

Uso de Redes Neurais

em Modelagem de Sistemas

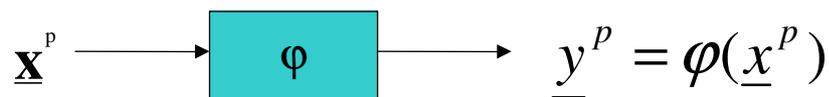
Luiz Calôba

caloba@ufrj.br , www.lps.ufrj.br/~caloba

I - Modelos Matemáticos de Sistemas Físicos

Sistema Real, Planta

$$(\underline{\mathbf{x}}^p, \underline{\mathbf{y}}^p) \quad p = 1, 2, \dots, P$$



Modelo, Simulador



Modelo Matemático

Sistema de equações tal que

saída modelo \approx saída planta

$$\underline{\tilde{y}}^p \cong \underline{y}^p \quad \forall p$$

equações (e seus parâmetros k) minimizam

$$F(k) = \underset{\forall p}{E} \left| \underline{y}^p - \underline{\tilde{y}}^p \right|^2$$

Modelos Fenomenológicos e Numéricos (Neurais)

Modelo Fenomenológico



Sistema de equações pré-determinado pela fenomenologia

Precisão das equações ?

Parâmetros das equações: conhecidos ou

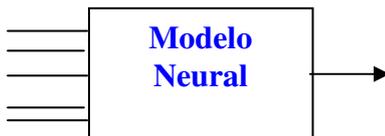
à determinar a partir de dados

experimentais

Cognitivo

Precisão ?

Modelo Neural



Sistema de equações:

$$y = \sum tgh(.) \quad \text{aproximador universal}$$

Parâmetros (sinapses) à ajustar

backpropagation $\Delta w = -\alpha \frac{\partial F}{\partial w} \quad F(w) = E_{\forall p} | \underline{y}^p - \tilde{y}^p |^2$

Preciso

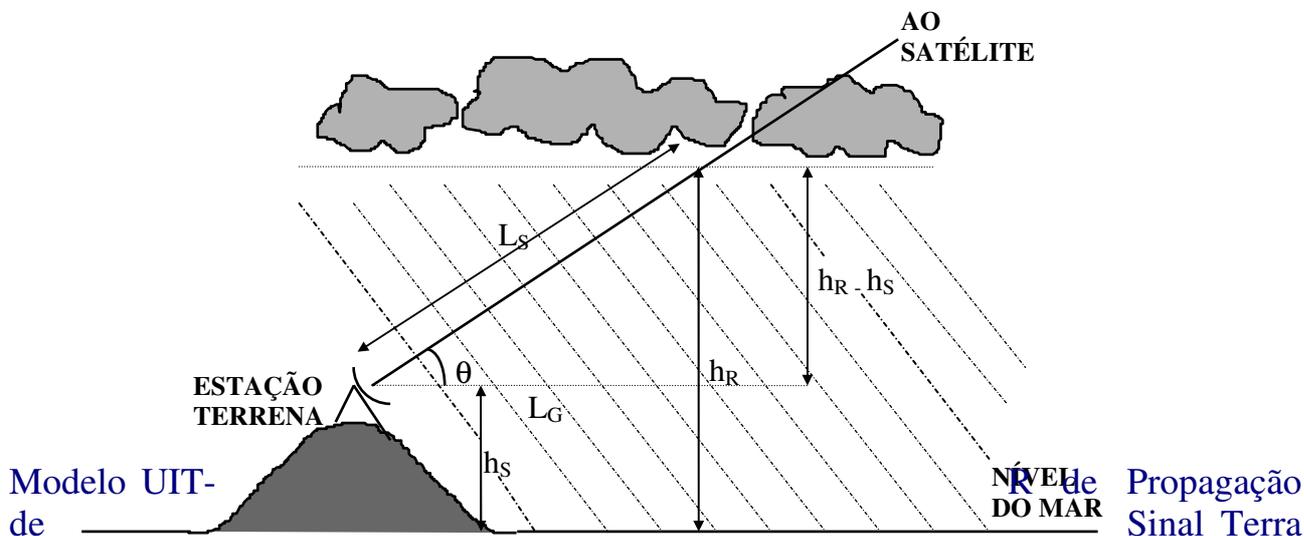
Informação fenomenológica ?

II - Modelos Fenomenológicos tratados como uma Rede Neural



Exemplo 1:

Atenuação de sinal em enlace terra-satélite



- Satélite - Equações

$$Z_1 = f(\varphi) = h_R$$

$$h_R = 5 - .075(\varphi - 23)$$

$$h_R = 5$$

$$h_R = 5 + .1(\varphi + 21)$$

$$h_R = 0 \quad \text{for } \varphi < -71$$

$$Z_2 = f(Z_1, h_R, \theta) = L_S,$$

$$L_S = \frac{(h_R - h_S)}{\text{sen } \theta} \quad \text{for } \theta \geq 5^\circ$$

$$L_S = \frac{2(h_R - h_S)}{\left(\text{sen}^2 \theta + \frac{2(h_R - h_S)}{R_e} \right)^{1/2} + \text{sen } \theta} \quad \text{else}$$

$$Z_3 = f(Z_2, \theta) = L_G = L_S \cos \theta$$

$$Z_4 = f(f, \tau, \theta, R_{0.01\%}) = \gamma_R = k(R_{0.01\%})^\alpha$$

$$Z_5 = f(Z_3, Z_4, f) =$$

$$r_{0.01} = \frac{1}{1 + 0.78 \sqrt{\frac{L_G \gamma_R}{f}} - 0.38(1 - e^{-2L_G})}$$

$$Z_9 = f(Z_4, Z_7, Z_8, f, \theta) =$$

$$v_{0.01} = \frac{1}{1 + \sqrt{\text{sen } \theta} \left(31(1 - e^{-(\theta/(1+\tau))}) \frac{\sqrt{L_R \gamma_R}}{f^2} - 0.45 \right)}$$

$$Z_{10} = f(Z_7, Z_9) = L_E = L_R v_{0.01}$$

$$Z_{11} = f(Z_4, Z_{10}) = A_{0.01} = \gamma_R L_E$$

$$Z_{12} = f(\theta, \varphi, p) = \beta$$

$$\text{if } p \geq 1\% \text{ or } |\varphi| \geq 36^\circ, \beta = 0$$

$$\text{if } p < 1\% \text{ and } |\varphi| < 36^\circ \text{ and } \theta \geq 25^\circ,$$

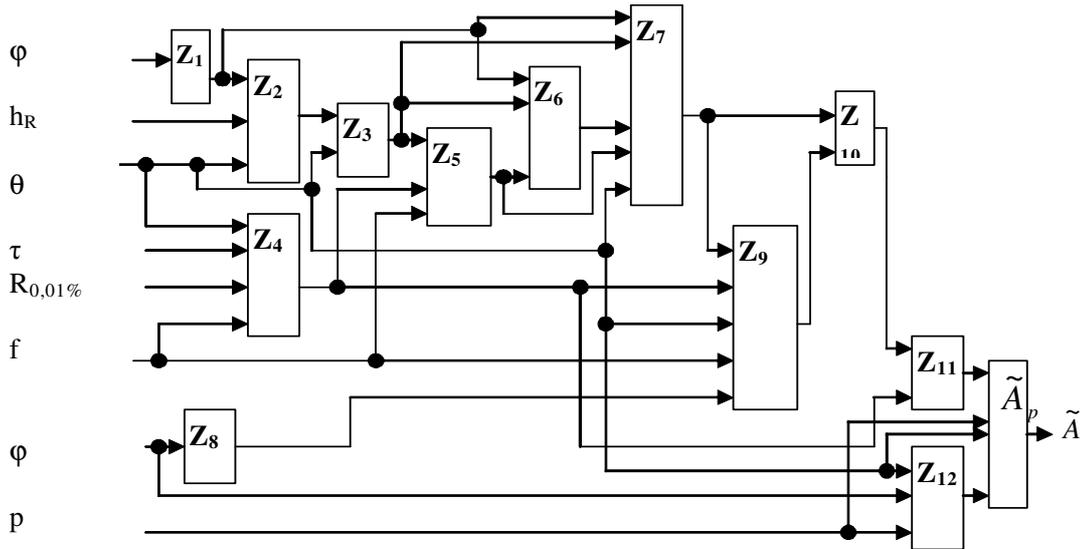
$$\beta = -0.005 (|\varphi| - 36)$$

$$\text{else, } \beta = -0.005 (|\varphi| - 36) + 1.8 - 4.25 \text{ sen } (\theta)$$

$$\tilde{A}_p = f(z_{11}, z_{12}, p, \theta) =$$

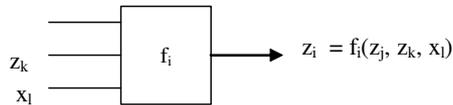
$$= A_{0.01} \left(\frac{p}{0.01} \right)^{-(0.655 + 0.033 \ln(p) - 0.045 \ln(A_{0.01}) - \beta(1-p) \text{sen}(\theta))}$$

Modelo UIT-R de Propagação de Sinal Terra - Satélite - Diagrama de Blocos



Composto por

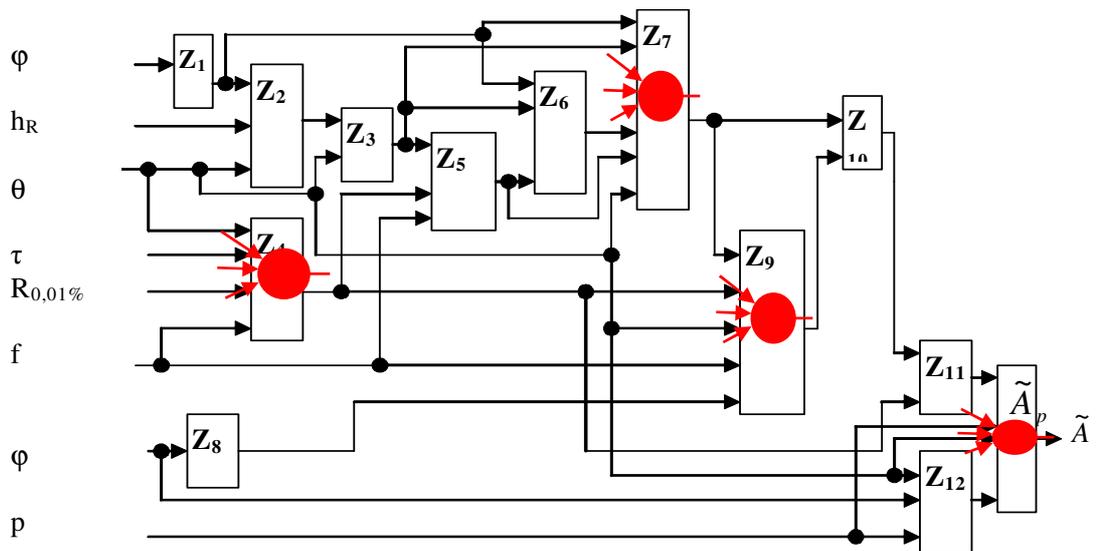
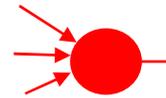
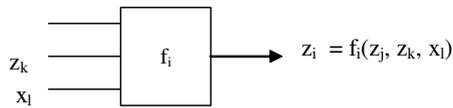
Blocos ou sub-modelos
(equações)



Bloco
ou sub-modelo



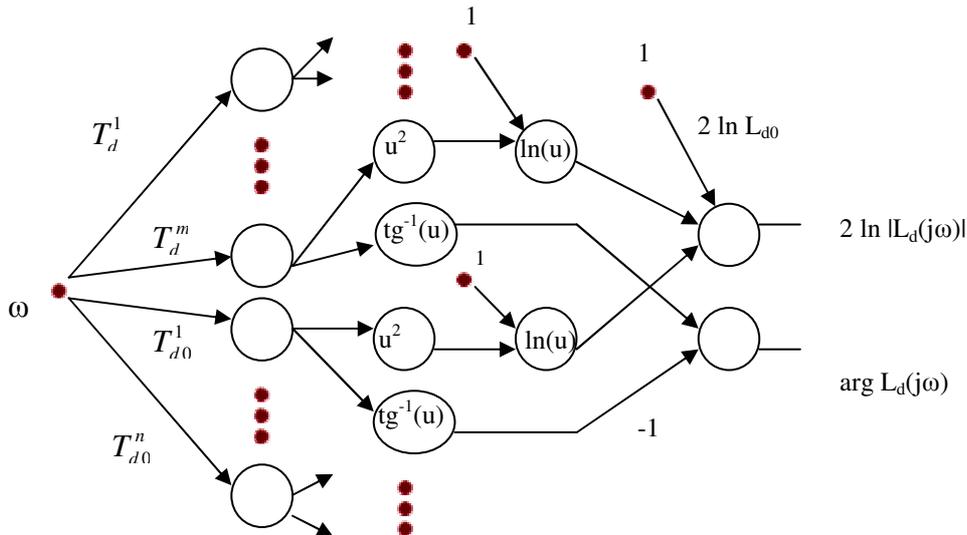
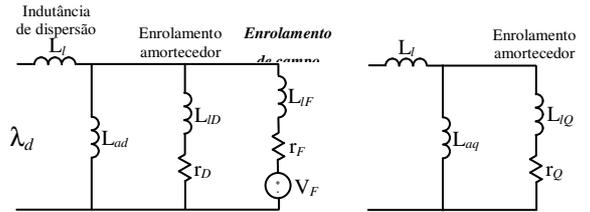
Neuronio



Ex 2: Circuito equivalente de um transformador

$$2 \ln |L_d(j\omega)| = 2 \ln L_{d0} + \sum_{j=1}^m \ln(1 + \omega^2 T_d^j{}^2) - \sum_{i=1}^n \ln(1 + \omega^2 T_{d0}^i{}^2)$$

$$\arg[L_d(j\omega)] = \sum_{j=1}^m \text{tg}^{-1}(\omega T_d^j) - \sum_{i=1}^n \text{tg}^{-1}(\omega T_{d0}^i)$$



II.1 -Ajuste dos parâmetros do modelo fenomenológico

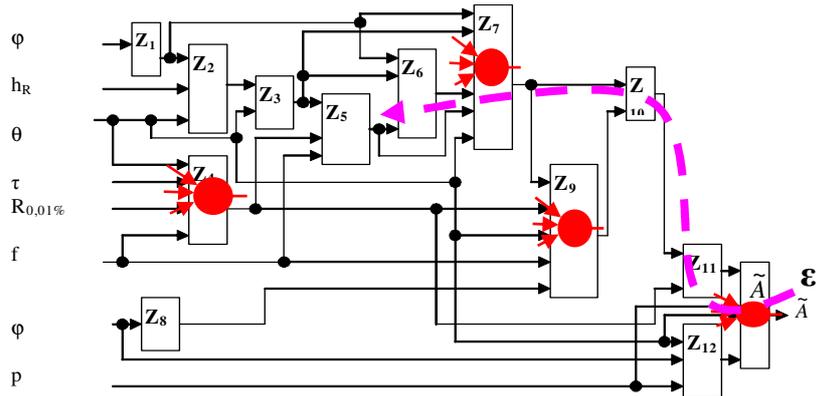
via retropropagação do erro

$$F_p = \epsilon_p^2 = (y - \tilde{y})^2$$

$$\tilde{y} = f_1(Z_5)$$

$$Z_5 = f_2(k)$$

$$\begin{aligned} \Delta k &= -\alpha \frac{\partial F_p}{\partial k} = -\alpha \frac{\partial F}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial k} \\ &= 2\alpha \epsilon \frac{\partial \tilde{y}}{\partial k} = 2\alpha \epsilon \frac{\partial \tilde{y}}{\partial Z_5} \frac{\partial Z_5}{\partial k} \end{aligned}$$



o erro ϵ na saída do modelo é retropropagado até cada bloco (ou parâmetro) de interesse

$$\varepsilon \frac{\partial \tilde{y}}{\partial Z_5} \quad \text{ou} \quad \varepsilon \frac{\partial \tilde{y}}{\partial k}$$

as derivadas podem ser calculadas analítica ou numericamente

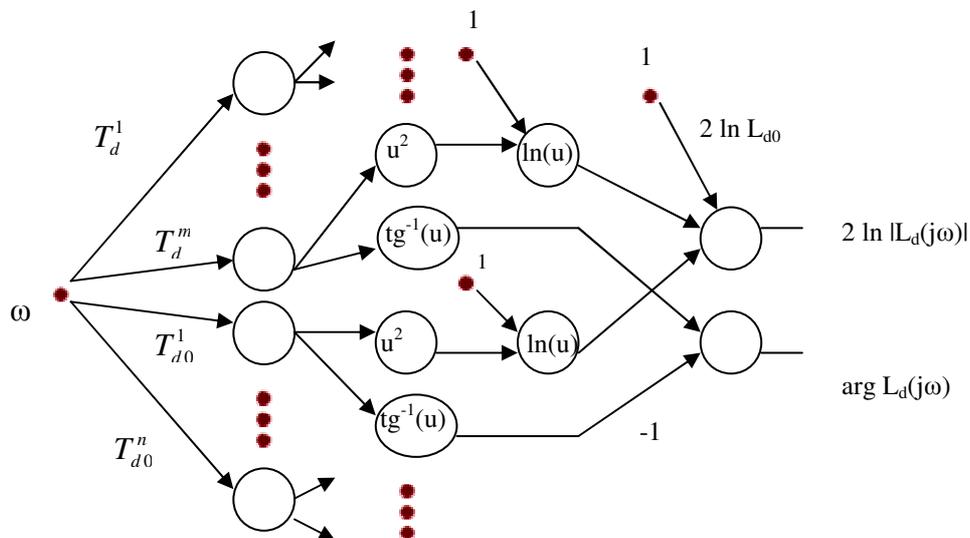
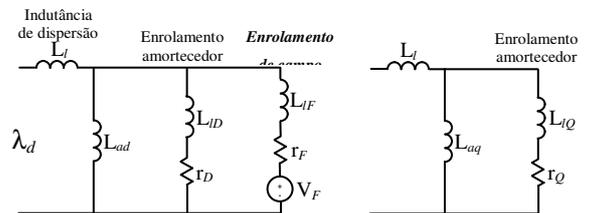
$$\frac{\partial \tilde{y}}{\partial k} \cong \frac{\Delta \tilde{y}}{\Delta k} \quad \frac{\partial \tilde{y}}{\partial Z_5} \cong \frac{\Delta \tilde{y}}{\Delta Z_5} \quad \frac{\partial Z_5}{\partial k} \cong \frac{\Delta Z_5}{\Delta k}$$

valores iniciais dos parâmetros podem ser críticos, para que a convergência se de para os parâmetros da planta.

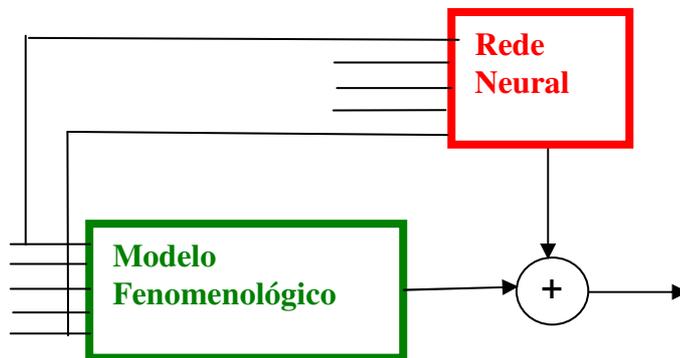
Aplicação: Circuito equivalente de um transformador

$$2 \ln |L_d(j\omega)| = 2 \ln L_{d0} + \sum_{j=1}^m \ln(1 + \omega^2 T_d^{j2}) - \sum_{i=1}^n \ln(1 + \omega^2 T_{d0}^{i2})$$

$$\arg[L_d(j\omega)] = \sum_{j=1}^m \text{tg}^{-1}(\omega T_d^j) - \sum_{i=1}^n \text{tg}^{-1}(\omega T_{d0}^i)$$

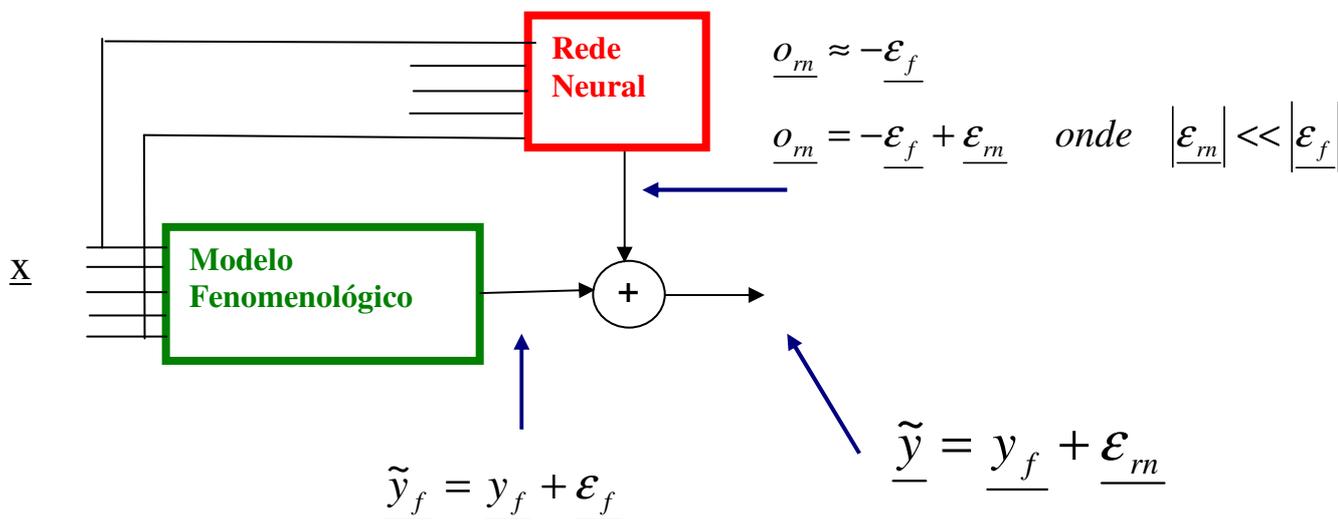


III - Modelos Híbridos Neural-Fenomenológicos



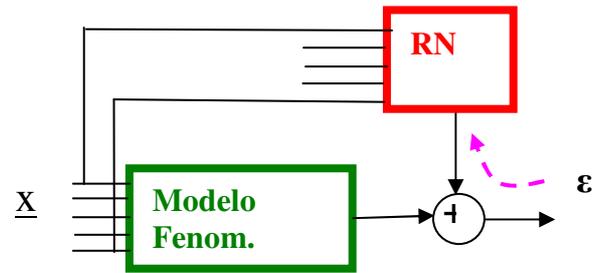
Informação do Processo através do Modelo Fenomenológico
 +
 Precisão numérica através da Rede Neural

Qual a função da RN, como atua ?



Corrigindo o erro do modelo fenomenológico

Como treinar a Rede Neural ?



1 - propagação do sinal para a frente: os dois blocos (modelo fenomenológico e rede neural) atuam

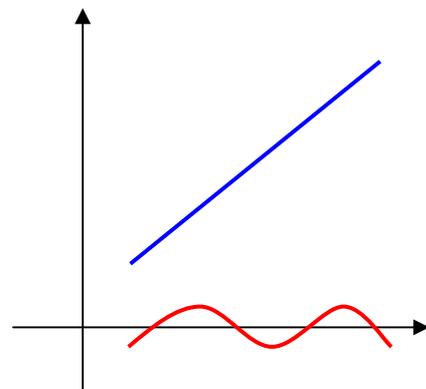
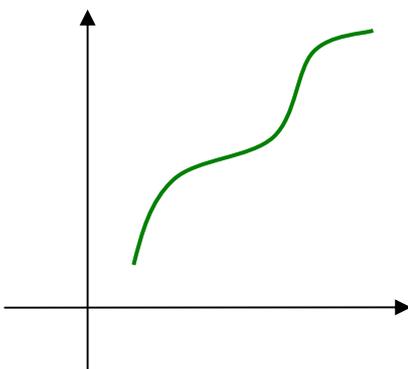
2 - erro na saída do sistema completo $\varepsilon = y - \tilde{y} = y - (\tilde{y}_f + o_{rn})$

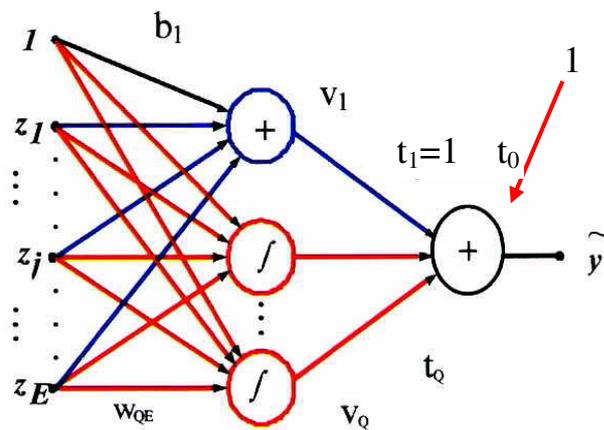
3 - erro retropropagado para a saída da RN = erro na saída do sistema completo, ε

III.1 - Um caso particular:

Sistemas com Não Linearidades Fracas

Modelo linear (fenomenológico ?) + complemento não linear



Modelo linear + complemento não linear

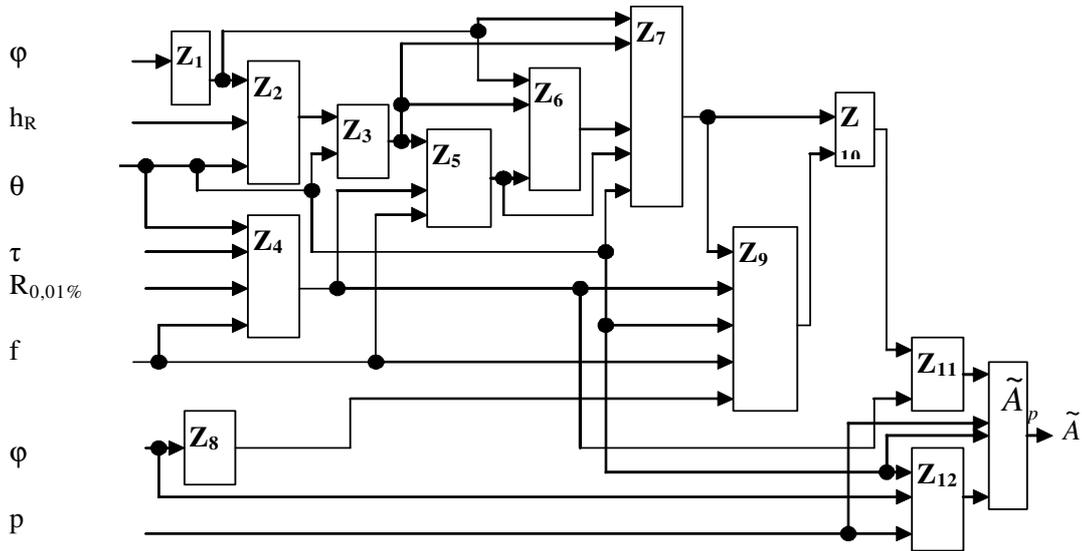
1 - Fazer as sinapses vermelhas nulas. Treinar apenas as sinapses azuis. $t_1=1$

2- Congelar as sinapses azuis. Treinar as sinapses vermelhas (com as azuis atuando).

IV Sub-modelos Híbridos Neural-Fenomenológicos

Modelos Fenomenológicos tratados como uma Rede Neural

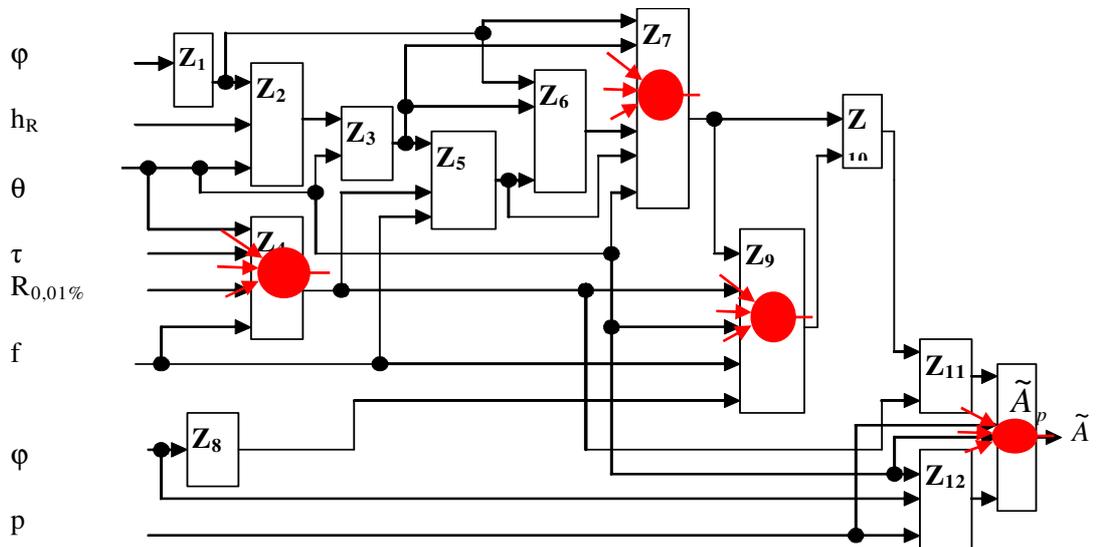
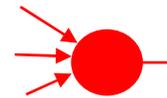
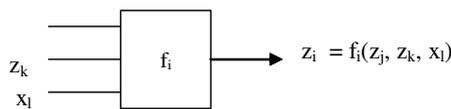
Modelo Fenomenológico → Sistema de equações → Diagrama de Blocos → Rede Neural



Modelo é composto por blocos ou sub-modelos



ou “neuronios”



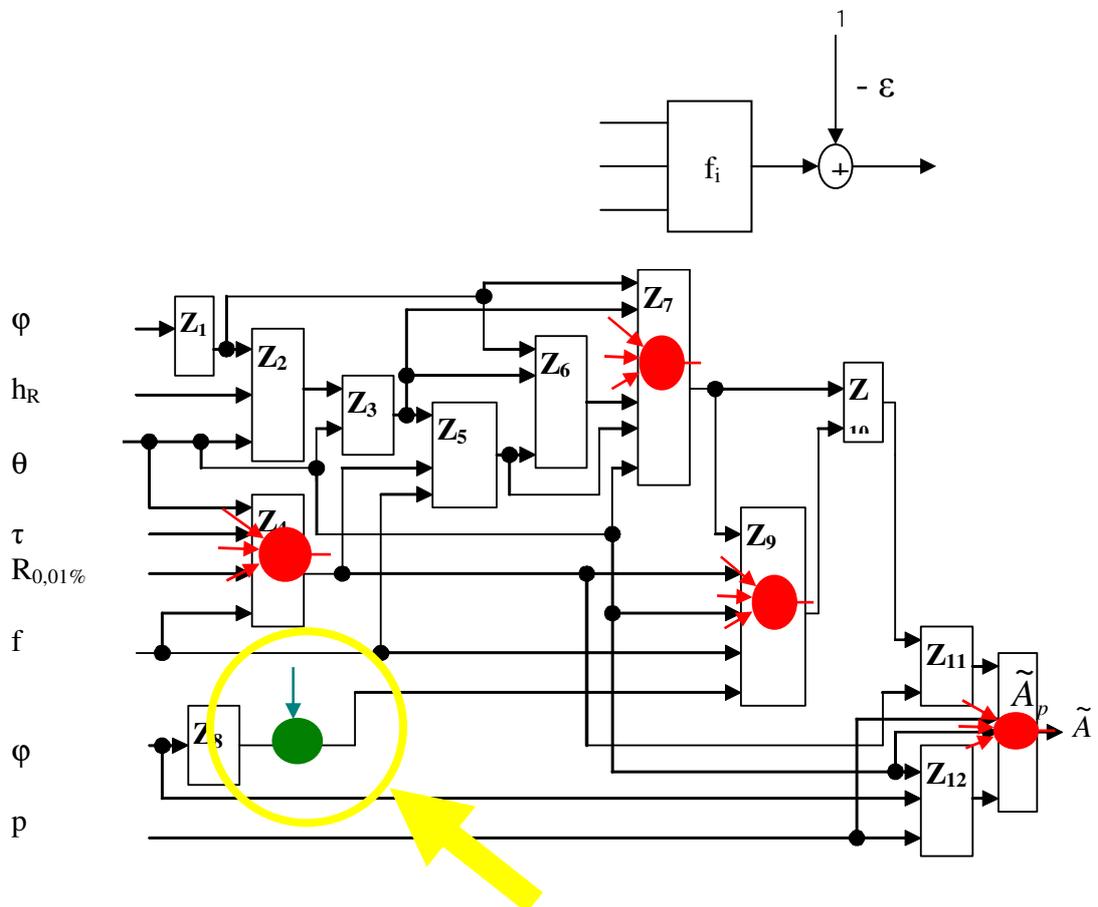
No modelo fenomenológico

o erro varia de bloco para bloco.

Como identificar os blocos com os maiores erros ?

Bloco com erros

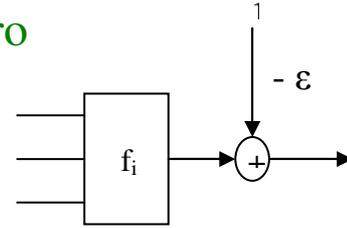
sinapses “mágicas” para correção do erro



Critério para erro do bloco:

$$\sigma_i = \sigma [\delta_i(p)]$$

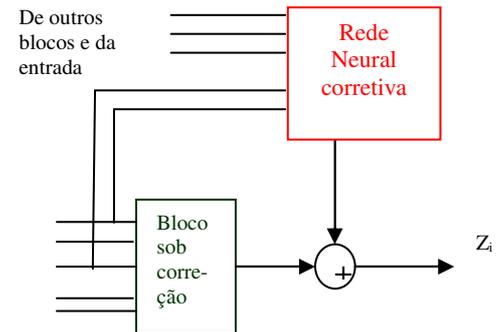
Bloco com uma
sinapse “mágica” para correção do erro



Bloco com uma
rede neural para correção do erro.

Que variáveis usar
como entrada da RN ?

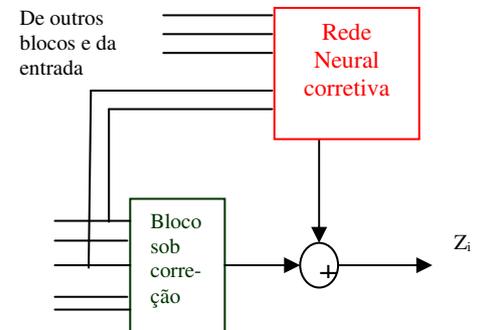
– correlatas com o erro do bloco



O erro do bloco vem de:

Função de transferência do bloco incorreta

- o erro do bloco está correlacionado com algumas entradas do próprio bloco
- use estas entradas na RN corretiva



Falta informação na entrada do bloco

- o erro do bloco não está correlacionado com as entradas do bloco
- teste outras variáveis disponíveis quanto à correlação com o erro do bloco. Se a correlação existe, use estas variáveis na entrada da rede neural de correção do bloco

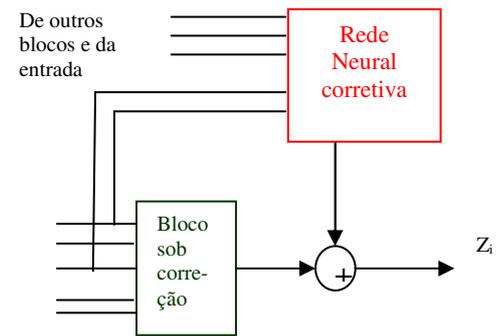
Entradas da RN corretiva do bloco

- entradas do próprio bloco

– correlatas com o erro do bloco

- outras variáveis disponíveis quando do cálculo do bloco

– correlatas com o erro do bloco

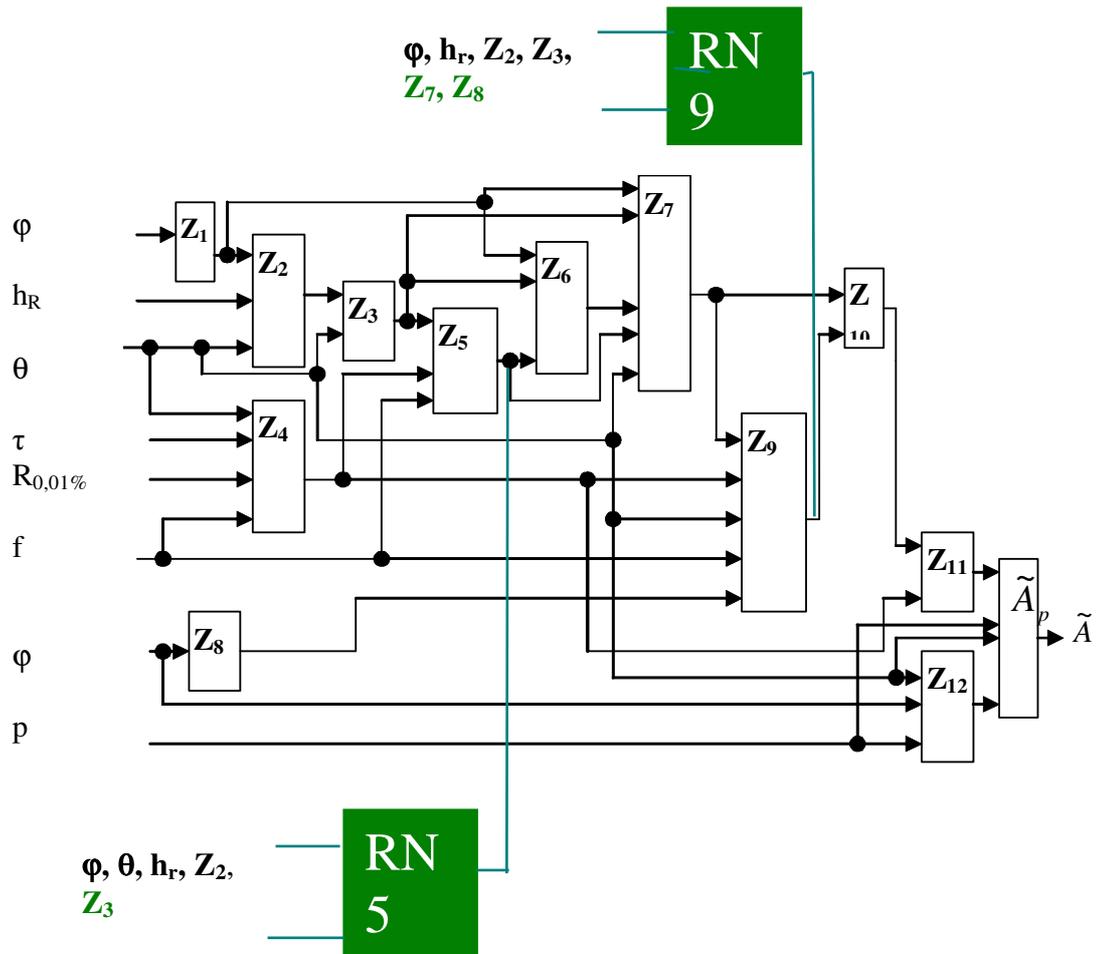


ERROS DOS SUBMODELOS

DESVIO PADRÃO DO ERRO RETROPROPAGADO (δ_i) PARA CADA SUB-MODELO E PARA DIFERENTES PERCENTUAIS DE TEMPO

	1	2	3	4	5	6	7	8	9	10	11	12	\tilde{A}
0.001	0.21	0.09	0.10	0.78	1.25	0.00	0.15	0.02	1.16	0.19	0.15	1.17	0.36
0.01	0.13	0.06	0.08	0.49	0.85	0.00	0.11	0.00	0.78	0.14	0.11	0.00	0.29
0.05	0.19	0.09	0.12	0.48	1.72	0.00	0.18	0.00	1.37	0.15	0.15	0.76	0.35

A tabela acima evidencia que **os sub-modelos 5 and 9 são os críticos.**

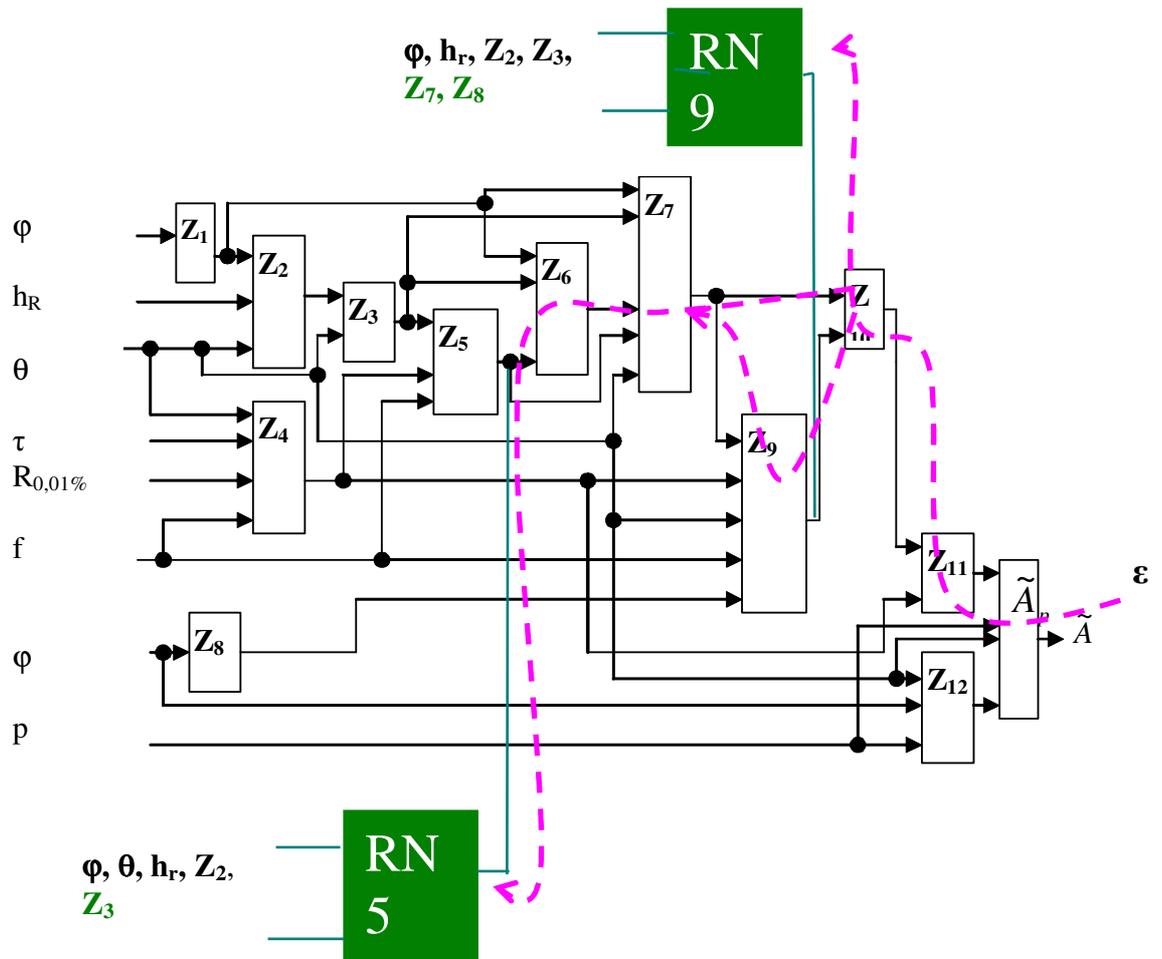


Treinamento das RN's

1 - como no caso do modelo híbrido, na propagação do sinal para a frente todos os blocos (sub-modelos fenomenológicos e redes neurais locais) atuam

2 - o erro a ser minimizado é o erro na saída do sistema completo ε

3 - como no caso de ajuste dos parâmetros do modelo fenomenológico, o erro ε é retropropagado para a saída de cada RN, e daí para cada sinapse da mesma. As derivadas podem ser calculadas analítica ou numericamente



RESULTADOS

ERRO RELATIVO RMS (%) PARA TODOS OS PARES.

Modelo	Fenom. UIT-R	Híbrido UIT-R-Neural	Neural
Percent. Tempo			
0,001%	38	21	17
0,01%	29	21	20
0,05%	35	23	20

Redes Neurais na Modelagem de Sistemas Dinâmicos (e Séries Temporais)



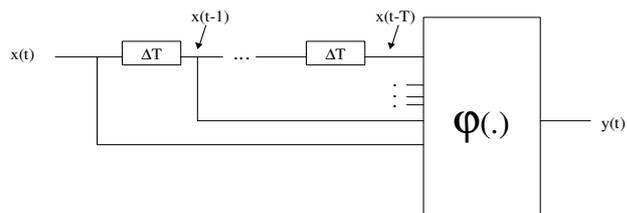
Sistema SISO

Modelos Não Recursivos

Modelo MA (moving average)

Filtro FIR

$$y(t) = \sum_{i=0}^T b_i x(t-i)$$



Time Delay Neural Network

Modelo NMA

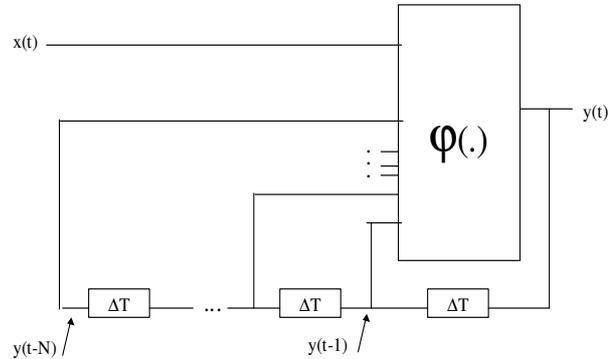
$$y(t) = \varphi[x(t), x(t-1), \dots, x(t-N)]$$

Modelos Recursivos

Modelo AR (auto regressivo)

Filtro IIR

$$y(t) = b_0 x(t) + \sum_{i=1}^N a_i y(t-i)$$



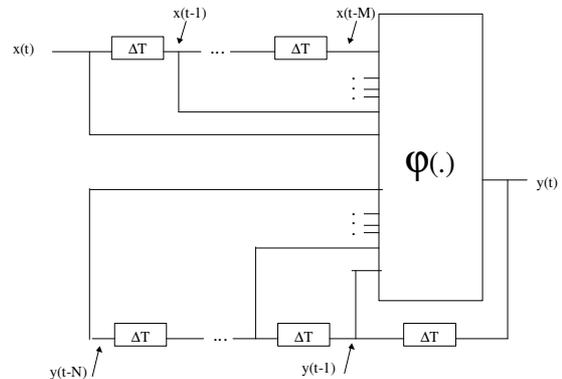
Modelo NAR

$$y(t) = \varphi[x(t), y(t-1), \dots, y(t-N)]$$

Modelos Recursivos

Modelo ARMA

$$\tilde{y}(t) = \sum_{i=0}^M b_i x(t-i) + \sum_{i=1}^N a_i \tilde{y}(t-i)$$



Modelo NARMA

$$y(t) = \varphi[x(t), x(t-1), \dots, x(t-M), y(t-1), \dots, y(t-N)]$$

Dimensão dos Modelos

Planta:



linearizada nos pontos de operação

“singularidades”,
 “frequências naturais”,
 “constantes de tempo”

Dimensionando o MA (NMA)

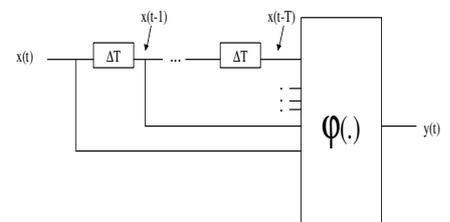
$$T \cdot \Delta T \geq 4 \text{ Max}(\tau_i) \qquad T \cdot \Delta T \geq \frac{2}{\pi f_{\min}}$$

$f_{\min} = 1 / (2\pi \text{Max}(\tau_i))$ é a mais baixa frequência natural do sistema, em Hz.

$$f_s = \frac{1}{\Delta T} \geq 2 f_{\max}$$

f_{\max} = maior frequência operada pelo sistema

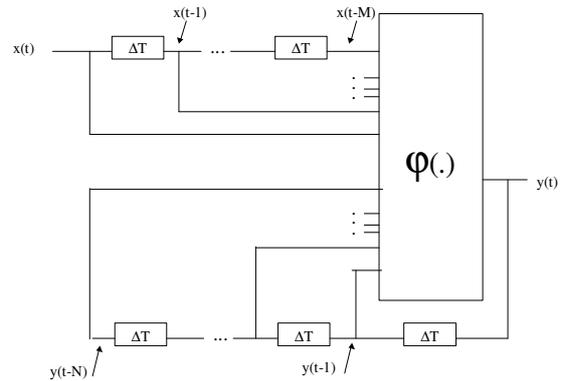
$$T \cong \frac{4}{\pi} \frac{f_{\max}}{f_{\min}}$$



Dimensionando o ARMA (NARMA)

N – ordem prevista (2,3)

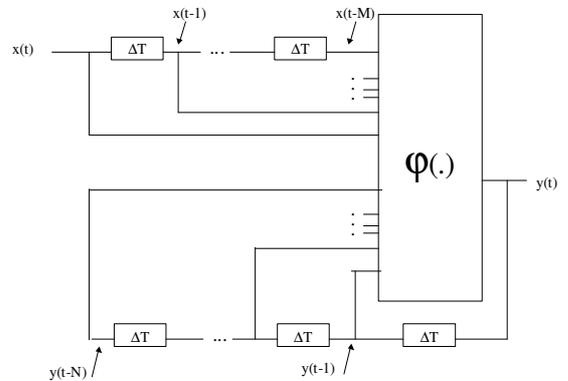
$$M \leq N$$



NARMA muito menor que o MA mas pode instabilizar !

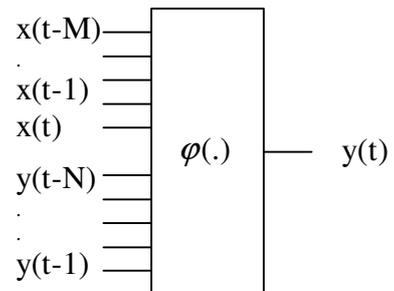
Como treinar a RN ?

Rede em Operação
(em modo paralelo)



Rede em Treinamento

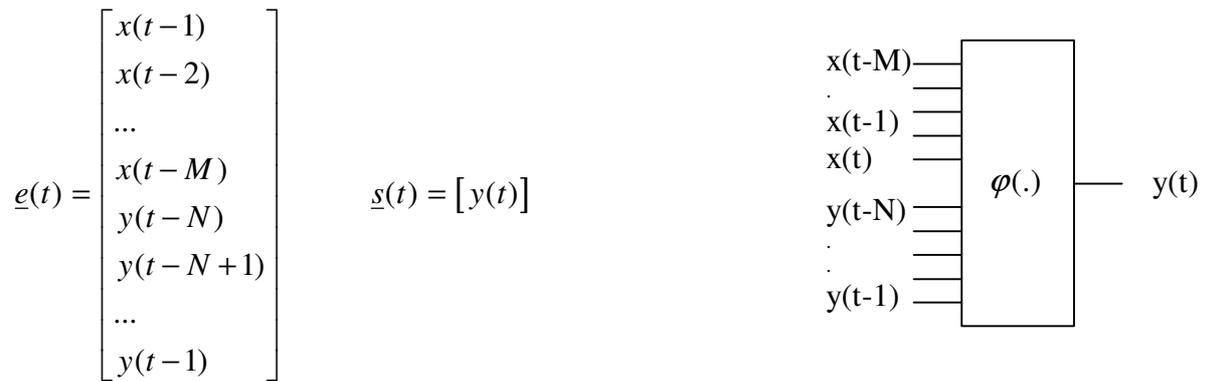
**Não é dinâmico, a rede
é treinada em separado !!**



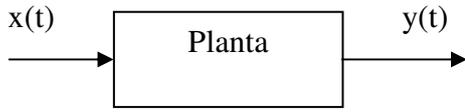
Treinamento da RN

Não é dinâmico, a rede é treinada em separado !!

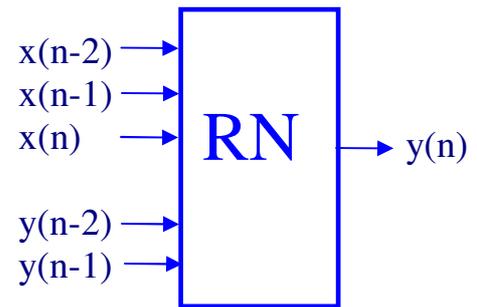
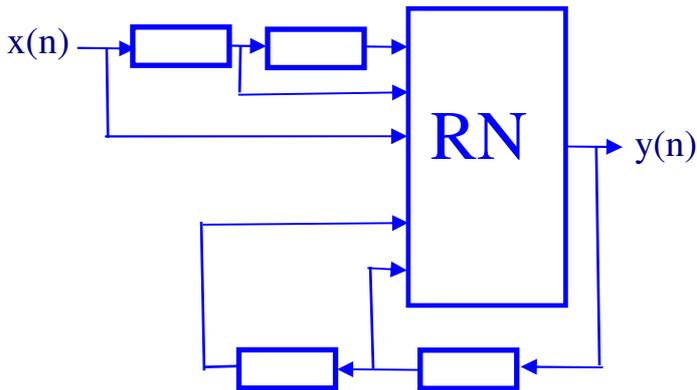
Par entrada-saída ($\underline{e}(t)$, $\underline{s}(t)$) correspondente ao instante t:



Exemplo:



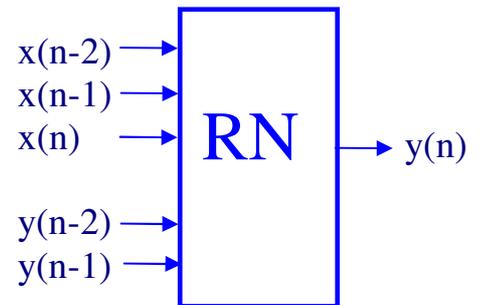
$M=3$
 $N=2$



Construção dos pares entrada-saída para o treinamento:

t, n	x(n)	y(n)
...
52	x(52)	y(52)
53	x(53)	y(53)
54	x(54)	y(54)
55	x(55)	y(55)
56	x(56)	y(56)
57	x(57)	y(57)
...

$t = n = 56$



$$\left[\begin{array}{l} x(54) \\ x(55) \\ e(56) = x(56) \\ y(54) \\ y(55) \end{array} \right]$$

$$s(56) = y(56)$$

Cuidados usuais a tomar com RNs

- 1 - Escolha das Variáveis de entrada
- 2 – Escalamento das Variáveis
- 3 - Pares entrada – saída
- 4 - Dimensionando a rede
- 5 – Inicializando a Rede
- 6 – Treinamento: Superfícies de Erro
- 7 – Overtraining (overfitting, sobre-treinamento)
- 8 – Critica durante o treinamento
- 9 - Validação Cruzada
- 10 – Critica pos treinamento
- 11 - Outros tipos de erro
- 12 – Outros Processos de Treinamento

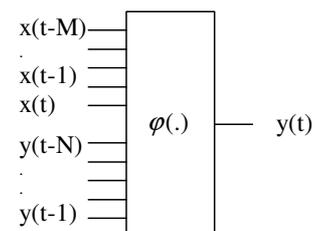
Cuidados adicionais a tomar no caso de sistemas dinâmicos

- 1 - Escolha das Variáveis de entrada da rede neural

Quantas (quais) entradas / **saídas** atrasadas tomar ?

Entradas – existem atrasos puros ?

Saídas realimentadas - **ordem do sistema**



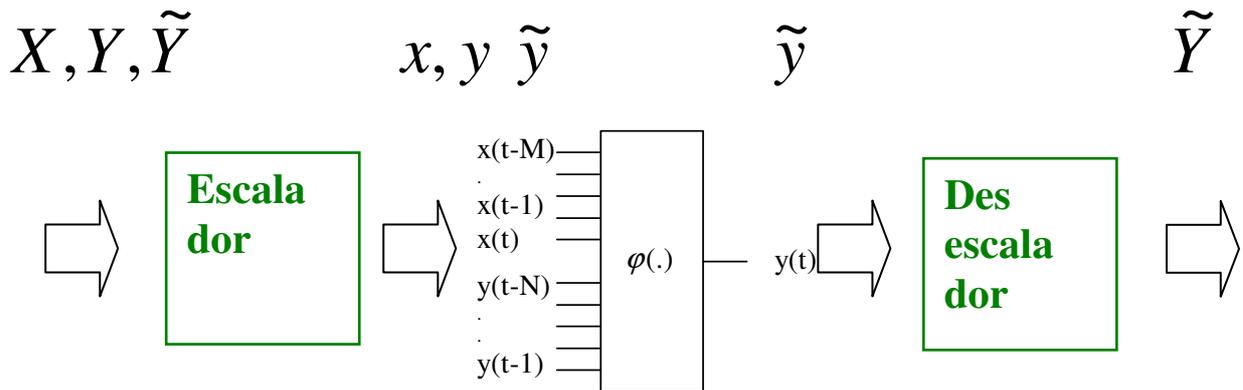
poucas: não representa
demasia: pode instabilizar

Tentativa e erro – ordem do sistema

Possível solução : Poda de saídas realimentadas
na construção da RN automática / manual

2 – Escalamento das Variáveis

Rede **em operação** trabalha imersa em um escalador / desescalador



O escalador e o desescalador podem ser embutidos na primeira e última camada da rede treinada.

3 - Pares entrada – saída:

Mapeamento unívoco

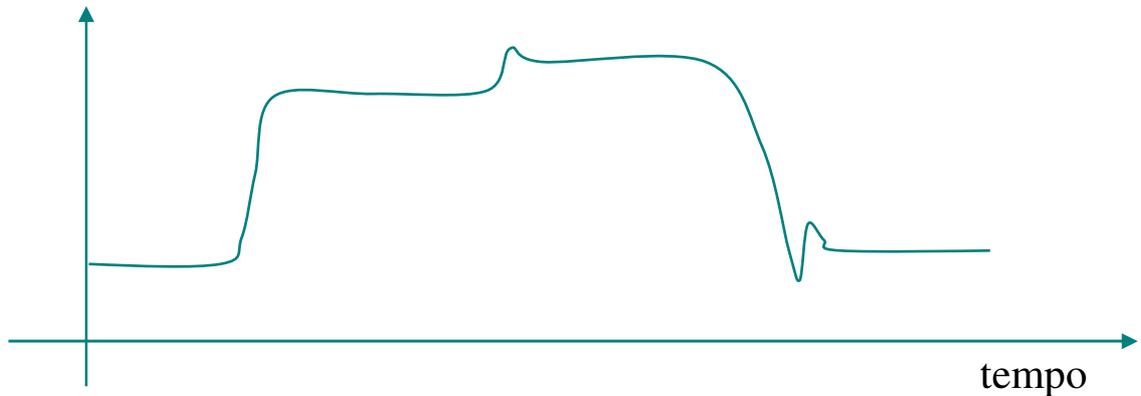
$$x_1 \gg y_1 \quad \text{e} \quad x_2 \gg y_2$$

$$\text{se } x_2 = x_1 \quad \text{então } y_2 = y_1$$

o modelo (inverso) é unívoco ?

Efeito de população local reduzida:

transição vs regime permanente



Efeito ? alto erro na região de baixa população

Correção ? replicar população

4 – Obtenção dos dados da planta

Pontos de operação para a modelagem:

Grande excursões (amplitudes) - explorar as não linearidades

Grande faixa de frequências – explorar todas as singularidades

entrada Ideal: ruído branco de grande amplitude

Pontos de operação reais:

Grandes excursões – busca do rendimento máximo

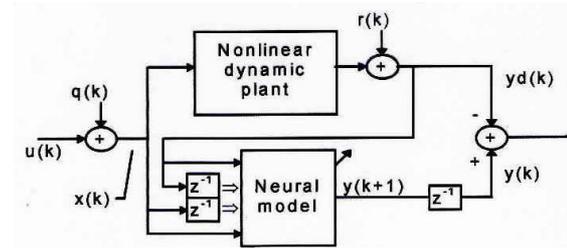
Operação suave – baixas frequências

E a caracterização das altas frequências ?

Aplicar um sinal piloto na entrada:

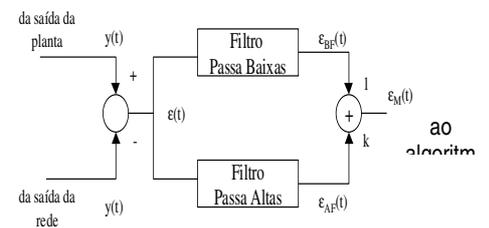
Altas frequências, baixa amplitude

$q(t)$ = onda quadrada de período aleatório

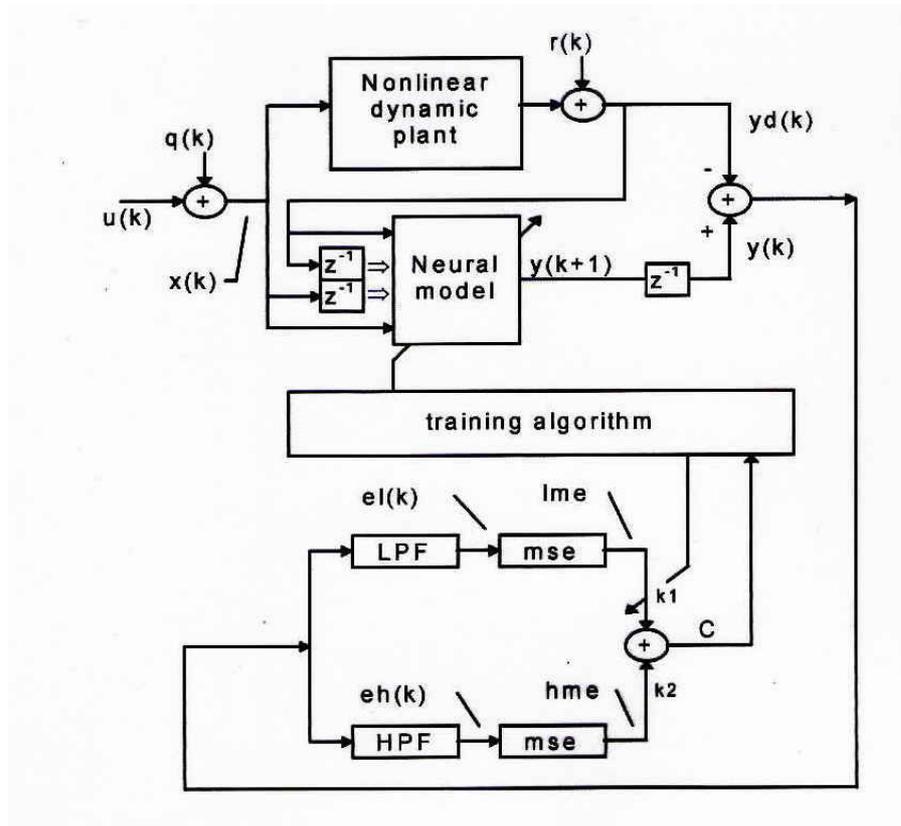


Enfatizar o erro em altas frequências na saída:

$$\epsilon_M(t) = \epsilon_{LF}(t) + k \epsilon_{HF}(t)$$



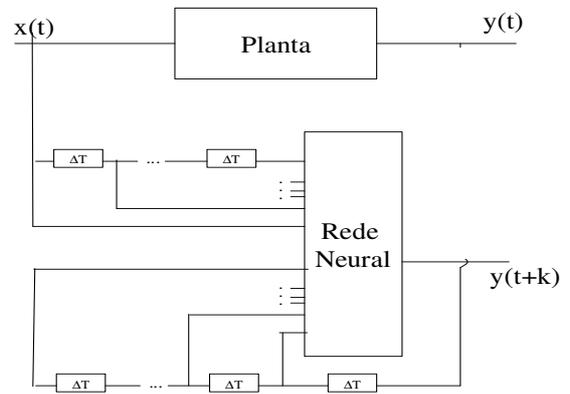
Treinamento:



Operação da Rede

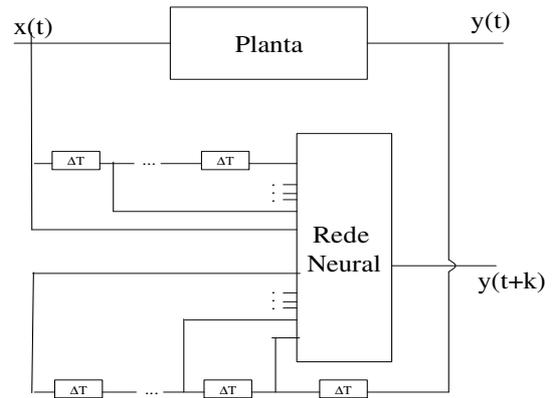
Operação em Paralelo

Independente da planta,
 possível instabilidade

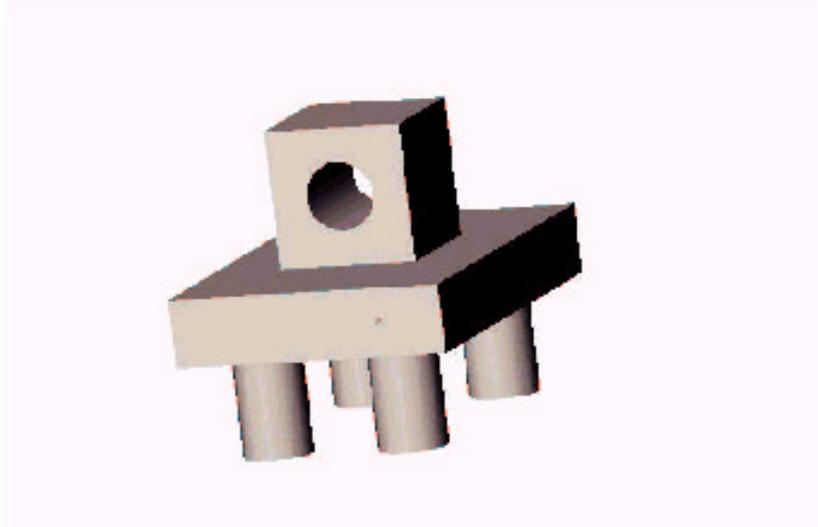


Operação em Série – Paralelo

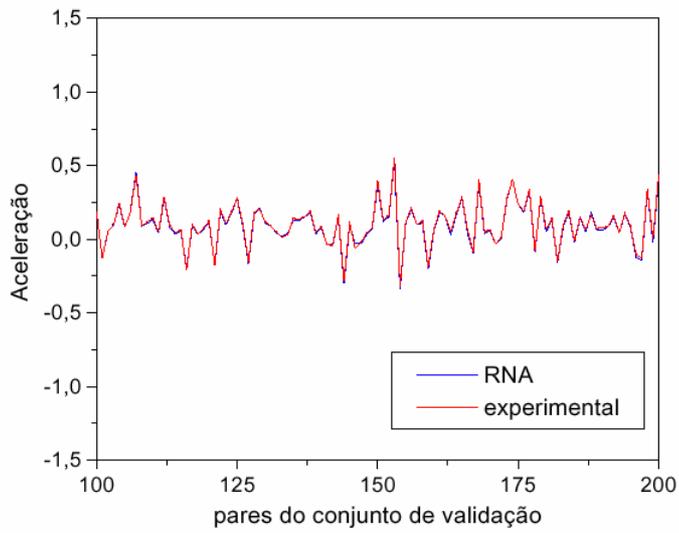
Depende da planta,
 estabilidade garantida
 para que serve ? **Preditores** $y(t+k)$



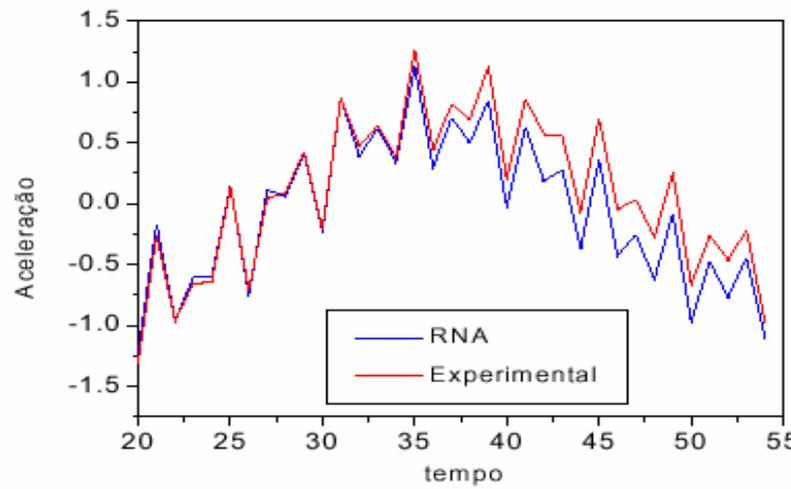
Exemplo 1:



Operação em serie - paralelo



Operação em paralelo



Exemplo 2:

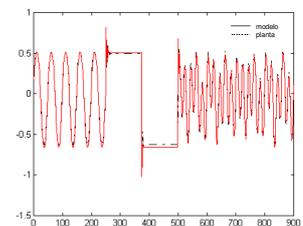
$$y(t) = \frac{y(t-1) y(t-2) y(t-3) x(t-2) [y(t-3) - 1] + x(t-1)}{1 + y(t-3)^2 + y(t-2)^3}$$

Entrada para treinamento e teste

$$x(t) = \text{sen}(.1t) + .02 \text{sq_rnd}(t)$$

Entrada para validação

$$x(k) = \begin{cases} \text{sen}(\pi k / 25) & , k < 250 \\ +1 & , 250 \leq k < 375 \\ -1 & , 375 \leq k < 500 \\ 0,3\text{sen}(\pi k / 25) + 0,1\text{sen}(\pi k / 32) + 0,6\text{sen}(\pi k / 10) & , 500 \leq k < 900 \end{cases}$$

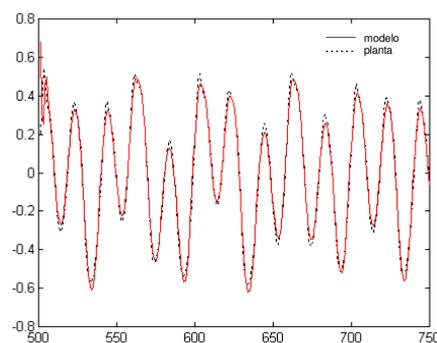
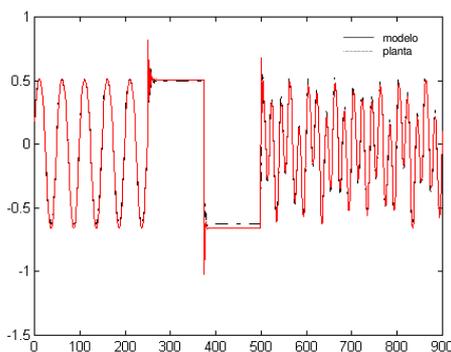


Rede Neural: 7 neurônios na camada intermediária

Ordem do modelo (poda):

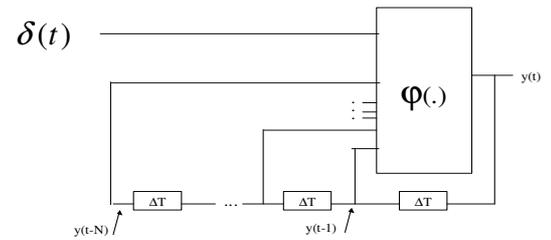
	Inicial	Final	Correta
N	5	3	3
M	4	2	2

Resultados



Uso em Series Temporais (modelo AR)

$S(t) =$ Tendencia +
 Sazonalidade +
 Ciclos senoidais +
 Outros artefatos determinísticos +
 Componentes não lineares +
 Ruído não correlato

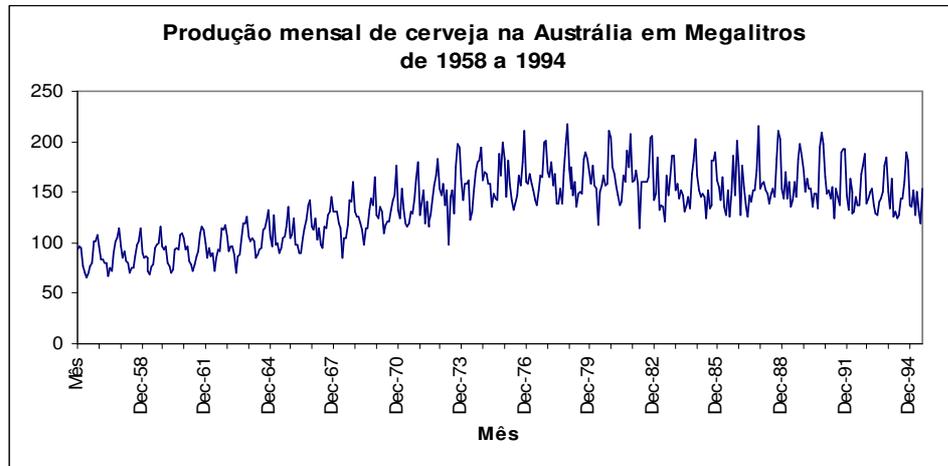


Redes Neurais
predizem este
termo !

Trabalhar sobre a serie residual !

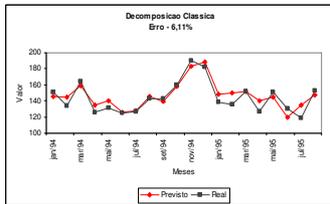
Series Temporais

Exemplo:

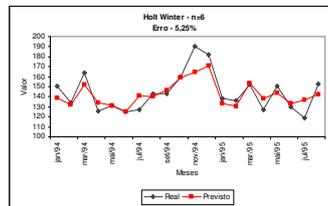


Resultados (período de teste):

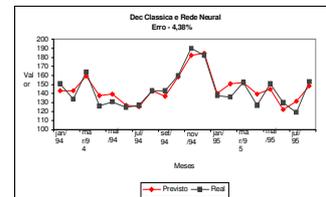
Convencional



Holt-Winters



RNs



	Decomp.	Holt Winter	RN
1º Ano	4.71%	4.81%	3.68%
20 Meses	6.11%	5.25%	4.35%