



Sistemas Inteligentes

UNIDADE 6 – Redes Neurais Artificiais (Perceptron MultiCamadas/PMC)

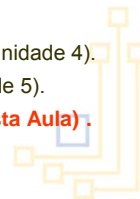
– Aplicação em Sistemas Variantes no Tempo –

Prof. Ivan Nunes da Silva



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 Unidade 4).
 - **Problemas envolvendo classificação de padrões** (Unidade 5).
 - **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.

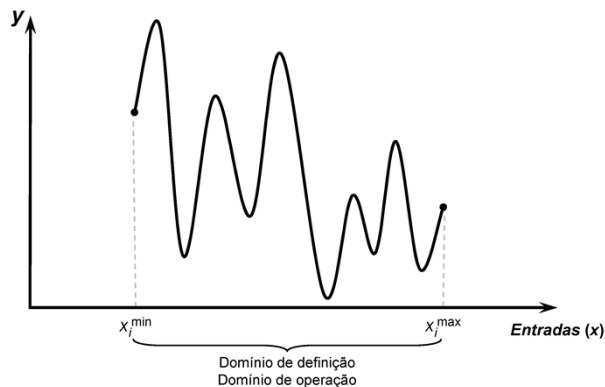


3

2. Problemas Variantes no Tempo

(Aproximação de Funções) X (Sistemas Variantes) // (I)

- Para problemas de **Aproximação de Funções**:
 - O domínio de definição referente aos dados de treinamento/teste são delimitados pelos valores mínimos e máximos associados a cada uma de suas variáveis de entrada.
 - O domínio de operação em que a efetiva aplicação da rede **PMC** estará sujeita, após ter sido treinada, deve obrigatoriamente coincidir com o seu domínio de definição.

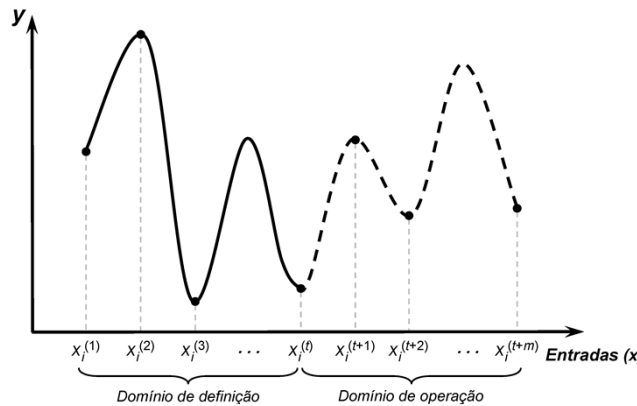


4

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.



5

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*).



6

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:

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

- 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).

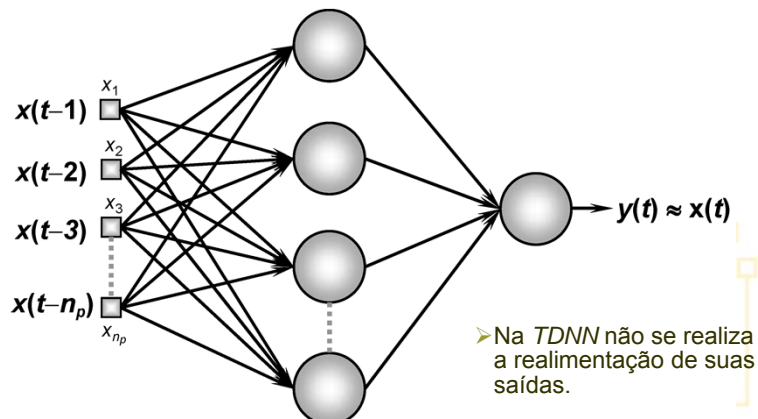


7

3. Configuração TDNN

Configuração topológica do modelo (I)

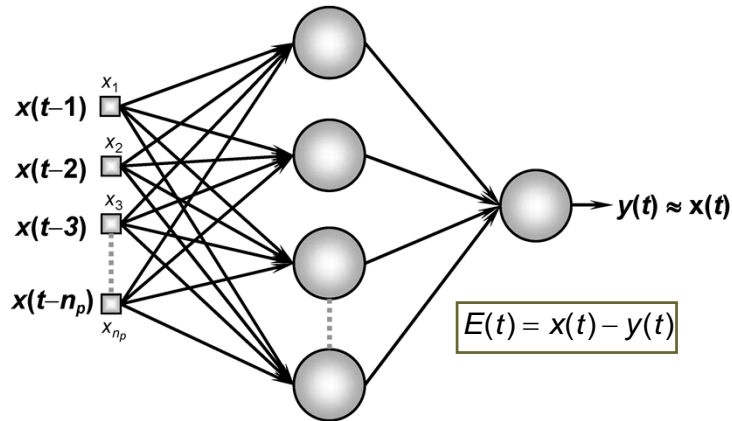
- De fato, a configuração **TDNN** estará aqui introduzindo linha de atrasos de tempo somente na camada de entrada.
- Esta linha funciona como uma memória, garantindo que amostras anteriores, que refletem o comportamento temporal do processo, sejam sempre inseridas dentro da rede.



8

3. Configuração TDNN

Configuração topológica do modelo (II)



- A rede recebe as n_p entradas $\{x(t-1), x(t-2), \dots, x(t-n_p)\}$.
- A sua resposta prediz o respectivo valor esperado para $x(t)$, cujo resultado é explicitado pelo valor $y(t)$ fornecido pelo neurônio de saída.
- Assim, no treinamento, a rede tentará ajustar as suas matrizes de pesos visando minimizar o erro $E(t)$ produzido entre $x(t)$ e $y(t)$.

9

3. Configuração TDNN

Composição do conjunto de treinamento e teste (I)

- O treinamento da rede **PMC** com configuração **TDNN** 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:

$$E(t) = x(t) - y(t), \text{ onde } (n_p + 1) \leq t \leq N$$

➤ 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:

$$x(t) = [0,11 \quad 0,32 \quad 0,53 \quad 0,17 \quad 0,98 \quad 0,67 \quad 0,83 \quad 0,79]^T$$

$\begin{matrix} t=1 & t=2 & t=3 & t=4 & t=5 & t=6 & t=7 & t=8 \\ x(1) & x(2) & x(3) & x(4) & x(5) & x(6) & x(7) & x(8) \end{matrix}$

10

3. Configuração TDNN

Composição do conjunto de treinamento e teste (II)

$$\mathbf{x}(t) = [0,11 \quad 0,32 \quad 0,53 \quad 0,17 \quad 0,98 \quad 0,67 \quad 0,83 \quad 0,79]^T$$

$x(1) \quad x(2) \quad x(3) \quad x(4) \quad x(5) \quad x(6) \quad x(7) \quad x(8)$

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

$$E(t) = \mathbf{x}(t) - \mathbf{y}(t), \text{ onde } (n_p + 1) \leq t \leq N$$

- 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 $4 \leq t \leq 8$.

relação entradas/saídas				conjunto de treinamento						
	x_1	x_2	x_3	saída desejada		x_1	x_2	x_3	d	
$t = 4$	$x(3)$	$x(2)$	$x(1)$	$x(4)$	$4 \leq t \leq 8$ (ordem 3) $n_p = 3$	$\mathbf{x}^{(1)}$	0,53	0,32	0,11	$d^{(1)} = 0,17$
$t = 5$	$x(4)$	$x(3)$	$x(2)$	$x(5)$		$\mathbf{x}^{(2)}$	0,17	0,53	0,32	$d^{(2)} = 0,98$
$t = 6$	$x(5)$	$x(4)$	$x(3)$	$x(6)$		$\mathbf{x}^{(3)}$	0,98	0,17	0,53	$d^{(3)} = 0,67$
$t = 7$	$x(6)$	$x(5)$	$x(4)$	$x(7)$		$\mathbf{x}^{(4)}$	0,67	0,98	0,17	$d^{(4)} = 0,83$
$t = 8$	$x(7)$	$x(6)$	$x(5)$	$x(8)$		$\mathbf{x}^{(5)}$	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.

11

3. Configuração TDNN

Composição do conjunto de treinamento e teste (III)

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

$$\mathbf{x}(t) = [\overset{\text{janela 1 (t=4)}}{\boxed{0,11 \quad 0,32 \quad 0,53 \quad 0,17}} \quad 0,98 \quad 0,67 \quad 0,83 \quad 0,79]^T$$

$$\mathbf{x}(t) = [0,11 \quad \overset{\text{janela 2 (t=5)}}{\boxed{0,32 \quad 0,53 \quad 0,17 \quad 0,98}} \quad 0,67 \quad 0,83 \quad 0,79]^T$$

$$\mathbf{x}(t) = [0,11 \quad 0,32 \quad \overset{\text{janela 3 (t=6)}}{\boxed{0,53 \quad 0,17 \quad 0,98 \quad 0,67}} \quad 0,83 \quad 0,79]^T$$

$$\mathbf{x}(t) = [0,11 \quad 0,32 \quad 0,53 \quad \overset{\text{janela 4 (t=7)}}{\boxed{0,17 \quad 0,98 \quad 0,67 \quad 0,83}} \quad 0,79]^T$$

$$\mathbf{x}(t) = [0,11 \quad 0,32 \quad 0,53 \quad 0,17 \quad \overset{\text{janela 5 (t=8)}}{\boxed{0,98 \quad 0,67 \quad 0,83 \quad 0,79}}]^T$$

- 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).

12

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 $\{t \geq 9\}$, tal procedimento é realizado da seguinte maneira:

predição de valores futuros

	x_1	x_2	x_3	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 $\{x_1=x(10), x_2=x(9), x_3=x(8)\}$
- (...)

$$\mathbf{x}(t) = [0,11 \quad 0,32 \quad 0,53 \quad 0,17 \quad 0,98 \quad 0,67 \quad 0,83 \quad 0,79]^T$$

$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$
$x(1)$	$x(2)$	$x(3)$	$x(4)$	$x(5)$	$x(6)$	$x(7)$	$x(8)$

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 $n_p = 4$ → Utiliza os 4 últimos valores para prever o valor atual.
 - Para $n_p = 5$ → Utiliza os 5 últimos valores para prever o valor atual.
 - (...)
- Entretanto, quanto maior a ordem de predição assumida, menor será também a quantidade de amostras para o processo de treinamento.

$$\mathbf{x}(t) = [0,11 \quad 0,32 \quad 0,53 \quad 0,17 \quad 0,98 \quad 0,67 \quad 0,83 \quad 0,79]^T$$

$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$
$x(1)$	$x(2)$	$x(3)$	$x(4)$	$x(5)$	$x(6)$	$x(7)$	$x(8)$

$$E(t) = x(t) - y(t), \text{ onde } (n_p + 1) \leq t \leq N$$

14

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:

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

- 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 valor n_q expressa a ordem de contexto, ou seja, a quantidade de saídas passadas que serão também utilizadas na estimação de $x(t)$.

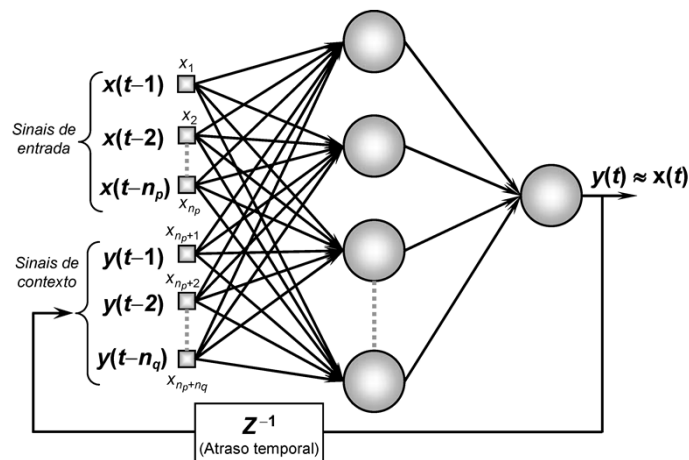
15

4. Configuração Recorrente

Configuração topológica do modelo (I)

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

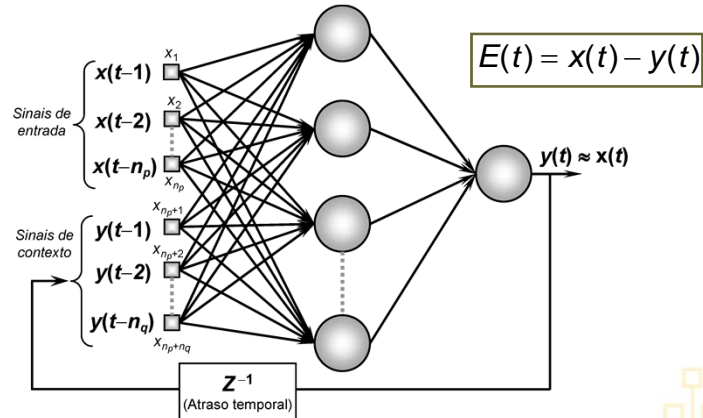
$$x(t) = f(x(t-1), x(t-2), \dots, x(t-n_p), y(t-1), y(t-2), \dots, y(t-n_q))$$



16

4. Configuração Recorrente

Configuração topológica do modelo (II)



- A rede recebe as n_p entradas $\{x(t-1), x(t-2), \dots, x(t-n_p)\}$ e n_q saídas anteriores $\{y(t-1), y(t-2), \dots, y(t-n_q)\}$.
- A sua resposta prediz o valor esperado para $x(t)$, cujo resultado é explicitado pelo valor $y(t)$ fornecido pelo neurônio de saída.
- Assim, no treinamento, a rede tentará ajustar as suas matrizes de pesos visando minimizar o erro $E(t)$ produzido entre $x(t)$ e $y(t)$.

17

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:

$$E(t) = x(t) - y(t), \text{ onde } (n_p + 1) \leq t \leq N$$

➤ 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:

$$\mathbf{x}(t) = \begin{bmatrix} 0,11 & 0,32 & 0,53 & 0,17 & 0,98 & 0,67 & 0,83 & 0,79 \end{bmatrix}^T$$

$x(1) \quad x(2) \quad x(3) \quad x(4) \quad x(5) \quad x(6) \quad x(7) \quad x(8)$

18

4. Configuração Recorrente

Composição do conjunto de treinamento e teste (II)

$$\mathbf{x}(t) = [0,11 \quad 0,32 \quad 0,53 \quad 0,17 \quad 0,98 \quad 0,67 \quad 0,83 \quad 0,79]^T$$

$x(1) \quad x(2) \quad x(3) \quad x(4) \quad x(5) \quad x(6) \quad x(7) \quad x(8)$

- 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:

$$E(t) = x(t) - y(t), \text{ onde } (n_p + 1) \leq t \leq N$$

- 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 $4 \leq t \leq 8$.

relação entradas/saídas							conjunto de treinamento								
	x_1	x_2	x_3	x_4	x_5	saída desejada	saída da rede		x_1	x_2	x_3	x_4	x_5	d	
$t=4$	$x(3)$	$x(2)$	$x(1)$	0	0	$x(4)$	$y(4)$	$4 \leq t \leq 8$ (ordem 3) $n_p = 3$ $n_q = 2$	$x^{(1)}$	0,53	0,32	0,11	0	0	$d^{(1)} = 0,17$
$t=5$	$x(4)$	$x(3)$	$x(2)$	$y(4)$	0	$x(5)$	$y(5)$		$x^{(2)}$	0,17	0,53	0,32	$y(4)$	0	$d^{(2)} = 0,98$
$t=6$	$x(5)$	$x(4)$	$x(3)$	$y(5)$	$y(4)$	$x(6)$	$y(6)$		$x^{(3)}$	0,98	0,17	0,53	$y(5)$	$y(4)$	$d^{(3)} = 0,67$
$t=7$	$x(6)$	$x(5)$	$x(4)$	$y(6)$	$y(5)$	$x(7)$	$y(7)$		$x^{(4)}$	0,67	0,98	0,17	$y(6)$	$y(5)$	$d^{(4)} = 0,83$
$t=8$	$x(7)$	$x(6)$	$x(5)$	$y(7)$	$y(6)$	$x(8)$	$y(8)$		$x^{(5)}$	0,83	0,67	0,98	$y(7)$	$y(6)$	$d^{(5)} = 0,79$

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

19

4. Configuração Recorrente

Composição do conjunto de treinamento e teste (III)

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

	x_1	x_2	x_3	x_4	x_5	saída estimada
$t=9$	$x(8)$	$x(7)$	$x(6)$	$y(8)$	$y(7)$	$x(9) \approx y(9)$
$t=10$	$x(9)$	$x(8)$	$x(7)$	$y(9)$	$y(8)$	$x(10) \approx y(10)$
$t=11$	$x(10)$	$x(9)$	$x(8)$	$y(10)$	$y(9)$	$x(11) \approx y(11)$
(...)	(...)	(...)	(...)	(...)	(...)	(...)

- Para obter $x(9)$ → Entradas $\{x_1=x(8), x_2=x(7), x_3=x(6), x_4=y(8), x_5=y(7)\}$
- Para obter $x(10)$ → Entradas $\{x_1=x(9), x_2=x(8), x_3=x(7), x_4=y(9), x_5=y(8)\}$
- Para obter $x(11)$ → Entradas $\{x_1=x(10), x_2=x(9), x_3=x(8), x_4=y(10), x_5=y(9)\}$
- (...)

$$\mathbf{x}(t) = [0,11 \quad 0,32 \quad 0,53 \quad 0,17 \quad 0,98 \quad 0,67 \quad 0,83 \quad 0,79]^T$$

$x(1) \quad x(2) \quad x(3) \quad x(4) \quad x(5) \quad x(6) \quad x(7) \quad x(8)$

20

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 $\{n_q\}$ 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:
 - Para $n_p = 4$ e $n_q = 2$ → 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.

21

Fim da Apresentação



22