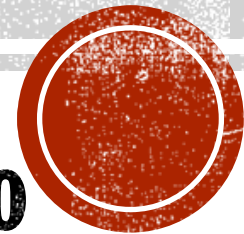


REDES NEURAIS ARTIFICIAIS

AULA 6 – REDE *PERCEPTRON MULTICAMADAS* APLICAÇÃO EM SISTEMAS VARIANTES NO TEMPO



Prof. Rodrigo Palácios

rodrigopalacios@utfpr.edu.br

PMC - 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.**
 - **Problemas envolvendo classificação de padrões.**
 - **Problemas envolvendo sistemas variantes no tempo.**

PMC - PROBLEMAS VARIANTES NO TEMPO

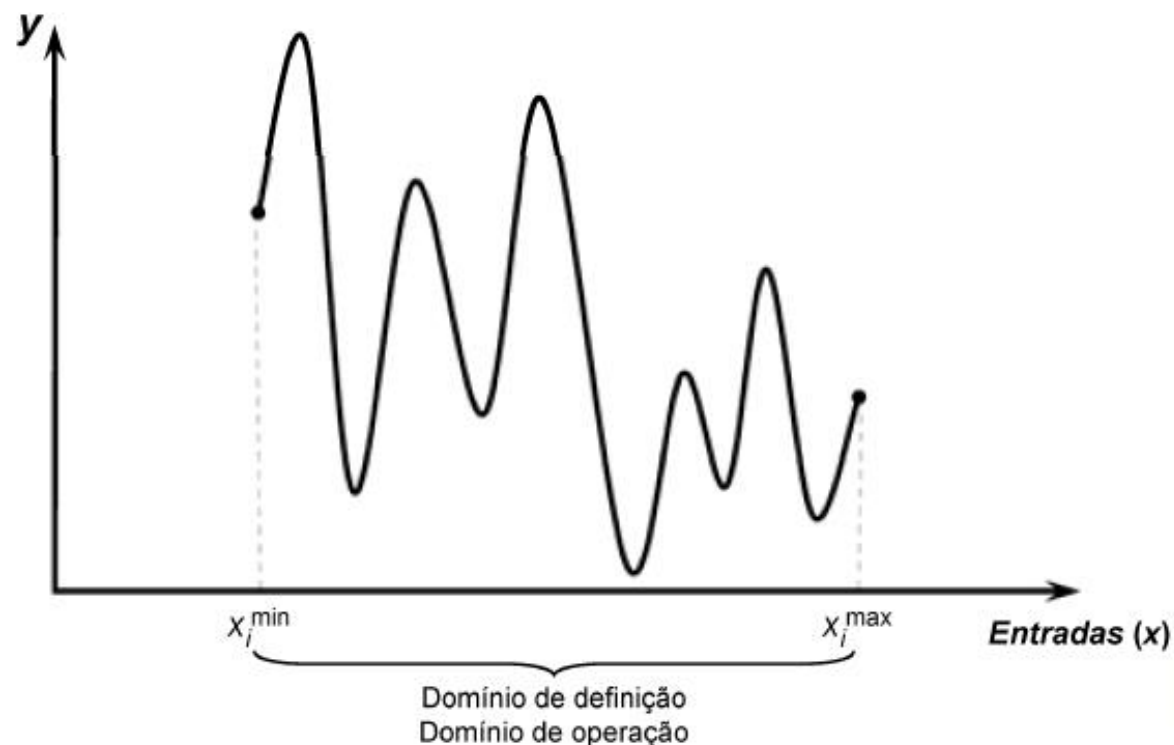
■ Aspectos de definição

- 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.

PMC - PROBLEMAS VARIANTES NO TEMPO

■ Para problemas de Aproximação de Funções:

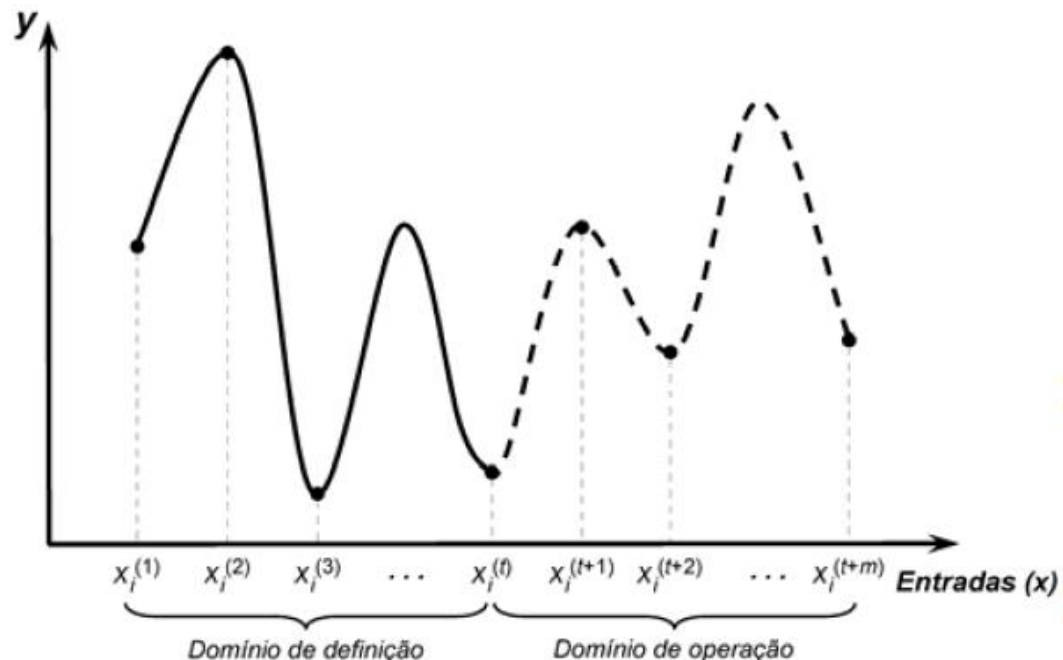
- O domínio de definição referente aos dados de treinamento/teste são delimitados pelos valores mínimos e máximos associados a cada uma de suas variáveis de entrada.
- O domínio de operação em que a efetiva aplicação da rede **PMC** estará sujeita, após ter sido treinada, deve obrigatoriamente coincidir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



PMC - PROBLEMAS VARIANTES NO TEMPO

■ 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.



PMC - 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*).

PMC - 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:

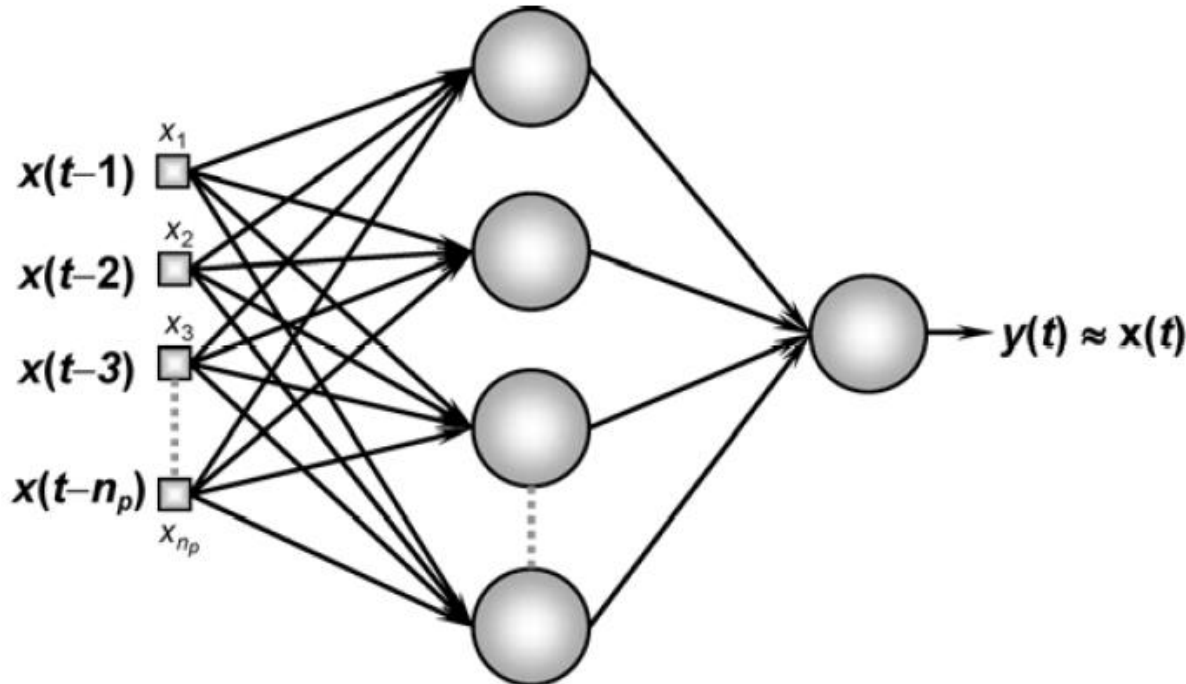
$$x(t) = f(x(t-1), x(t-2), \dots, x(t-n_p))$$

- 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).

PMC - CONFIGURAÇÃO TDNN

■ Configuração topológica do modelo

- 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.

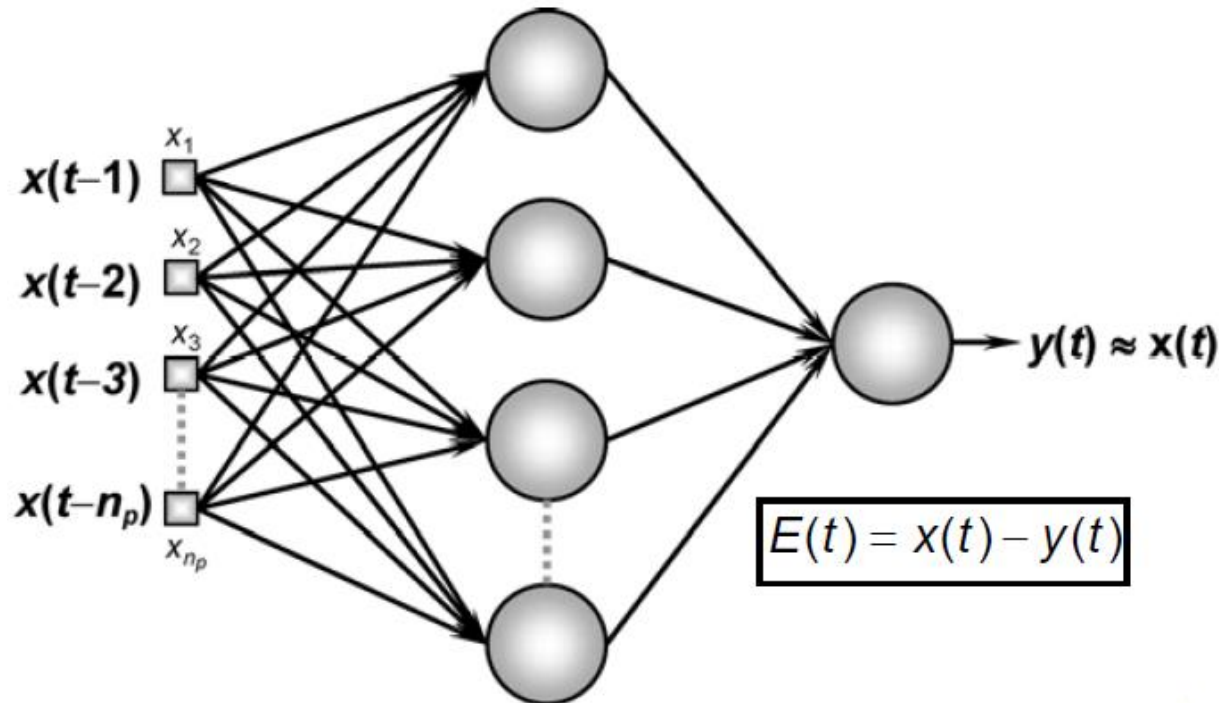


- Na *TDNN* não se realiza a realimentação de suas saídas

PMC - CONFIGURAÇÃO TDNN

■ Configuração topológica do modelo

- 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 $y(t)$.



PMC - CONFIGURAÇÃO TDNN

▪ Composição do conjunto de treinamento e teste

- 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:


$$x(t) = \begin{matrix} t=1 & t=2 & t=3 & t=4 & t=5 & t=6 & t=7 & t=8 \\ [0,11 & 0,32 & 0,53 & 0,17 & 0,98 & 0,67 & 0,83 & 0,79]^T \\ x(1) & x(2) & x(3) & x(4) & x(5) & x(6) & x(7) & x(8) \end{matrix}$$

PMC - CONFIGURAÇÃO TDNN

■ Composição do conjunto de treinamento e teste

$$x(t) = \begin{matrix} t=1 & t=2 & t=3 & t=4 & t=5 & t=6 & t=7 & t=8 \\ [0,11 & 0,32 & 0,53 & 0,17 & 0,98 & 0,67 & 0,83 & 0,79]^T \\ x(1) & x(2) & x(3) & x(4) & x(5) & x(6) & x(7) & x(8) \end{matrix}$$

- 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 $(n_p + 1) \leq t \leq N$
- 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 | | | | | $4 \leq t \leq 8$  (ordem 3) $n_p = 3$ | conjunto de treinamento | | | | |
|-------------------------|--------|--------|--------|----------------|--|-------------------------|-------|-------|-------|---------------------------|
| | x_1 | x_2 | x_3 | saída desejada | | | x_1 | x_2 | x_3 | d |
| $t = 4$ | $x(3)$ | $x(2)$ | $x(1)$ | $x(4)$ | | $\mathbf{x}^{(1)}$ | 0,53 | 0,32 | 0,11 | $\mathbf{d}^{(1)} = 0,17$ |
| $t = 5$ | $x(4)$ | $x(3)$ | $x(2)$ | $x(5)$ | | $\mathbf{x}^{(2)}$ | 0,17 | 0,53 | 0,32 | $\mathbf{d}^{(2)} = 0,98$ |
| $t = 6$ | $x(5)$ | $x(4)$ | $x(3)$ | $x(6)$ | | $\mathbf{x}^{(3)}$ | 0,98 | 0,17 | 0,53 | $\mathbf{d}^{(3)} = 0,67$ |
| $t = 7$ | $x(6)$ | $x(5)$ | $x(4)$ | $x(7)$ | | $\mathbf{x}^{(4)}$ | 0,67 | 0,98 | 0,17 | $\mathbf{d}^{(4)} = 0,83$ |
| $t = 8$ | $x(7)$ | $x(6)$ | $x(5)$ | $x(8)$ | | $\mathbf{x}^{(5)}$ | 0,83 | 0,67 | 0,98 | $\mathbf{d}^{(5)} = 0,79$ |

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

PMC - CONFIGURAÇÃO TDNN

■ Composição do conjunto de treinamento e teste

- 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.

$$\begin{aligned} \mathbf{x}(t) &= [\underbrace{0,11 \quad 0,32 \quad 0,53 \quad 0,17}_{\text{janela 1 (t=4)}} \quad 0,98 \quad 0,67 \quad 0,83 \quad 0,79]^T \\ \mathbf{x}(t) &= [0,11 \quad \underbrace{0,32 \quad 0,53 \quad 0,17 \quad 0,98}_{\text{janela 2 (t=5)}} \quad 0,67 \quad 0,83 \quad 0,79]^T \\ \mathbf{x}(t) &= [0,11 \quad 0,32 \quad \underbrace{0,53 \quad 0,17 \quad 0,98 \quad 0,67}_{\text{janela 3 (t=6)}} \quad 0,83 \quad 0,79]^T \\ \mathbf{x}(t) &= [0,11 \quad 0,32 \quad 0,53 \quad \underbrace{0,17 \quad 0,98 \quad 0,67 \quad 0,83}_{\text{janela 4 (t=7)}} \quad 0,79]^T \\ \mathbf{x}(t) &= [0,11 \quad 0,32 \quad 0,53 \quad 0,17 \quad \underbrace{0,98 \quad 0,67 \quad 0,83 \quad 0,79}_{\text{janela 5 (t=8)}}]^T \end{aligned}$$

- 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).


PMC - CONFIGURAÇÃO TDNN

■ Composição do conjunto de treinamento e teste

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

$$x(t) = \begin{matrix} t=1 & t=2 & t=3 & t=4 & t=5 & t=6 & t=7 & t=8 \\ [0,11 & 0,32 & 0,53 & 0,17 & 0,98 & 0,67 & 0,83 & 0,79]^T \\ x(1) & x(2) & x(3) & x(4) & x(5) & x(6) & x(7) & x(8) \end{matrix}$$

PMC - CONFIGURAÇÃO TDNN

■ Composição do conjunto de treinamento e teste

- Assim, para o exemplo anterior, conclui-se que a rede sempre realiza uma predição de um passo à frente, calculando sequencialmente 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 estimativa 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.

$$\mathbf{x}(t) = \begin{matrix} t=1 & t=2 & t=3 & t=4 & t=5 & t=6 & t=7 & t=8 \\ [0,11 & 0,32 & 0,53 & 0,17 & 0,98 & 0,67 & 0,83 & 0,79]^T \\ x(1) & x(2) & x(3) & x(4) & x(5) & x(6) & x(7) & x(8) \end{matrix}$$

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

PMC - ASPECTOS DO TREINAMENTO

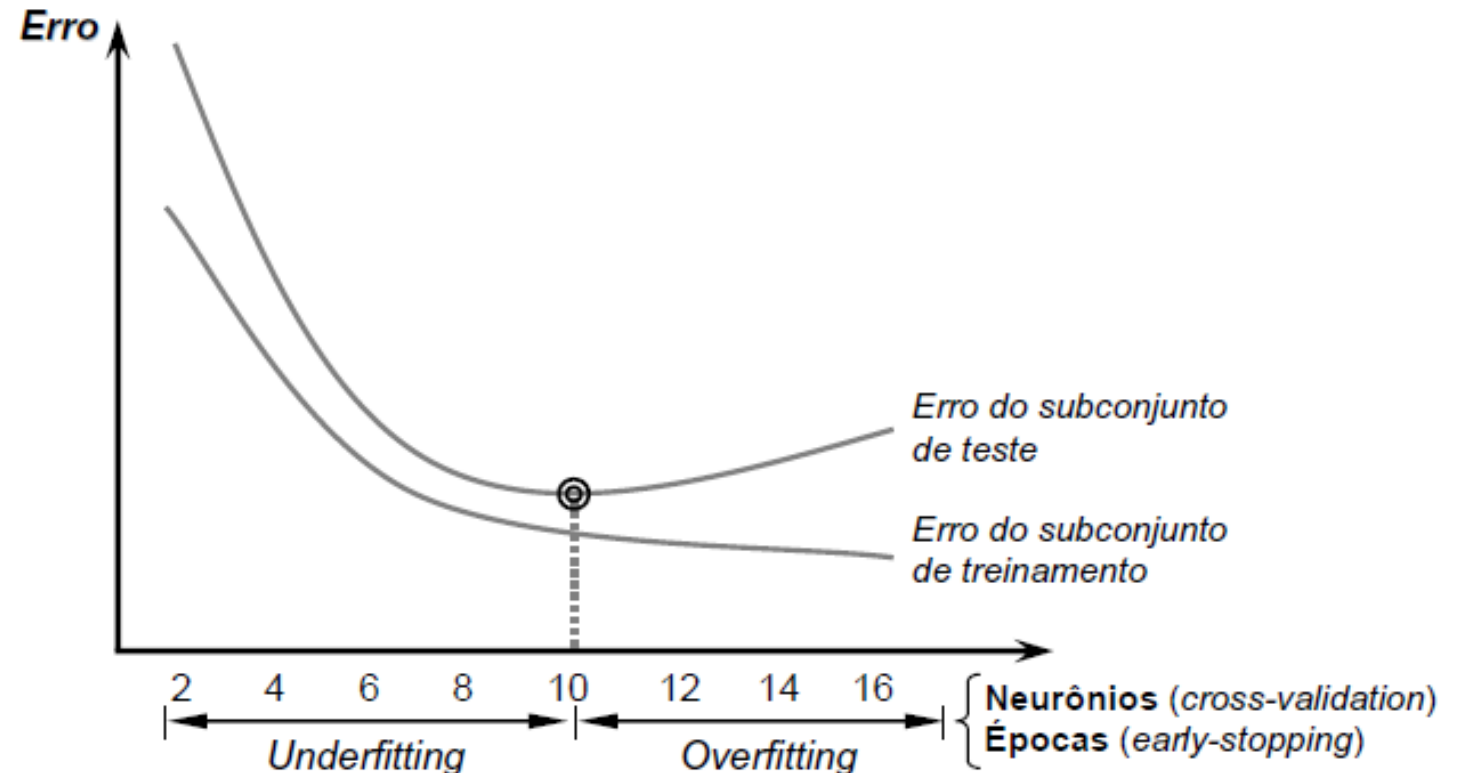
■ Inclusão de parada antecipada (“*early stop*”)

- 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

PMC - ASPECTOS DO TREINAMENTO

■ Interpretação gráfica

- Tem-se aqui que uma RNA apropriada teria um total de 10 neurônios em sua camada escondida, pois um número menor acarretaria em erro grande nas amostras de testes em virtude da insuficiência de neurônios (**underfitting**).
- Em contraposição, para uma quantidade maior que 10, o PMC começaria também a produzir um erro considerável para as amostras dos subconjuntos de teste (**overfitting**).



PMC - ASPECTOS DE CONVERGÊNCIA

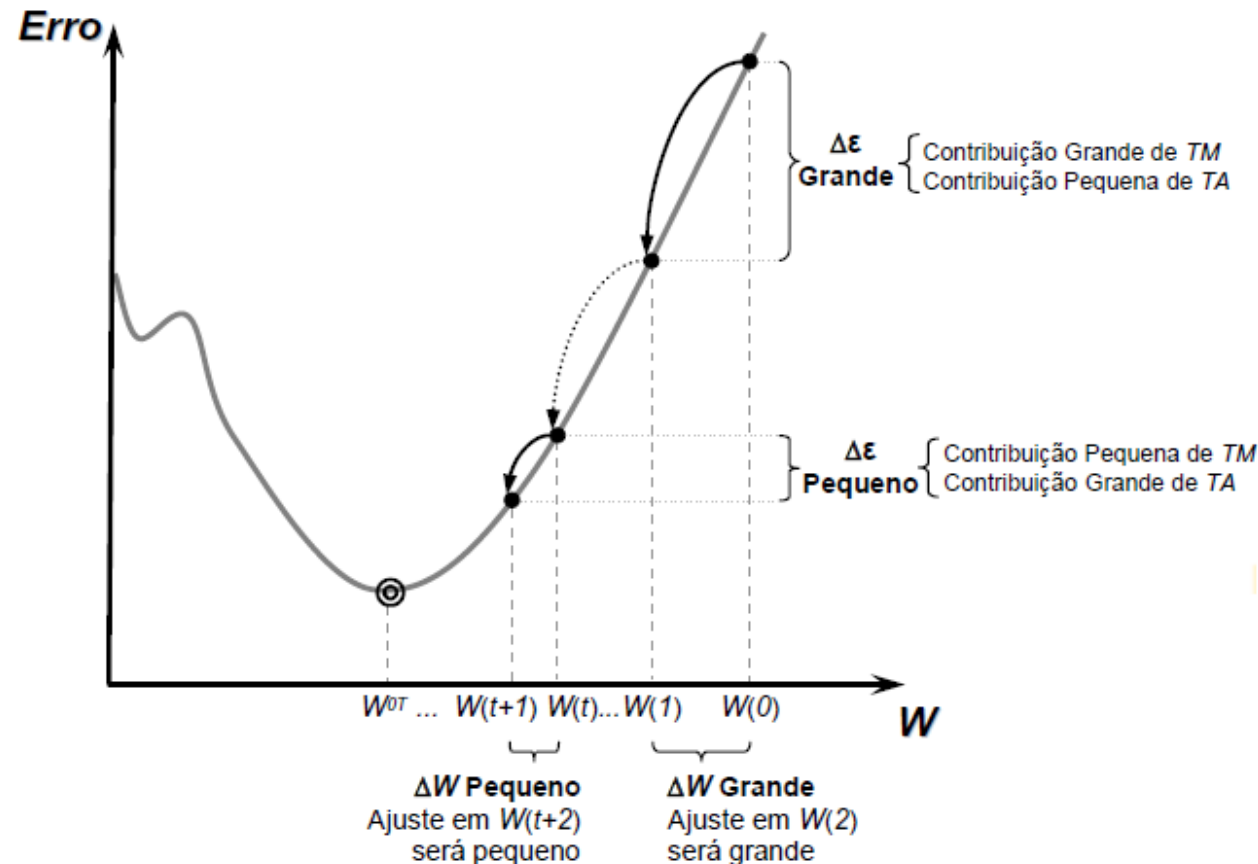
- **Inserção do termo de *momentum***

- 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.

PMC - ASPECTOS DE CONVERGÊNCIA

■ Interpretação gráfica (termo de *momentum*)

- A figura seguinte ilustra a contribuição do termo de *momentum* (TM) e do termo de aprendizagem (TA) visando a convergência em direção ao mínimo W^{OT} da função erro quadrático



REFERÊNCIA

- SILVA, Ivan Nunes da e SPATTI, Danilo Hernane e FLAUZINO, Rogério Andrade. Redes neurais artificiais para engenharia e ciências aplicadas. . São Paulo: Artliber Editora. . Acesso em: 23 dez. 2023. , 2010