

Previsão de séries temporais mediante redes neurais

Aluna: João Pedro Mano
Orientadora: Celia Anteneodo

Introdução

Atualmente, com o crescimento dos mercados financeiros, e com as mudanças bruscas nos preços dos ativos, os profissionais da área necessitam de métodos eficientes para anteciparem-se aos acontecimentos. Diversas técnicas têm sido usadas a fim de prever o comportamento desses ativos. Por isso, o uso de sistemas inteligentes de análise de decisão tem sido cada vez mais freqüente nessa área.

Uma das técnicas de inteligência computacional, comumente usada na tentativa de prever séries temporais, é o treinamento de redes neurais. Estas são baseadas na arquitetura e aprendizagem do cérebro humano.

Essas redes foram apresentadas pela primeira vez em 1943 pelo neurofisiologista McCulloch, e pelo matemático Walter Pitts[1].

O modelo de Pitts e McCulloch era constituído de apenas uma saída e uma função de entrada, calculada através da soma ponderada de diversos valores.

Esse modelo assemelha-se com o neurônio humano, onde sinais são recebidos pelo mesmo com intensidades diferentes, provocando uma determinada reação por parte da célula nervosa.

Os sistemas mais atuais são mais complexos. Eles são formados por estruturas semelhantes à figura acima, interligadas, formando uma rede desses “neurônios artificiais”.

O sistema utilizado nas simulações deste trabalho são agrupamentos de estruturas semelhantes ao modelo de Pitts e McCulloch, chamado de “multi layered perceptron model”(MDL). Nele, uma série de dados entra pelas camadas de entrada, interagem entre si nas camadas ocultas, e depois saem pela camada de saída.

O aprendizado dessas redes vem dos erros cometidos pelas mesmas. O resultado de saída delas é comparado com o valor desejado de saída (conhecido a priori). O erro cometido pelo sistema serve para que os pesos das ligações entre os neurônios artificiais sejam re-determinados, até que o valor de saída se assemelhe com o desejado.

As redes neurais têm alto poder de aprendizagem. Muitas técnicas são usadas com essa finalidade. A usada neste estudo foi a “retropropagação”, apresentada por [2].

A eficiência desse sistema, em séries que apresentam certa regularidade, ou um certo padrão, é extremamente alta. Entretanto, em séries mais voláteis como a usada na segunda parte deste estudo, a previsão obtida, apesar de ter uma qualidade menor que em séries mais regulares, auxilia na tomada de decisão.

Desenvolvimento

A Rede Neural

– Estrutura da Rede

A rede citada por [2] é composta de diversas camadas, denotadas por $l = 0, \dots, L$, onde a camada $l = 0$ é chamada de entrada, e a camada $l = L$ de camada de saída, e as demais são conhecidas como camadas ocultas. Na rede construída para as simulações desse estudo, há apenas uma camada oculta.

Os valores referentes à série temporal, entram pela camada “input” e são transformados para valores pertencentes ao intervalo $[0,1]$, através da seguinte fórmula abaixo:

$$x_j = f(u_j) \quad (1)$$

Sendo $f(x)$ a função abaixo:

$$f(x) = (x - \min) / (\max - \min) \quad (2)$$

Onde \max = maior valor da série temporal utilizada, e \min = menor valor da mesma.

Assim, os valores transformados na camada de entrada, são transmitidos para a camada seguinte de forma ponderada. Ou seja, cada dado tem um peso diferente na entrada da camada posterior. Inicialmente, os pesos são atribuídos de forma aleatória, assumindo valores também no intervalo $[0,1]$. Posteriormente, esses pesos serão reavaliados durante o período de aprendizagem do sistema, até assumirem seus valores devidos.

Na camada oculta, os neurônios artificiais assumem valores calculados da seguinte forma:

$$u_i = \sum_{j=1}^n w_{ij} \cdot x_j \quad (3)$$

Onde w_{ij} é o peso do valor do neurônio de entrada j no neurônio i da camada oculta.

Da mesma forma que na camada de entrada, os valores da camada oculta também são transformados. Entretanto, utiliza-se uma função $g(y)$ diferente da primeira.

$$g(y) = \frac{1}{1 + e^{-y}} \quad (4)$$

De forma análoga, os valores da camada de saída da rede neural utilizada são calculados através da soma ponderada dos valores de saída da camada oculta. Os valores de saída do sistema são transformados através da fórmula:

$$x_L = u_L (\max - \min) + \min \quad (5)$$

- O aprendizado da Rede

Usando a rede descrita acima, inicialmente os valores obtidos diferem muito dos desejados, devido à atribuição aleatória dos pesos iniciais. Assim, o sistema precisa passar por um processo de aprendizagem para, através dos dados conhecidos, chegar ao valor correto desses pesos.

Para tal, existem diversos modos de fazer com que a rede aprenda. Neste estudo usou-se o retropropagação, método baseado na aprendizagem através dos erros.

O quanto o peso deve ser alterado num determinado tempo t , é dado da seguinte maneira:

$$\Delta w_{ij}^{(l)}(t) = \eta \delta_i^{(l)} x_j^{(l-1)} + \alpha \Delta w_{ij}^{(l)}(t-1) \quad (6)$$

onde, η é a taxa de aprendizado do sistema, e α é o momentum. Ambos pertencem ao intervalo (0,1) e são fixados no início do processo. Para cada série usada existe uma combinação ótima dessas variáveis.

O δ_i é definido de duas maneiras. Para a camada output, ele é determinado como:

$$\delta_i^{(L)} = (d_i - x_i^{(L)}) g'(u_i^{(L)}) \quad (7)$$

onde $g'(u_i^{(L)})$ é a derivada no ponto $u_i^{(L)}$ da função $g(x)$, descrita acima, e d_i é o valor que deveria ser obtido pelo previsão do sistema.

Para as demais camadas, o $\delta_i^{(l)}$ é definido como:

$$\delta_i^{(l)} = g'(u_i^{(l)}) \sum_{k=1}^{n(l+1)} \delta_k^{(l+1)} w_{ki}^{(l+1)} \quad (8)$$

Determinada a variação dos pesos no tempo t o novo valor do peso é calculado pela soma do anterior com a variação:

$$w_{ij}^{(l,novo)} = w_{ij}^{(l)} + \Delta w_{ij}^{(l)}(t) \quad (9)$$

O processo de aprendizagem é realizado diversas vezes com a finalidade de obter os pesos corretos.

Séries Regulares

O processo descrito acima tem grande poder de aprendizagem em séries que apresentam um determinado padrão. Para testar o poder da rede criada, a série abaixo, que é extremamente regular, foi usada:

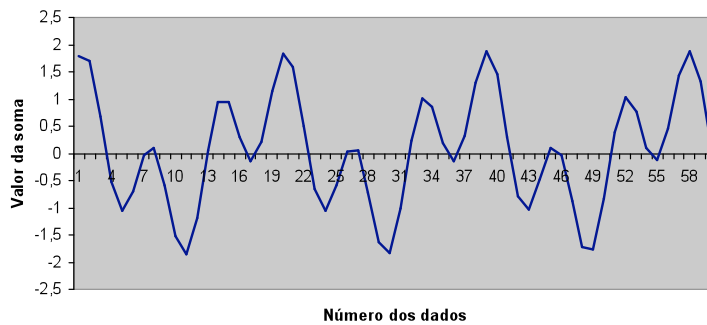


Figure 3 – O Gráfico acima representa a soma de senos e cossenos.

A rede criada para realizar essa previsão teve as seguintes características, descritas na tabela abaixo:

Tabela 1 – A tabela abaixo mostra as características da rede neural criada para prever a série regular.

# de Neurônios da Oculta	# de dados de entrada	# de dados usados no aprendizado	# de repetições do processo de aprendizado
10	10	40	6000

O sistema criado é composto de três camadas, a de entrada, a oculta e a de saída. Contando com dez neurônios de entrada, dez na camada oculta, e um neurônio de saída, onde sai a previsão obtida pelo algoritmo.

Para realizar cada uma das seis mil repetições do processo de aprendizado, um número aleatório i é gerado pelo sistema, e o i -ésimo termo da série temporal é o primeiro dado de entrada, sendo seguido pelos nove subsequentes a ele, completando assim os dez dados de entrada utilizados pela rede. Com eles então, a rede previa o décimo primeiro dado – a partir do i -ésimo -, comparava com o verdadeiro e calculava o erro, dando, assim, início ao processo de aprendizado por retropropagação.

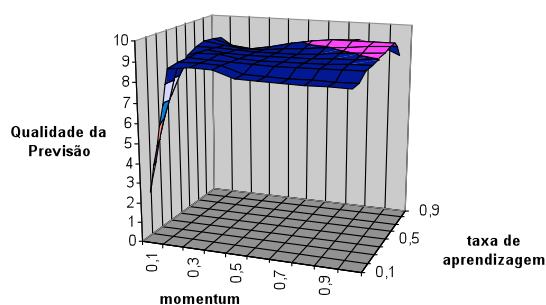
Após as seis mil repetições, a rede aprende o padrão da série temporal em estudo, e seus pesos são corrigidos, a fim de correlacionar os dados de entrada para obter previsões eficientes. Depois disso, com os novos pesos, chega ao fim a fase de aprendizado, e o sistema usa o que “aprendeu” para gerar os resultados da previsão, neste caso o sistema repetiu o processo de previsão vinte vezes.

A qualidade da previsão, depende também da escolha correta dos valores que as variáveis momentum e a taxa de aprendizado devem assumir. Assim, com a finalidade de obter a combinação ótima dessas variáveis η e α , o sistema foi testado para diversos desses parâmetros, dentro do intervalo (0,1). O erro de cada combinação é dado da seguinte maneira:

$$Erro = \sqrt{\frac{\sum_{i=1}^n (a_i - \hat{a}_i)^2}{n}} \quad (10)$$

onde a_i é o valor real da série, e \hat{a}_i é o valor estimado pelo algoritmo.

A qualidade da previsão referente a cada valor de η e α pode ser visualizada no gráfico abaixo:

Figure 4– Qualidade de previsão com cada par (η, α) de valores.

Pelo gráfico, nota-se que a combinação mais eficiente dos dois valores foi $\eta = 0,9$ e $\alpha = 0,8$.

Com a determinação desses valores, o resultado obtido pelo sistema pode ser visualizado abaixo:

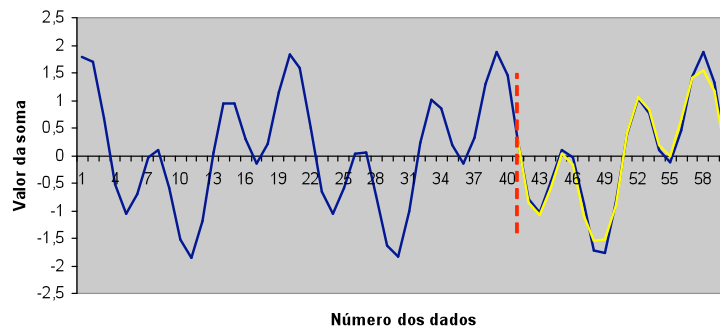


Figure 5 – Gráfico da série temporal e da previsão realizada pela rede neural.

Claramente, o resultado da previsão da série pela rede neural é extremamente eficiente. Os valores obtidos pelo sistema diferem numa escala bem reduzida dos reais, confirmando a qualidade de previsão do algoritmo.

Séries Irregulares

A grande maioria das séries financeiras, objeto principal desse estudo, não apresentam uma regularidade tão clara quanto a série testada acima. Por isso, a previsão das mesmas torna-se extremamente complicada.

A dinâmica dos mercados financeiros torna quase que “aleatório” o valor de um ativo. E a correlação entre o valor de amanhã e os valores até o dia de hoje, nem sempre são tão evidentes. Assim, o uso de redes neurais para a previsão dessas séries não é tão eficiente como para as que apresentam regularidade. Entretanto, o uso de sistemas neurais pode servir de método auxiliar para a tomada de decisão quanto aos futuros valores dessas séries muito voláteis.

O gráfico abaixo, do comportamento da ação da Petrobrás no mercado Dow Jones, mostra a dificuldade de localizar um certo padrão nesse tipo de série financeira.

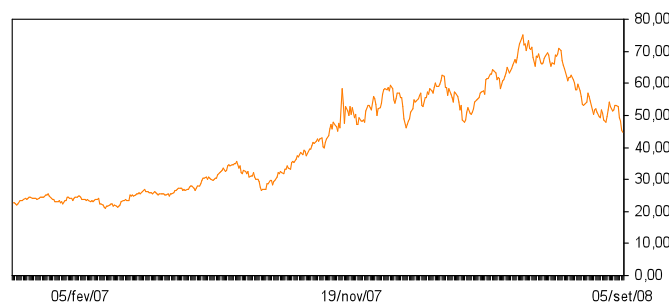


Figure 6 – Preço diário de fechamento (em US\$) do Petróleo Brasileiro na bolsa de Nova York.

Pela observação do gráfico fica clara a inexistência de uma regularidade nesse tipo de série temporal. Na tentativa de prever o comportamento da série acima, uma rede semelhante à usada no exemplo anterior foi construída, mas com características um pouco diferentes.

A tabela abaixo descreve as características dessa nova rede criada.

Tabela 3 – Características da Rede Neural usada na previsão do comportamento da ação da Petrobrás.

# de Neurônios na camada Input	# de Neurônios na camada oculta	# de dados usados na aprendizado	# de Repetições do processo de aprendizado	Taxa de aprendizado	Momentum
28	30	399	6000	0,1	0,9

A rede descrita acima difere um pouco das redes usuais, onde o número de neurônios artificiais da camada de entrada é igual ou maior do que o da camada oculta. Entretanto, a rede construída nesse molde apresentou resultado superior na previsão da tendência principal do comportamento do ativo.

Usando a rede para prever o comportamento dos 40 dias seguintes ao último dado fornecido ao sistema, ela produziu os seguintes resultados, descritos no gráfico abaixo:

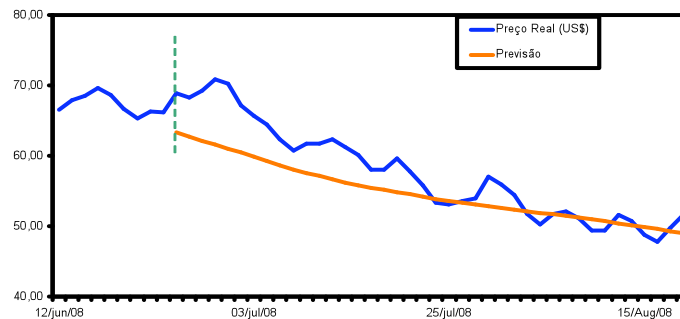


Figura 7 – Previsão do preço(em US\$) do Petróleo Brasileiro.

Claramente, a previsão do sistema ignorou as flutuações locais da série, e também os valores exatos, prevendo apenas a tendência forte de queda do ativo. Então, como a tendência foi captada pelo sistema de forma eficiente, os valores da previsão devem passar por uma transformação linear, a fim de ajudarem também na previsão de valores exatos. Para tal propósito, em vez de compararmos a previsão com os valores de fechamento do Petróleo Brasileiro, os valores máximos e mínimos são mais úteis, porque revelam os preços atingidos durante o pregão. A comparação da previsão com esses valores pode ser vista abaixo:

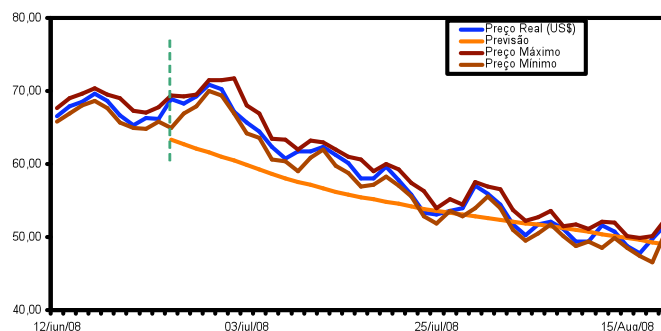


Figure 8 – Gráficos dos Valores (em US\$) do Petróleo Brasileiro.

A previsão realizada pela rede neural, para ajudar nos valores exatos do preço e não apenas na tendência forte, deve sofrer uma transformação linear $h(x)$, que não altere a variação entre seus valores. A função $h(x)$ pode ser descrita da seguinte forma:

$$h(x) = \Phi x \quad (11)$$

onde,

$$\Phi = \frac{V_f}{P_1} \quad (12)$$

onde, V_f = a media entre o último valor conhecido dos preços máximos e mínimos, e P_1 = ao primeiro valor da previsão feita pela rede neural.

Além disso como os ativos são extremamente voláteis, é interessante construir uma margem de variação que a previsão possa assumir. Nesse estudo, essa variação foi baseada na variância da variação dos valores de fechamento entre dois dias.

O Canal de Variação Média é formado pela previsão transformada conforme descrita acima, permitindo ao preço do ativo X_i assumir valores do intervalo $[P_i - \sigma^2, P_i + \sigma^2]$.

Como no mercado acionário diversos fatores pontuais podem causar quedas e subidas não coerentes com o padrão captado pelo sistema, três canais de variação, semelhantes ao descrito, foram construídos tornando a previsão mais adaptável ao comportamento real dessas séries.

O Canal de Variação de Mínima, por sua vez, é formado pela previsão transformada conforme descrita acima, permitindo ao preço do ativo, X_i , assumir valores do intervalo $[P_i - 3\sigma^2, P_i - \sigma^2]$, e o Canal de Variação Máxima, é o intervalo $[P_i + \sigma^2, P_i + 3\sigma^2]$.

Os gráficos abaixo ilustram a construção desses canais:

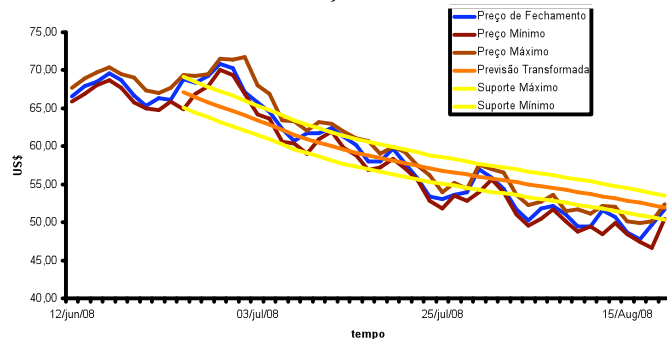


Figure 9 – Canal de Variação Média.

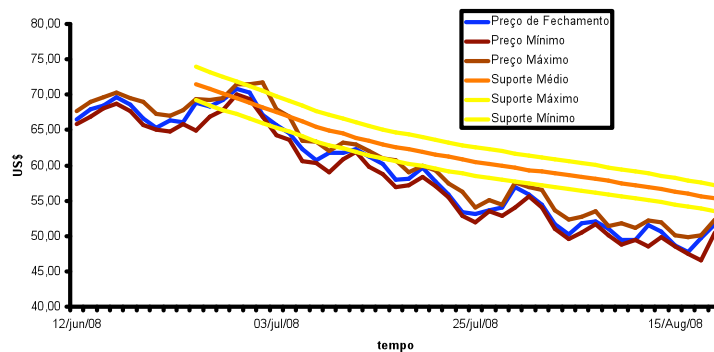


Figure 10 – Canal de Variação Máxima.

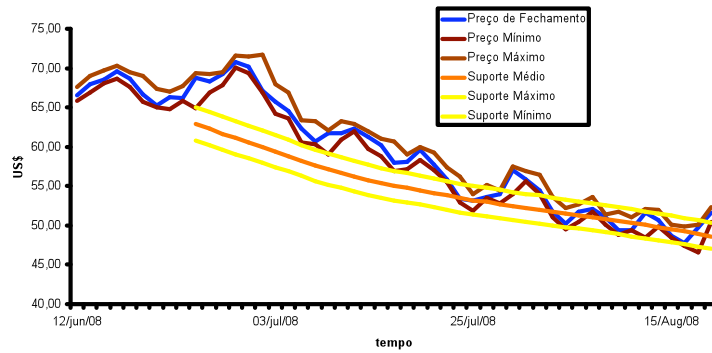


Figure 11 – Canal de Variação Mínima.

Os canais descritos acima, abrangem praticamente a totalidade dos valores assumidos pelo preço do ativo, no período de 40 dias após a primeira previsão. Se os três canais forem agrupados de forma correta, a previsão transformada, feita pela rede neural, oferece canais por onde o preço do ativo variará durante um certo numero de dias.

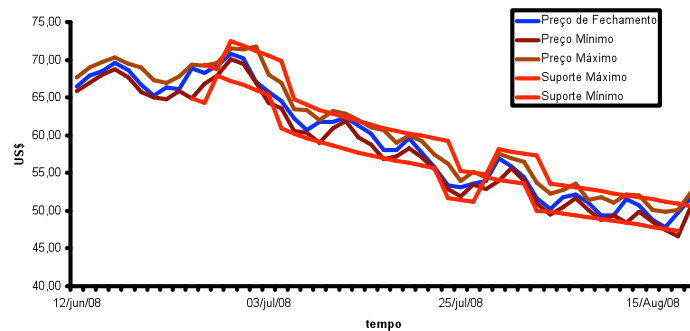


Figure 12 – Canais de Variação Combinados.

A combinação entre os três canais, fornece, além da tendência geral, suportes para os valores máximos e mínimos obtidos no dia. No exemplo abaixo, o mesmo critério para a determinação dos suportes foi usado. A série abaixo também representa o preço de ativos do Petróleo Brasileiro na Bolsa de Nova York. O período relacionado à previsão feita pelo sistema é um pouco anterior ao do primeiro exemplo.

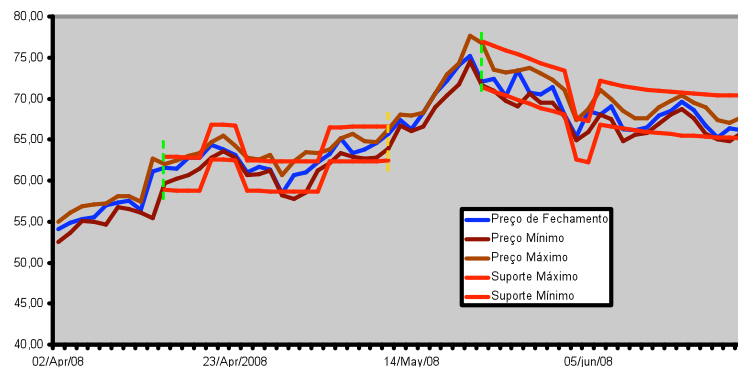


Figure 13 – Previsão para um outro trecho da série temporal referente ao preço(em US\$) do Petróleo Brasileiro.

No gráfico acima, as linhas pontilhadas verde e amarela representam, respectivamente, o início e término da previsão. Nesse exemplo, o sistema previu – a partir da primeira linha verde pontilhada – o comportamento praticamente “constante” do preço do ativo, entretanto, a previsão dada pelo sistema foi travada no dia 12 de maio de 2008, porque o preço de fechamento do ativo foi igual ao suporte máximo do Canal de Variação Máxima, denotando que a previsão da rede não era mais eficiente. A alta vertiginosa foi provocada pela valorização do barril de petróleo, e uma série de notícias boas que influenciaram no preço do ativo, entre elas a elevação da classificação da empresa pela agência de análise de risco americana Standard & Poor’s, fatos que fizeram com que o preço do ativo valorizasse tanto. Esse fatos, são imprevisíveis para o sistema, que trabalha apenas com a análise de dados antigos, por isso a falha do sistema durante esse pequeno período.

No dia 21 de maio, quando o fechamento foi menor do que o dia anterior, avisando que a alta vertiginosa podia ter chegado ao fim, outra previsão foi realizada, seguindo as mesmas características das previsões anteriores. Essa, como as anteriores, traçou o comportamento futuro do ativo de forma eficiente.

Conclusões

O uso de redes neurais na previsão de séries que apresentam alguma regularidade é extremamente eficiente. Entretanto a maioria das séries utilizadas em finanças não seguem um padrão muito claro. Mesmo assim, o sistema conseguiu traçar uma tendência geral. Nos dois trechos estudados ele conseguiu prever o que aconteceria com a série, desprezando as pequenas flutuações.

A utilização desses sistemas de forma isolada pode não conseguir prever séries dessa natureza de forma adequada. Se, entretanto, usado como mecanismo de apoio aliado a outros conhecimentos, pode chegar a conclusões ainda mais eficazes. Uma das idéias nessa linha, é a junção da lógica fuzzy com essas redes, nos chamados sistemas Neuro-Fuzzy, aliando o poder de aprendizado das redes neurais, com o poder de agir em ambientes de incerteza da lógica fuzzy.

Agradecimentos

Aproveito para agradecer à Professora Celia Anteneodo, do Departamento de Física da Puc-Rio, pela ajuda e orientação durante todo este estudo, e ao CNPQ, pela bolsa de iniciação científica que me foi concedida em Julho de 2008 para realizar este trabalho.

Referências

- [1] Warren McCulloch and Walter Pitts, "A Logical Calculus of Ideas Immanent in Nervous Activity", Bulletin of Mathematical Biophysics 5:115-133, 1943.
- [2] David E. Rumelhart, Geoffrey E. Hilton, Ronald J. Williams. "Learning representation by back-propagation errors." Nature, 323(9):533-536, Outubro 1986.
- [3] T. Kolarick, G. Rudorfer, "Time Series Forecasting Using Neural Networks", APL Quode Quad, 25:86-94, 1994.