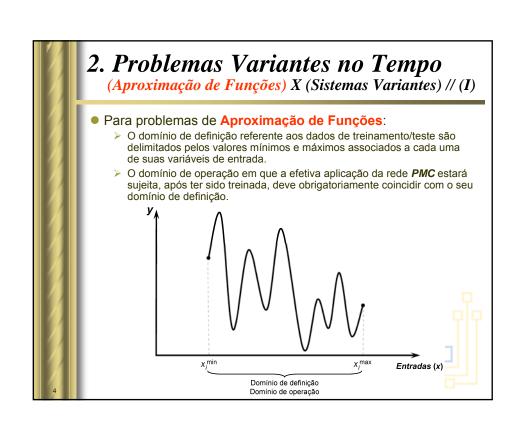


1. Aplicabilidades de PMC (Resumo) Principais classes de problemas

- Redes PMC podem ser consideradas as mais utilizadas na solução de problemas advindos das mais variadas áreas de conhecimento.
- Redes PMC são as mais amplamente empregadas em diferentes temáticas envolvendo as engenharias como um todo, em especial a Engenharia Elétrica.
- Existe ainda aplicações de redes PMC em medicina, biologia, química, física, economia, geologia, ecologia e psicologia.
- Considerando esses leques de aplicabilidades em que as redes PMC são passíveis de serem utilizadas, destacam-se três classes de problemas que acabam concentrando grande parte de suas aplicações, isto é:
 - > Problemas envolvendo aproximação funcional (Fim da Aula 05).
 - Problemas envolvendo classificação de padrões (Aula 06).
 - Problemas envolvendo sistemas variantes no tempo (Esta Aula).



2. Problemas Variantes no Tempo Aspectos de definição (I) Como o próprio nome sugere, sistemas dinâmicos ou sistemas variantes no tempo são aqueles cujos comportamentos são dependentes do tempo. A resposta atual depende dos valores das respostas anteriores. Como exemplos de aplicação, têm-se os seguintes: Previsão de consumo de energia para os próximos meses. Previsão de valores futuros para ações do mercado financeiro frente a um horizonte semanal. Previsão de temperatura, precipitação atmosférica, etc. Em contraste aos problemas de Aproximação de Funções ou Classificação de Padrões (considerados estáticos), tem-se: As saídas dos sistemas denominados dinâmicos, assumindo um instante de tempo qualquer, dependem de seus valores anteriores de saída e de entrada.



2. Problemas Variantes no Tempo (Aproximação de Funções) X (Sistemas Variantes) // (II) Para problemas de Sistemas Variantes no Tempo: Aqui, ambos os domínios são regidos pelo tempo, sendo que o domínio de operação se inicia após o seu domínio de definição. Como a saída atual depende das saídas/entradas anteriores, utiliza-se então dados de treinamento/teste para ajustar seus parâmetros internos. Em seguida, a rede estará apta para estimar valores futuros que estarão pertencendo ao seu domínio de operação

2. Problemas Variantes no Tempo Topologias de PMC para sistemas variantes no tempo • Em se tratando de aplicação de redes PMC no mapeamento de problemas envolvendo sistemas variantes no tempo, duas configurações topológicas podem ser utilizadas: • Configuração TDNN → rede PMC com entradas atrasadas no tempo (Time Delay Neural Network - TDNN). • Configuração RECORRENTE→ rede PMC com saídas recorrentes às entradas (Recurrent Perceptron).

3. Configuração TDNN

Aspectos de estrutura de modelo

- Redes PMC em configuração TDNN (entradas atrasadas no tempo) são enquadradas dentro da arquitetura feedforward de camadas múltiplas:
 - Inexiste qualquer realimentação das saídas de neurônios de camadas posteriores em direção aos neurônios da primeira camada.
- A previsão de valores futuros a partir do instante t, associados ao comportamento do processo, é computada em função do conhecimento de seus valores temporariamente anteriores, isto é:

$$x(t) = f(x(t-1), x(t-2), ..., x(t-n_p))$$

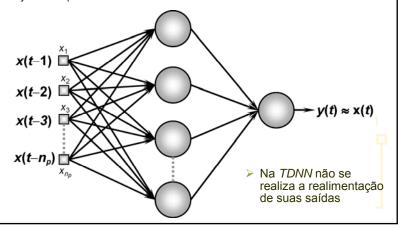
- \triangleright Onde n_p é a ordem do estimador, ou seja, a quantidade de medidas (amostras) passadas que serão necessárias para a estimação do valor x(t).
- O modelo representado pela expressão acima é também conhecido como Auto-Regressivo (AR).

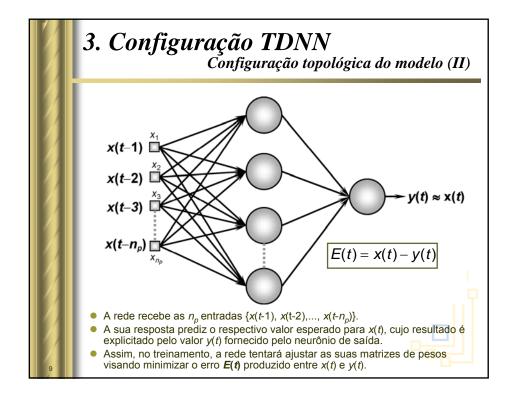
7

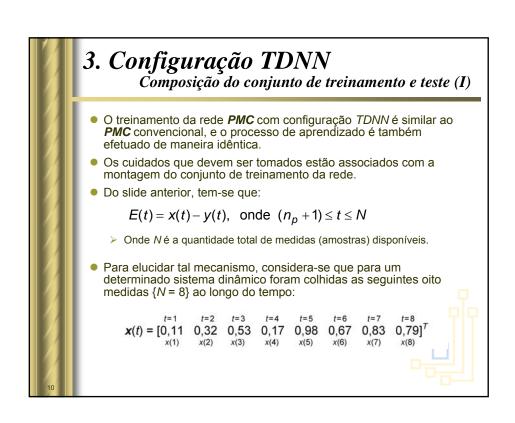
3. Configuração TDNN

Configuração topológica do modelo (I)

- De fato, a configuração TDNN estará aqui introduzindo linha de atrasos de tempo somente na camada de entrada.
- Esta linha funciona como uma memória, garantindo que amostras anteriores, que refletem o comportamento temporal do processo, sejam sempre inseridas dentro da rede.







3. Configuração TDNN Composição do conjunto de treinamento e teste (II)

$$\boldsymbol{x}(t) = \begin{bmatrix} t=1 & t=2 & t=3 & t=4 & t=5 & t=6 & t=7 & t=8 \\ 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 \end{bmatrix}^T$$

 Como exemplo, assume-se que o processo possa ser mapeado com uma ordem de predição igual a três $\{n_p=3\}$. Então, tem-se:

$$E(t) = x(t) - y(t)$$
, onde $(n_p + 1) \le t \le N$

Por intermédio da expressão acima, ter-se-á então um conjunto de treinamento composto por um total de 5 amostras, com parâmetro *t* variando de $4 \le t \le 8$.

	relag	ão er	ntrada	s/saídas			con	junto	de trei	namento
	X ₁	X2	Х3	saída desejada			<i>X</i> ₁	X ₂	<i>X</i> ₃	d
t = 4	x(3)	x(2)	x(1)	x(4)	4 ≤ <i>t</i> ≤ 8	x ⁽¹⁾	0,53	0,32	0,11	$d^{(1)} = 0,17$
t = 5	x(4)	x(3)	x(2)	x(5)	\Leftrightarrow	x ⁽²⁾	0,17	0,53	0,32	$d^{(2)} = 0.98$
t = 6	x(5)	x(4)	x(3)	x(6)	(ordem 3)	x ⁽³⁾	0,98	0,17	0,53	$d^{(3)} = 0,67$
t = 7	x(6)	x(5)	x(4)	x(7)	$n_p = 3$	x ⁽⁴⁾	0,67	0,98	0,17	$d^{(4)} = 0.83$
t = 8	x(7)	x(6)	x(5)	x(8)]	x ⁽⁵⁾	0,83	0,67	0,98	$d^{(5)} = 0,79$

onde o valor $\{x_0 = -1\}$, associado ao limiar do neurônio, deverá ser considerado em todos eles.

3. Configuração TDNN

Composição do conjunto de treinamento e teste (III)

Na realidade, procede-se no vetor x(t), uma operação de janela deslizante de largura n_p , movimentando-a de uma unidade para a direita em cada iteração de tempo.

$$\mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 &]^T \\ \mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 &$$

Após o treinamento da rede, basta realizar a inserção de amostras anteriores da série a fim de se executar a predição de seus valores futuros (posteriores).

6

3. Configuração TDNN

Composição do conjunto de treinamento e teste (IV)

- Considerando-se agora o seu domínio de operação, a rede já treinada pode ser usada para estimação de valores futuros.
- Para o exemplo anterior, considerando agora {t ≥ 9}, tal procedimento é realizado da seguinte maneira:

predição de	e valores futuros
-------------	-------------------

	X ₁	X2	X3	saída estimada
t = 9	x(8)	x(7)	x(6)	$x(9) \approx y(9)$
t = 10	x(9)	x(8)	x(7)	$x(10) \approx y(10)$
t = 11	x(10)	x(9)	x(8)	$x(11) \approx y(11)$
()	()	()	()	()

- ➤ Para obter x(9) → Entradas $\{x_1=x(8), x_2=x(7), x_3=x(6)\}$
- ➤ Para obter x(10) → Entradas $\{x_1=x(9), x_2=x(8), x_3=x(7)\}$
- ➤ Para obter x(11) → Entradas $\{x_1 = x(10), x_2 = x(9), x_3 = x(8)\}$
- · (...)

$$\mathbf{x}(t) = \begin{bmatrix} t = 1 & t = 2 & t = 3 & t = 4 & t = 5 & t = 6 & t = 7 & t = 8 \\ 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 \end{bmatrix}^{T}$$

$$\mathbf{x}(t) = \begin{bmatrix} 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 \end{bmatrix}^{T}$$

13

3. Configuração TDNN

Composição do conjunto de treinamento e teste (V)

- Assim, para o exemplo anterior, conclui-se que a rede sempre realiza uma predição de um passo à frente, calculando seqüencialmente o seu valor atual ou futuro a partir de seus 3 últimos valores.
- Contudo, há situações em que a ordem de predição deve ser incrementada a fim de assegurar uma maior precisão na estimação do comportamento futuro do processo.
- Como exemplo, para a ação de se utilizar uma ordem de predição igual a 4 $\{n_p = 4\}$, o próximo valor obtido pela rede estaria em função dos 4 últimos valores, sendo que sua topologia seria agora composta de 4 entradas.
- Generalizando, tem-se:
 - > Para $n_p = 4 \rightarrow$ Utiliza os 4 últimos valores para predizer o valor atual. > Para $n_p^0 = 5 \rightarrow$ Utiliza os 5 últimos valores para predizer o valor atual.
 - (...)
- Entretanto, quanto maior a ordem de predição assumida, menor será também a quantidade de amostras para o processo de treinamento.

$$\mathbf{x}(t) = \begin{bmatrix} t=1 & t=2 & t=3 & t=4 & t=5 & t=6 & t=7 & t=8 \\ 0.11 & 0.32 & 0.53 & 0.17 & 0.98 & 0.67 & 0.83 & 0.79 \end{bmatrix}^T$$

E(t) = x(t) - y(t), onde $(n_p + 1) \le t \le N$

4. Configuração Recorrente

Aspectos de estrutura de modelo

- Diferentemente da TDNN, a arquitetura com saídas recorrentes às entradas possibilitam a recuperação de respostas passadas a partir da realimentação de sinais produzidos em instantes anteriores.
- Pode-se dizer que a configuração recorrente possui memória, sendo capazes de "relembrar" saídas passadas a fim de produzir a resposta atual ou futura.
- Tais redes pertencem à classe de arquiteturas recorrentes ou realimentadas.
- A previsão de valores futuros a partir do instante t, associados ao comportamento do processo, é computada em função do conhecimento de seus valores temporariamente anteriores, assim como de valores anteriores produzidos por suas saídas, isto é:

$$x(t) = f(x(t-1), x(t-2), ..., x(t-n_p), y(t-1), y(t-2), ..., y(t-n_q))$$

- ρ Onde $n_{
 ho}$ é a ordem do estimador, ou seja, a quantidade de medidas (amostras) passadas que serão necessárias para a estimação do valor x(t).
- O valor n_q expressa a ordem de contexto, ou seja, a quantidade de saídas passadas que serão também utilizadas na estimação de x(t).

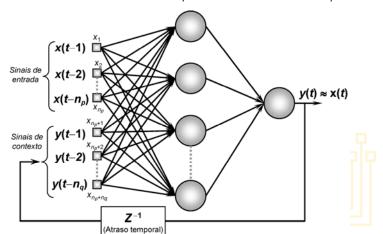
15

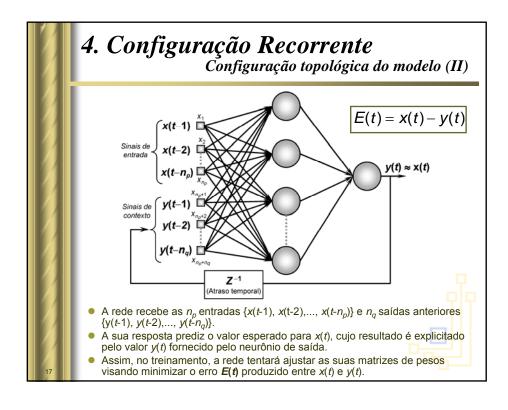
4. Configuração Recorrente

Configuração topológica do modelo (I)

 A figura seguinte ilustra um modelo de PMC recorrente que implementa o processo dinâmico explicitado por:

$$x(t) = f(x(t-1), x(t-2),..., x(t-n_p), y(t-1), y(t-2),..., y(t-n_q))$$





4. Configuração Recorrente Composição do conjunto de treinamento e teste (I) O treinamento da rede **PMC** com configuração Recorrente é similar ao PMC convencional, e o processo de aprendizado é também efetuado de maneira idêntica. Os cuidados que devem ser tomados estão associados com a montagem do conjunto de treinamento da rede. Do slide anterior, tem-se que: E(t) = x(t) - y(t), onde $(n_p + 1) \le t \le N$ > onde *N* é a quantidade total de medidas (amostras) disponíveis. Para elucidar tal mecanismo, considera-se que para um determinado sistema dinâmico foram colhidas as seguintes oito medidas $\{N = 8\}$ ao longo do tempo: t=3 t=5 t=6 $\mathbf{x}(t) = [0,11 \quad 0.32 \quad 0.53 \quad 0.17 \quad 0.98 \quad 0.67 \quad 0.83 \quad 0.79]^T$ x(2) x(3) x(5) x(6) x(7)

4. Configuração Recorrente Composição do conjunto de treinamento e teste (II) $x(t) = [0,11 \ 0.32 \ 0.53 \ 0.17 \ 0.98 \ 0.67$ $0,79]^{T}$ 0,83 x(1)x(2) x(3)x(4) x(5) x(6) x(7)x(8) Como exemplo, assume-se que o processo possa ser mapeado com ordem de predição igual a três $\{n_p=3\}$ e ordem de contexto igual a dois $\{n_q=2\}$. Então, tem-se: E(t) = x(t) - y(t), onde $(n_p + 1) \le t \le N$ Por intermédio da expressão acima, ter-se-á então um conjunto de treinamento composto por um total de 5 amostras, com parâmetro *t* variando de $4 \le t \le 8$. relação entradas/saídas conjunto de treinamento x₃ x₄ saida saida desejada da rede X2 X3 X4 X1 | See place of a rede | X(4) | Y(4) | Y(5) | Y(5) | X(6) | Y(6) | X(7) | Y(7) | X(8) | Y(8) | Y(8) | X(8) | Y(8) | X(9) $t = 4 \ x(3) \ x(2) \ x(1) \ 0 \ 0$ $t = 5 \ x(4) \ x(3) \ x(2) \ y(4) \ 0$ t = 6 | x(5) | x(4) | x(3) | y(5) | y(4)t = 7 | x(6) | x(5) | x(4) | y(6) | y(5) | x(7) | y(7)t = 8 |x(7)| x(6) |x(5)| y(7) |y(6)|onde o valor {x₀ = -1}, associado ao limiar do neurônio, deverá ser considerado em todos eles.

futuros.	
• Para o exemplo anterior, considerando agora $\{t \ge 9\}$, tal	
procedimento é realizado da seguinte maneira:	
X_1 X_2 X_3 X_4 X_5 saida estim.	ada
$t = 9$ $x(8)$ $x(7)$ $x(6)$ $y(8)$ $y(7)$ $x(9) \approx y(9)$	
$t = 10 x(9) x(8) x(7) y(9) y(8) x(10) \approx y(8)$	
$t = 11 \ x(10) \ x(9) \ x(8) \ y(10) \ y(9) \ x(11) \approx y(11)$	
() () () () () ()	

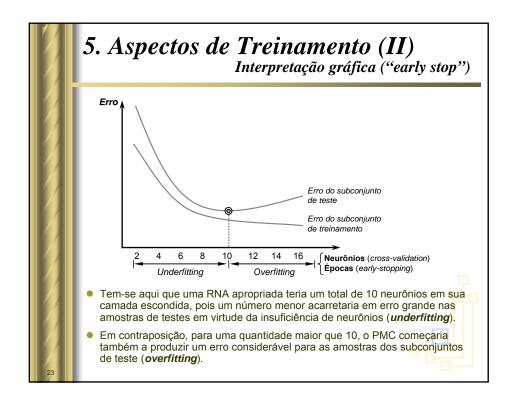
4. Configuração Recorrente Composição do conjunto de treinamento e teste (IV)

- Assim, para o exemplo anterior, a predição de seus valores futuros sempre levará em conta tanto as 3 entradas atrasadas no tempo como as 2 últimas saídas produzidas pela rede.
- Contudo, há situações em que tanto a ordem de predição como a ordem de contexto $\{n_a\}$ deve ser incrementada a fim de assegurar maior precisão na estimação do comportamento futuro do processo.
- O número de sinais de entrada p/ a rede será a soma de n_p com n_q.
- Generalizando, tem-se:
 - \succ Para n_p = 4 e n_q = 2 \Rightarrow Utiliza as 4 últimas entradas atrasadas no tempo, mais as 2 últimas saídas produzidas pela rede:
 - Rede terá 6 entradas.
 - Para n_p = 6 e n_q = 3 → Utiliza as 6 últimas entradas atrasadas no tempo, mais as 3 últimas saídas produzidas pela rede:
 - Rede terá 9 entradas.
- Então, quanto maior a ordem de predição e/ou ordem de contexto assumida, menor será também a quantidade de amostras para o processo de treinamento.

5. Aspectos de Treinamento (I)

Inclusão de parada antecipada ("early stop")

- Um dos problemas em aplicar o algoritmo backpropagation está justamente em se determinar o momento em que o mesmo já está generalizando de maneira satisfatória.
- Um procedimento eficaz a ser implementado na técnica crossvalidation é a parada antecipada ou prematura (early stopping).
- Aqui, o processo de aprendizagem para uma topologia candidata é constantemente checado pela aplicação dos subconjuntos de teste, sendo finalizado quando começar haver elevação do erro quadrático (frente aos subconjuntos de teste) entre épocas sucessivas.
- De fato, esta variação repentina sinaliza a tentativa de extração excessiva de características dos subconjuntos de treinamento, como por exemplo, os ruídos de medição.
- Em suma, a seleção de uma topologia de RNA deve levar em conta o compromisso de superar o *underfitting* e de evitar o *overfitting*.
- Com exemplo, considerando um PMC com apenas uma camada escondida, ilustra-se na figura seguinte o processo early stopping, assim como as situações de *underfitting* e *overfitting* em função do número de neurônios de sua camada escondida.



6. Aspectos de Convergência BP (I) Inserção do termo de momentum

- A inserção do termo de momentum se configura como uma das variações mais interessantes de ser efetuada no algoritmo backpropagation, pois, basta-se inserir um único parâmetro visando ponderar o quão as matrizes sinápticas foram alteradas entre duas iterações anteriores e sucessivas.
- Formalmente, considerando os neurônios pertencentes à L-ésima camada, tem-se:

$$W_{ji}^{(L)}(t+1) = W_{ji}^{(L)}(t) + \underbrace{\alpha \cdot (W_{ji}^{(L)}(t) - W_{ji}^{(L)}(t-1))}_{termo\ de \\ momentum} + \underbrace{\eta \cdot \delta_{j}^{(L)} \cdot Y_{i}^{(L-1)}}_{termo\ de \\ aprendizagem}$$

onde o parâmetro α é definido como sendo a taxa de **momentum** e seu valor está compreendido entre zero e um.

