZEVEDO, Fernando M.; Lourdes M. Brasil e Roberto C. L. Oliveira. *Redes neurais com aplicação em controle e em sistemas especialistas*. Florianópolis: Editora Bookstore; 2000;

BRAGA, Antônio; André C. de Carvalho e Teresa B. Ludemir. Redes neurais artificiais *in* REZENDE, Solange O. (organizadora). *Sistemas inteligentes*. Barueri: Editora Manole; 2003;

CHRYSSOLOURIS, George; Moshin Lee e Alvin Ramsey. Confidence interval prediction for neural network models. *IEEE Transactions on Neural Networks*; Vol. 7, No. 1, Janeiro, 1996, pg. 229-232;

DANTAS, Rubens A. *Engenharia de avaliações. Uma introdução a metodologia científica*. São Paulo: Editora Pini; 1999;

DE VEAUX, Richard D., J. Schumi, J. Schwinsberg e Lyle Ungar. Prediction intervals for neural networks via nonlinear regression. *Technometrics*, Vol. 40, No. 4, 1994, pg. 273-282;

DONALDSON, Janet R. e Robert Schnabel. Computational experience with confidence regions and confidence intervals for nonlinear least squares. Boulder: University of Colorado, Department of Computer Science; 1985;

DYBOWSKI, Richard e Stephen J. Roberts. Confidence intervals and prediction intervals for feed-forward neural networks *in* DYBOWSKI, R. e GANT, V. (organizadores). *Clinical applications of artificial neural networks*. Cambridge: Cambridge University Press; 2001;

GEMAN, Stuart; Elie Bienenstock e René Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, Vol. 4, 1992, pg. 1-58;

GONZÁLEZ, Marco A. S. e Carlos T. Formoso. Planta Genérica de Valores Utilizando Inteligência Artificial. In: *Congresso Brasileiro de Cadastro Técnico Multifinalitário*, 2000, Florianópolis. Anais eletrônicos...Florianópolis, 2000;

GUEDES, Jackson C. O Emprego da Inteligência Artificial na Avaliação de Bens. In *VIII Congresso Brasileiro de Engenharia de Avaliações e Perícias*, 1995, Florianópolis. Anais... Florianópolis: IBAPE/SC, 1995, pg. 368-374;

GUEDES, Jackson C. Duas Ferramentas Poderosas a Disposição do Engenheiro de Avaliações – Modelos Lineares Generalizados e Redes Neurais. In *XI Congresso Brasileiro de Engenharia de Avaliações e Perícias, 2001*, Guarapari. Anais... Vitória: IBAPE/ES, 2001. 1 CD;

GOLUB, G. H. e C. Reinsch. Singular value decomposition and least square solutions. *Numerical Mathematics*. Vol. 14, 1970, pg. 403-420;

HECHT-NIELSEN, Robert. Theory of backpropagation neural network. In *Proceedings of the International Joint Conference on Neural Networks* (pg. I:593-608). San Diego, EUA: SOS Printing, 1989;

HORNIK, Kur', Maxwell Stinchcombe e Halber White. Multilayer feedforward networks are universal approximators. *Neural Networks*. Vol. 02, 1989, pg. 359-366;

LEITE, José Galvão e Júlio da Motta Singer. Métodos assintóticos em estatística. Fundamentos e aplicações. *9º Simpósio Nacional de Probabilidade e Estatística*. ABE – Associação Brasileira de Estatística: São Paulo, julho de 1990;

LIMA, Elon Lages. Análise geométrica*. 7º Colóquio Brasileiro de Matemática*. Poços de Caldas, 1969;

KOVÁCS, Zsolt László. *Redes Neurais Artificiais. Fundamentos e Aplicações, um texto básico*. 4ª Edição. São Paulo: Editora Livraria da Física, 2006;

MELO, Brício de. Mínimos Quadrados Eficazes, Redes Neurais Eficientes. In *III Simpósio Brasileiro de Engenharia de Avaliações*, 2002, São Paulo. Anais eletrônicos...São Paulo, 2002;

NGUYEN, Nghiep e Al Cripps. Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks. *Journal of Real Estate Research*, V. 22, n. 3, p. 313-336;

OLIVEIRA, Mauri A.; Alessandra de A. Montini e Daniel R. Bergman. Construção de intervalos de predição para redes neurais via regressão multivariada e sua aplicação em séries heterocedásticas. *Facef Pesquisa*, Vol. 10, No. 3, setembro, 2007, pg. 271-283;

RIVALS, Isabelle e Léon Personnaz. Construction of confidence intervals in neural modeling using a linear Taylor expansion. *Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling*. 8-10 Julho, 1998, Leuwen, pg. 17-22;

RIVALS, Isabelle e Léon Personnaz. Construction of confidence intervals for neural networks based on least squares estimation. *Neural Networks*, Vol. 13, 2000, pg. 463-484;

RIVALS, Isabelle e León Personnaz. Jacobian conditioning analysis for model validation. *Neural Computation*, Vol 16, 2004, pg. 401-418;

UNGAR, Lyle H.; Richard D. de Veaux e Evelyn Rosengarten. Estimating prediction intervals for artificial neural networks. *Proceedings of the 9th Yale Workshop on Adaptive and Learning Systems*; 1996;

YANG, Luren, Tom Klavi, Mats Carlin, Sigmund Clausen e Paul F. M. De Groot. An evaluation of confidence bound estimation methods for neural networks. *ESIT 2000*, Aachen, Alemanha, 14 – 15 de setembro, 2000;

ZAPRANIS, Achilleas e Efstratios Livanis. Prediction intervals for neural network models. *ICCOMP'05 Proceedings of the 9th WSEAS International Conference on Computers*, Stevens Point, Wisconsin, EUA, 2005.