

# PREVISÃO DO IBOVESPA POR MEIO DA REDE NEURAL ARTIFICIAL GMDH. Daniel Charleaux Roque, Nair Cristina Margarido Brondino. - Inter-áreas - Departamento de Engenharia Elétrica, Departamento de Matemática - Faculdade de Ciências - Campus de Bauru.

O objetivo deste trabalho é apresentar os resultados obtidos através da utilização de uma rede neural artificial, GMDH (*Group Method of Data Handling*), na previsão da série temporal do índice econômico-financeiro IBOVESPA. O desempenho do modelo proposto será analisado a partir da coerência com o comportamento real da série, em termos dos valores previstos. Desta forma, os resultados aqui apresentados não serão discutidos sob o ponto de vista da teoria econômica.

Uma série temporal é definida como sendo qualquer conjunto de observações ordenadas no tempo. A análise de séries temporais, por sua vez, objetiva construir modelos com propósitos bem determinados para as séries. Entre as técnicas que desempenharam papel importante no desenvolvimento teórico dos métodos para previsão deste tipo, destaca-se a metodologia proposta por Box e Jenkins (Morettin & Toloi, 2004), desenvolvida no início da década de 70 e que foi proposta para possibilitar a construção de modelos auto-regressivos integrados de médias móveis. Estes modelos vêm sendo extensivamente estudados na literatura e têm apresentado resultados bastante satisfatórios. Entretanto, deparam-se com o fato de que em muitos problemas reais pode ocorrer a ausência de informações suficientes para gerar as equações matemáticas que governam o comportamento do fenômeno em estudo. Nestes casos, pode-se usar uma abordagem baseada em modelos, como é o caso das Redes Neurais Artificiais (RNA), em que a rede é treinada para abstrair as equações geradoras de uma série temporal.

A motivação para a utilização de tal ferramenta neste estudo é que, além de suas propriedades de não linearidade e habilidade em extrair informações relevantes dentro de um conjunto extenso de dados, as redes neurais têm sido utilizadas como uma boa alternativa aos modelos estatísticos convencionais. Sua utilização já vem sendo feita em estudos de previsões de índices financeiros, com algum sucesso (Oliveira, 2003). Entretanto, não foi constatada na literatura nacional a utilização de Redes Neurais do tipo GMDH neste tipo de previsão.

A rede neural GMDH é um tipo de rede auto-organizável, que, por sua estrutura, pode ser utilizada para fazer previsões de curto e médio prazo. Neste tipo de rede, uma relação geral entre entrada e saída pode ser descrita na forma de um modelo matemático, que também é chamado de função referência. Usualmente, esta descrição é considerada como uma forma discreta da Série Funcional de Volterra, apresentada na expressão (1).

$$y = a_0 + \sum_{i=1}^m a_i x_i + \sum_{i=1}^m \sum_{j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ijk} x_i x_j x_k + \dots \quad (1)$$

onde  $y$  é a saída,  $x = (x_1, x_2, \dots)$  é o vetor de entrada e  $a$  é o vetor de pesos. Como pode ser observado, a função dada em (1) é linear nos pesos, mas não linear nas variáveis. A forma parcial desta função, como um estado ou função soma, é desenvolvida em cada unidade simulada e então é ativada em paralelo para a construção mais complexa (Madala & Ivakhnenko, 1994).

Durante o processo de treinamento, cada unidade simulada  $k$  recebe um par de variáveis de entrada  $(x_i, x_j) \subset x$ ,  $i \neq j$  e gera uma função  $f()$ , que, por sua vez, é uma forma parcial da função de referência. A equação (2) mostra a função parcial  $f(x_i, x_j)$ .

$$f(x_i, x_j) = v_0^{(k)} + v_1^{(k)} x_i + v_2^{(k)} x_j + v_3^{(k)} x_i x_j + v_4^{(k)} x_i^2 + v_5^{(k)} x_j^2 \quad (2)$$

onde os  $v^{(k)}$  representam os pesos. A escolha de uma função como dada em (2) parece adequada, pois, segundo Oliveira (2003), “... o modelo linear pode falhar em capturar ou prever pontos de mudança acentuada nos dados...”. Desta forma, se  $m$  variáveis de entrada forem consideradas, então o número total de funções parciais como a mostrada em (2), que serão geradas na primeira camada intermediária,

será dado por  $\binom{m}{2}$ . Se denotarmos por  $o$  a saída desejada, então o erro estimado em cada unidade é dado por (3).

$$e_p = y_p - o_p; p \in N_A, \quad (3)$$

onde  $N_A$  denota o conjunto de treinamento.

O erro quadrático de treinamento para cada vetor de entrada é dado pela expressão (4). Os pesos  $v^{(k)}$  devem ser tais que conduzam ao erro quadrático mínimo e são obtidos pelo Método dos Mínimos Quadrados. Cada camada contém um grupo de unidades que estão interconectadas com as unidades da próxima camada. Os pesos em cada camada devem minimizar o erro  $E$ . O valor de uma função objetivo é usado como limiar para ativar ou desativar cada unidade. Esta função é comparada com a saída gerada quando outro subconjunto de dados (conjunto de teste), denotado por  $N_B$  é apresentado à rede. As unidades ativadas em uma camada são, então, utilizadas como entrada na próxima camada e assim por diante.

$$E = \sum_{p \in N_A} e_p^2 \quad (4)$$

Com base nesta teoria, foi construída pelos autores uma rotina no software MatLab, denominada GMDHNet. Após ler o vetor de entrada, consistindo dos índices diários IBOVESPA relativos ao período de 04 de julho de 1994 a 29 de setembro de 2000, tal rotina divide o conjunto de dados aleatoriamente em dois subconjuntos: um para treinamento da rede e um para teste, sendo que os últimos 2% dos valores da série temporal são armazenados para a fase de validação.

A memória é inserida na rede através de atrasos no tempo, que são implementados nos nós de entrada. Os nós da primeira camada escondida combinam as entradas duas a duas, de acordo com a função apresentada em (2). O tamanho do *lag* (e conseqüentemente o número de nós de entrada) é determinado através de simulações, a partir de uma comparação da saída da rede e aqueles dados reservados para a fase de validação. Nesta etapa, podem ser elaboradas previsões de quantos passos à frente se desejar.

As simulações realizadas indicaram que o modelo cujas previsões mais se aproximaram do comportamento real da série foi aquele que apresentou tamanho de *lag* igual a 9 e que será denominado de LAG9. Tais conclusões basearam-se em resultados gráficos, como os apresentados nas Figuras 1 e 2, para duas janelas de tempo distintas, que confrontaram a qualidade do modelo ajustado com o comportamento real da série para este conjunto. Como pode ser observado nestas figuras, o modelo ajustado conseguiu captar a tendência da série, mostrando coerência com o comportamento real desta.

Os resultados apresentados aqui indicaram que a utilização de uma Rede Neural GMDH para fazer previsões da série do índice IBOVESPA apresentou resultados satisfatórios, do ponto de vista da qualidade do ajuste, em termos numéricos. Entretanto, o presente trabalho limitou-se apenas a estudar a qualidade do método proposto, em termos de ajuste ao conjunto de dados, sem, contudo, analisar sua qualidade quanto a previsões. Desta forma, para trabalhos futuros, pretende-se estudar o comportamento das previsões fornecidas pelo modelo através do cálculo e análise das estatísticas de erro de previsão MAE (Erro Absoluto Médio - *Mean Absolute Error*), RMSE (Raiz do Erro Quadrático Médio - *Root Mean Square Error*) e TIC (Coeficiente de Desigualdade de Theil - *Theil Inequality Coefficient*), descritas no trabalho de Pindyck & Rubinfeld (2004), além de outros tipos de análises pertinentes.

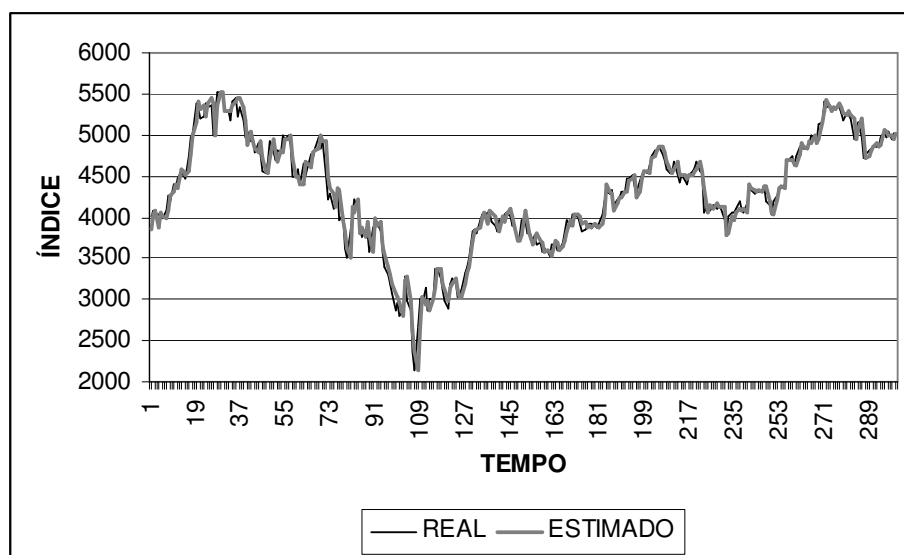


Figura 1 – Valores previstos versus dados reais para as 300 primeiras observações do conjunto de dados.

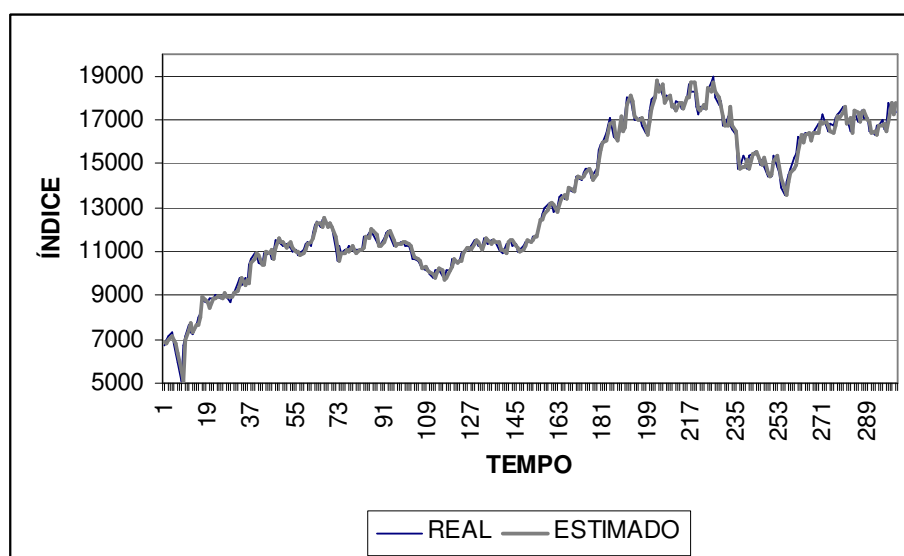


Figura 2 – Valores previstos versus dados reais para as 300 últimas observações do conjunto de dados.

### Referências Bibliográficas

- Madala, H. R.; Ivakhnenko, A. G. (1994). *Inductive Learning Algorithms for Complex System Modeling*. CRC Press. Florida.
- Morettin, P. A.; Toloi, C. M. C. (2004). *Análise de Séries Temporais*. Edgard Blücher Ltda. São Paulo.
- Oliveira, M. A. (2003). *Previsão de sucessões cronológicas econômico-financeiras por meio de redes neurais artificiais recorrentes de tempo real e de processos ARMA-GARCH: um estudo comparativo quanto à eficiência de previsão*. Dissertação de Mestrado. São Paulo: FEA/USP, 195 p.
- Pindyck, R.S; Rubinfeld, D.L. (2004). *Econometria*. Elsevier. Rio de Janeiro.