



Multilayer Perceptron (MLP)

Professor Ciniro Nametala
Bacharelado em Engenharia de Computação
Instituto Federal de Minas Gerais – Campus Bambuí

Código da disciplina: BiSuCOM.553
Oferta: 2-2025



<https://tinyurl.com/aula4rnaciniro>

Agenda

1. Introdução
2. Topologia vs. Arquitetura
3. MLP
 1. Análise matemática
 1. Fase forward
 2. Fase backward
 1. Backpropagation
4. Pseudocódigo de treinamento
5. Pseudocódigo de previsão
6. Implementação

Introdução

- As redes Perceptron de Múltiplas Camadas (PMC), ou do inglês, *Multilayer Perceptron* (MLP) são caracterizadas pela presença de pelo menos uma camada intermediária entre a camada de entrada e a camada de saída. Essa camada intermediária é também chamada de camada escondida.
- Possui, portanto, pelo menos duas camadas.
- Tornaram-se populares e foram responsáveis por resgatar as pesquisas em RNA's a partir do final da década de 80, após a publicação de Rumelhart (1986) que descreveu detalhadamente o algoritmo de treinamento Backpropagation ou Regra Delta Generalizada.

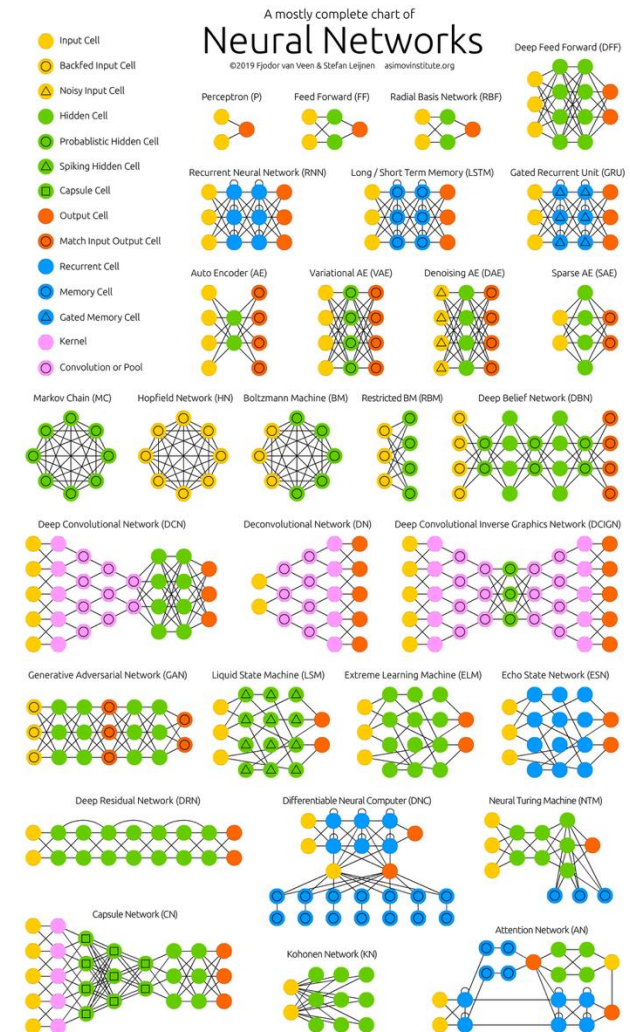
Introdução

- É uma das arquiteturas de RNA's mais versáteis que existe, pois permite realizar:
 - Aproximação universal de funções.
 - Reconhecimento de padrões.
 - Previsão em séries temporais.
 - Identificação e controle de processos.
 - Otimização de sistemas.
- Destaque para os três primeiros itens.
- É considerada também uma rede feedforward, contudo, que incorpora o conceito de backward na fase de treinamento.
- É uma rede de aprendizado supervisionado.
- Não possui qualquer realimentação.

Topologia vs. Arquitetura

Arquitetura em RNA diz respeito aos projetos padrão de um algoritmo de Rede Neural Artificial que se referem ao uso de técnicas específicas daquela arquitetura em questão. Por exemplo: Perceptron simples, MLP, Elman, Jordan, RBF, CNN, LSTM, GRU, Autoencoder, Transformer e outras.

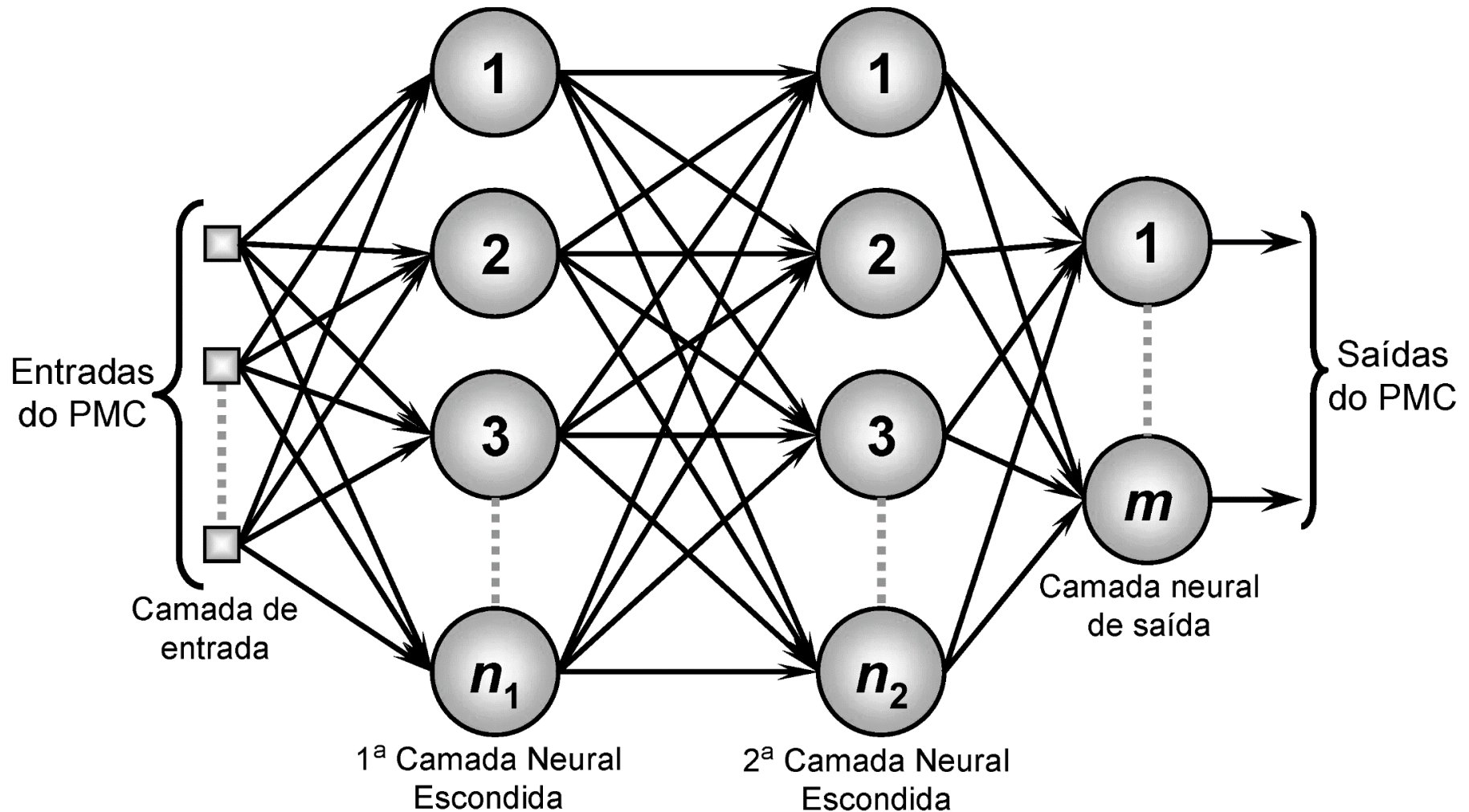
Topologia em RNA diz respeito a diferentes configurações usando uma mesma arquitetura. Por exemplo, redes de arquitetura MLP podem ser configuradas quantidades diferentes de camadas, neurônios, funções de ativação diferentes e hiperparâmetros diversos.



MLP no treinamento de carros virtuais numa pista



Esquema de uma MLP



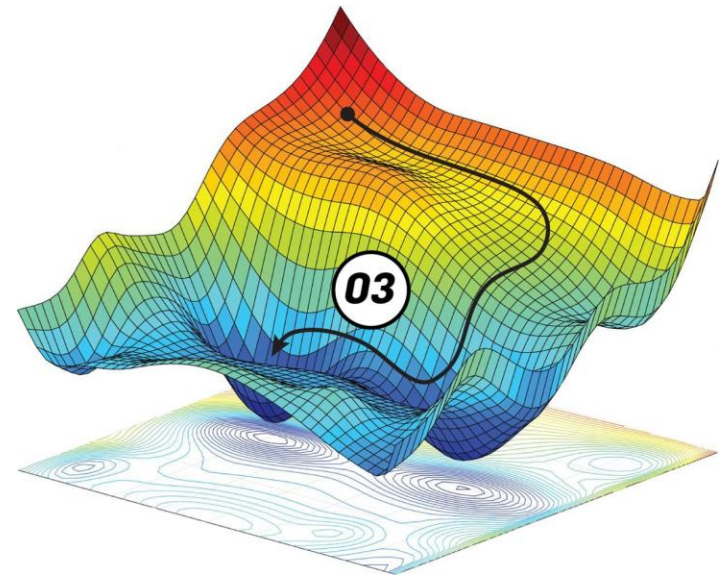
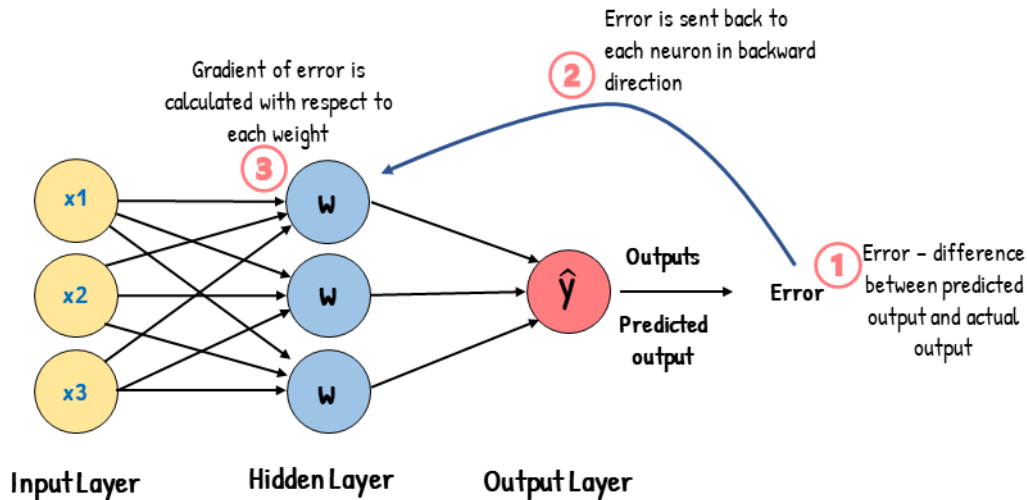
Esquema de uma MLP

- Diferente da Perceptron simples a MLP pode mapear um ou mais entradas para mais de uma saída.
- É uma arquitetura que realiza qualquer tipo de separação em espaços n-dimensionais.
- O processo de conhecimento está distribuído em várias camadas, em especial, na camada intermediária.
- A configuração topológica vai depender principalmente de:
 - Classe do problema a ser tratado.
 - Disposição espacial das amostras de treinamento.
 - Valores iniciais atribuídos aos pesos sinápticos.

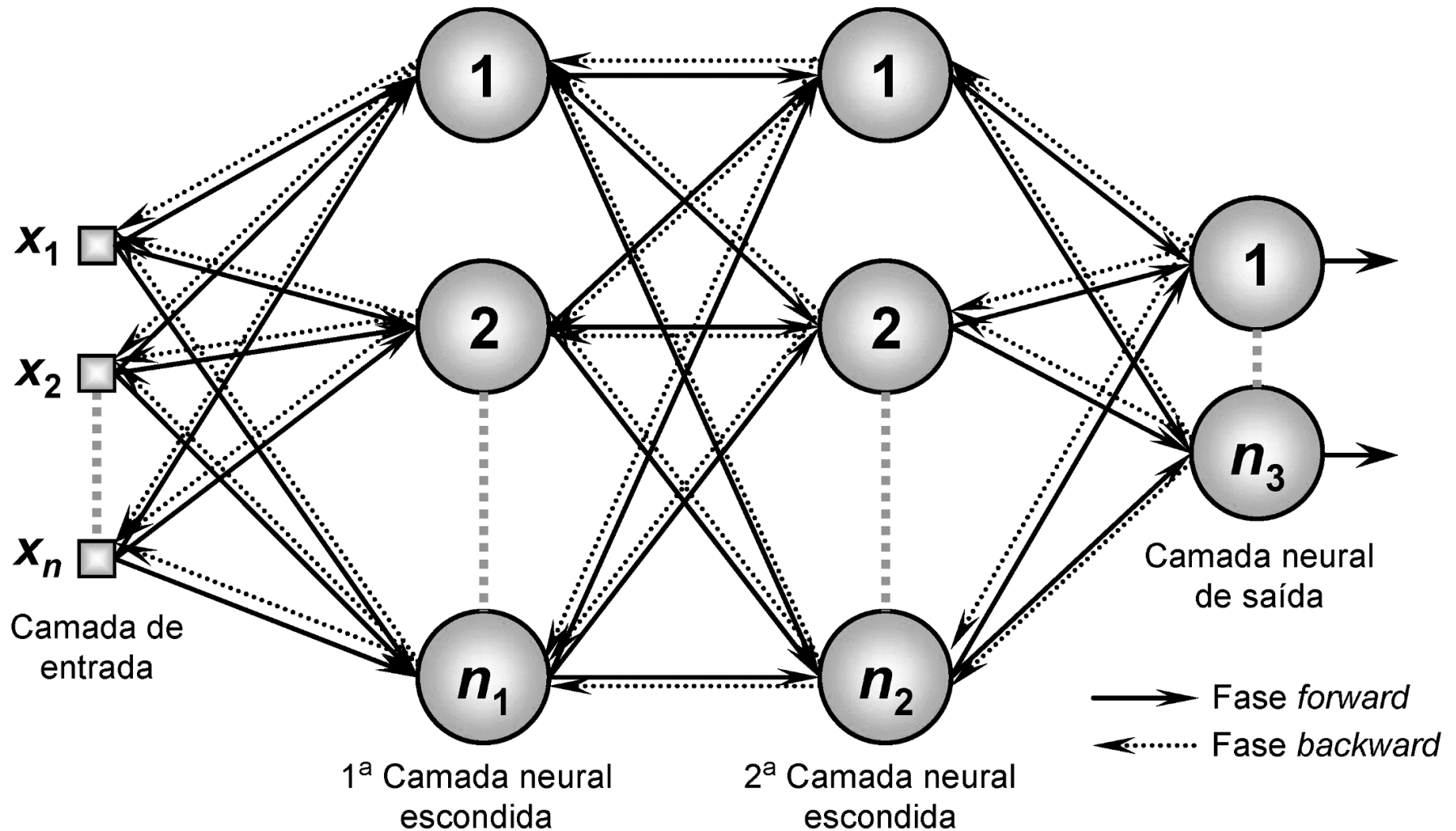
Backpropagation

Existem vários algoritmos de otimização de pesos em RNAs, na MLP, o clássico é o algoritmo Backpropagation, também conhecido como Regra Delta Generalizada.

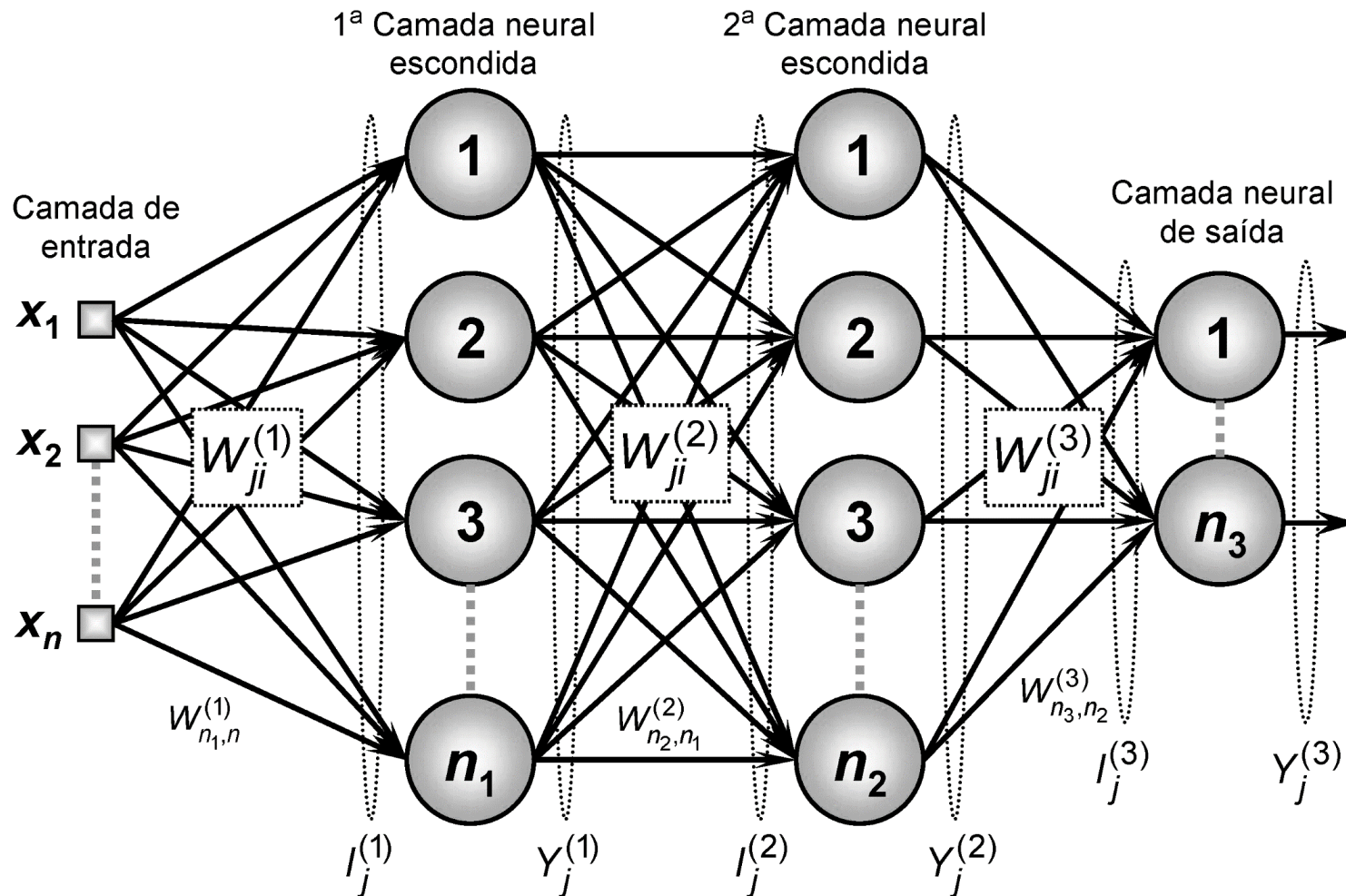
Backpropagation



Fases forward e backward



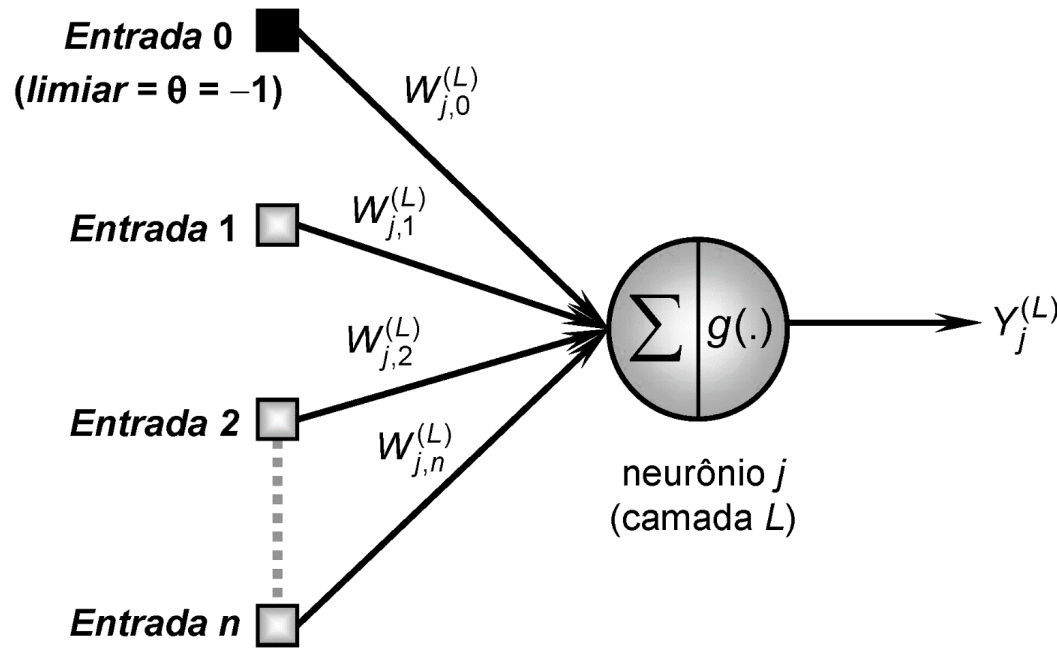
Análise matemática



Análise matemática

Fase forward

- Cada um dos j neurônios pertencentes a uma camada L pode ser configurado conforme a figura a seguir.
- A função de ativação neste caso deve ser **CONTÍNUA E DIFERENCIÁVEL EM TODO O SEU DOMÍNIO**
- O mais comum é usar uma **função tangente hiperbólica**.



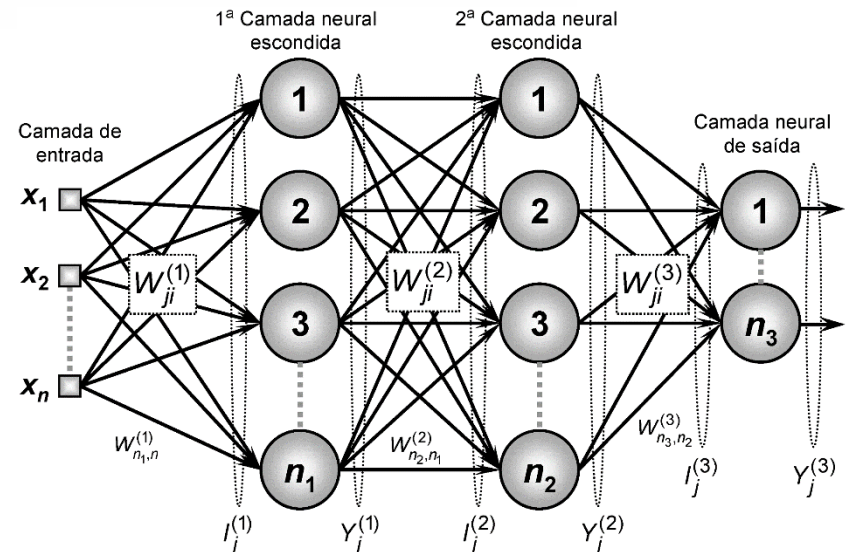
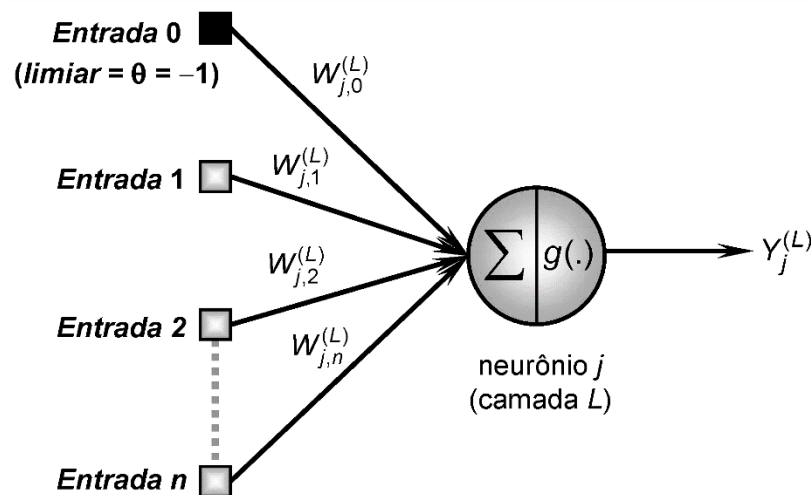
- $W_{ji}^{(L)}$ são matrizes de pesos cujos elementos denotam o valor do peso sináptico conectando o j -ésimo neurônio da camada (L) ao i -ésimo neurônio da camada ($L-1$). Para a topologia ilustrada na figura 5.3, tem-se:

- $W_{ji}^{(3)}$ é o peso sináptico conectando o j -ésimo neurônio da camada de saída ao i -ésimo neurônio da camada 2;

- $W_{ji}^{(2)}$ é o peso sináptico conectando o j -ésimo neurônio da camada escondida 2 ao i -ésimo neurônio da camada 1;

- $W_{ji}^{(1)}$ é o peso sináptico conectando o j -ésimo neurônio da camada 1 ao i -ésimo sinal da camada de entrada.

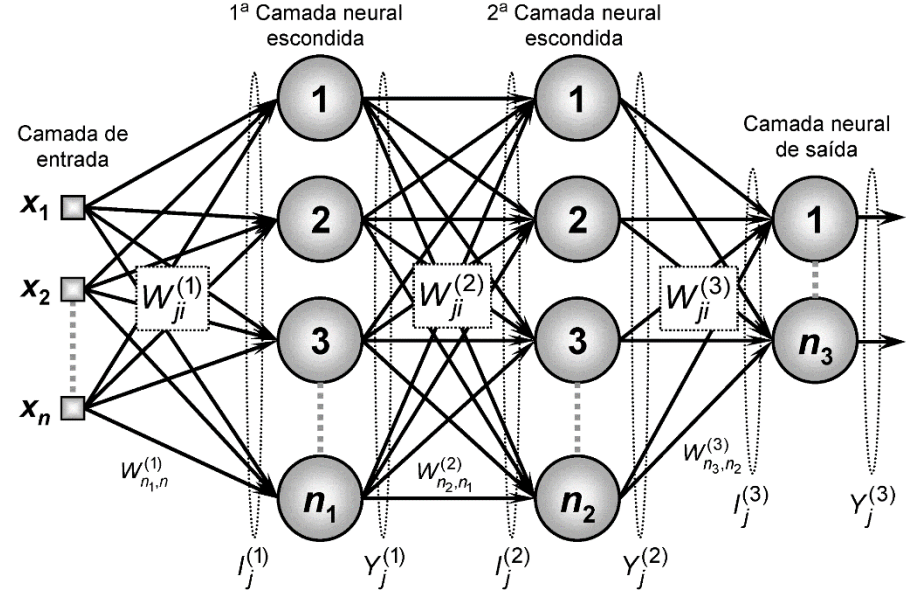
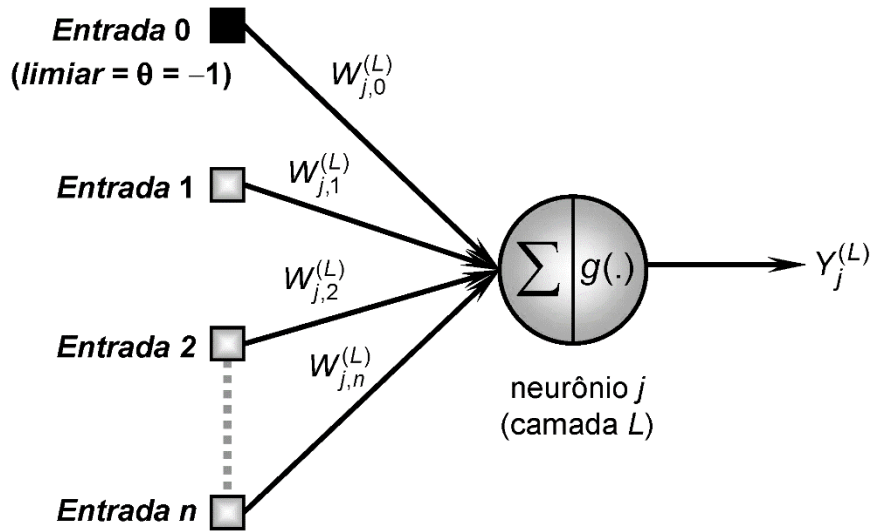
- $I_{ji}^{(L)}$ são vetores cujos elementos denotam a entrada ponderada em relação ao j -ésimo neurônio da camada L , os quais são definidos por:



$$I_j^{(1)} = \sum_{i=0}^n W_{ji}^{(1)} \cdot x_i \Leftrightarrow I_j^{(1)} = W_{j,0}^{(1)} \cdot x_0 + W_{j,1}^{(1)} \cdot x_1 + \dots + W_{j,n}^{(1)} \cdot x_n \quad (5.1)$$

$$I_j^{(2)} = \sum_{i=0}^{n_1} W_{ji}^{(2)} \cdot Y_i^{(1)} \Leftrightarrow I_j^{(2)} = W_{j,0}^{(2)} \cdot Y_0^{(1)} + W_{j,1}^{(2)} \cdot Y_1^{(1)} + \dots + W_{j,n_1}^{(2)} \cdot Y_{n_1}^{(1)} \quad (5.2)$$

$$I_j^{(3)} = \sum_{i=0}^{n_2} W_{ji}^{(3)} \cdot Y_i^{(2)} \Leftrightarrow I_j^{(3)} = W_{j,0}^{(3)} \cdot Y_0^{(2)} + W_{j,1}^{(3)} \cdot Y_1^{(2)} + \dots + W_{j,n_2}^{(3)} \cdot Y_{n_2}^{(2)} \quad (5.3)$$

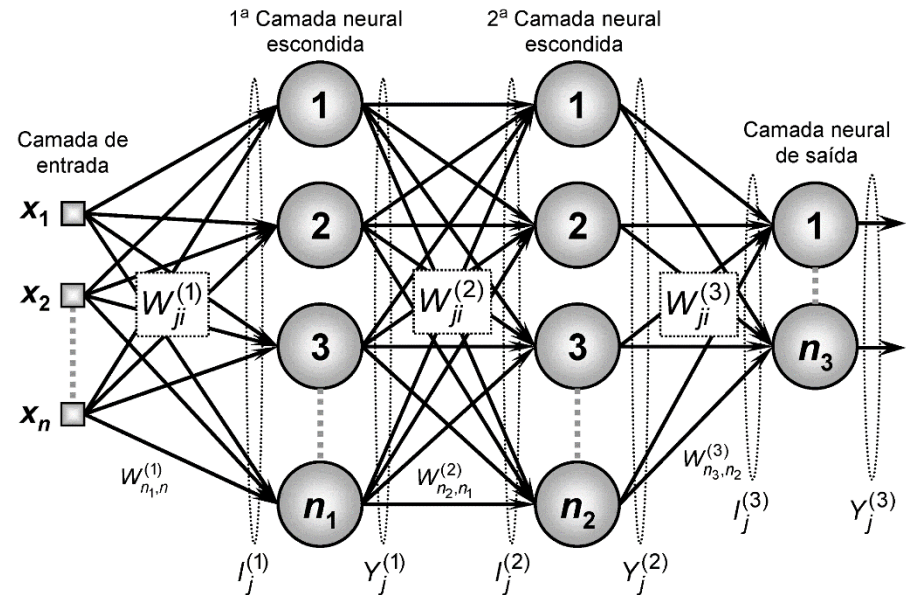
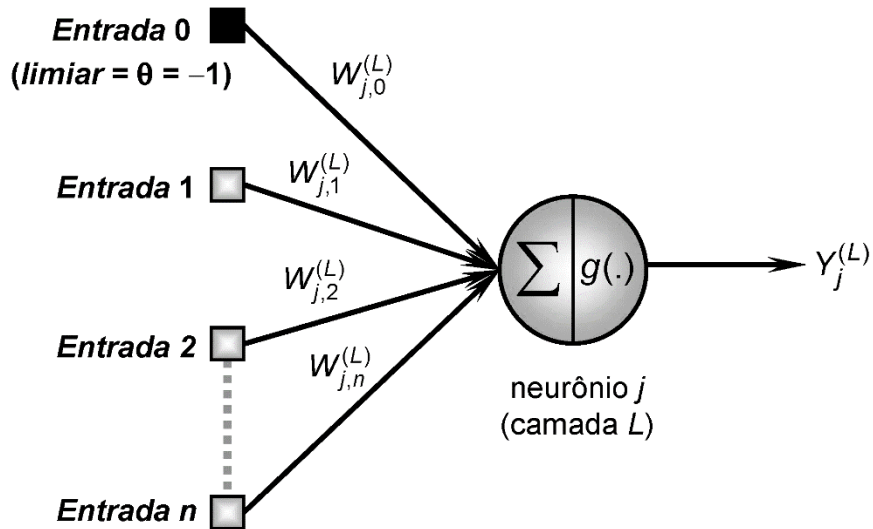


- $Y_j^{(L)}$ são vetores cujos elementos denotam a saída do j -ésimo neurônio em relação à camada L , os quais são definidos por:

$$Y_j^{(1)} = g(I_j^{(1)}) \quad (5.4)$$

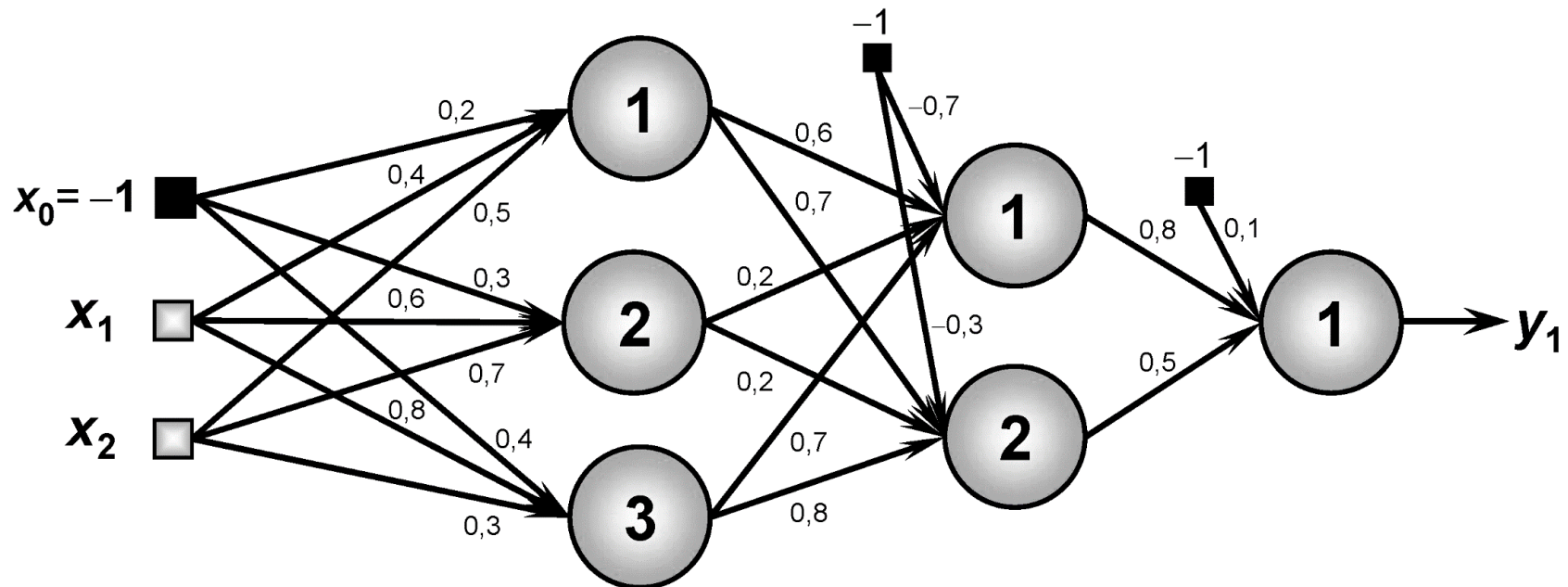
$$Y_j^{(2)} = g(I_j^{(2)}) \quad (5.5)$$

$$Y_j^{(3)} = g(I_j^{(3)}) \quad (5.6)$$

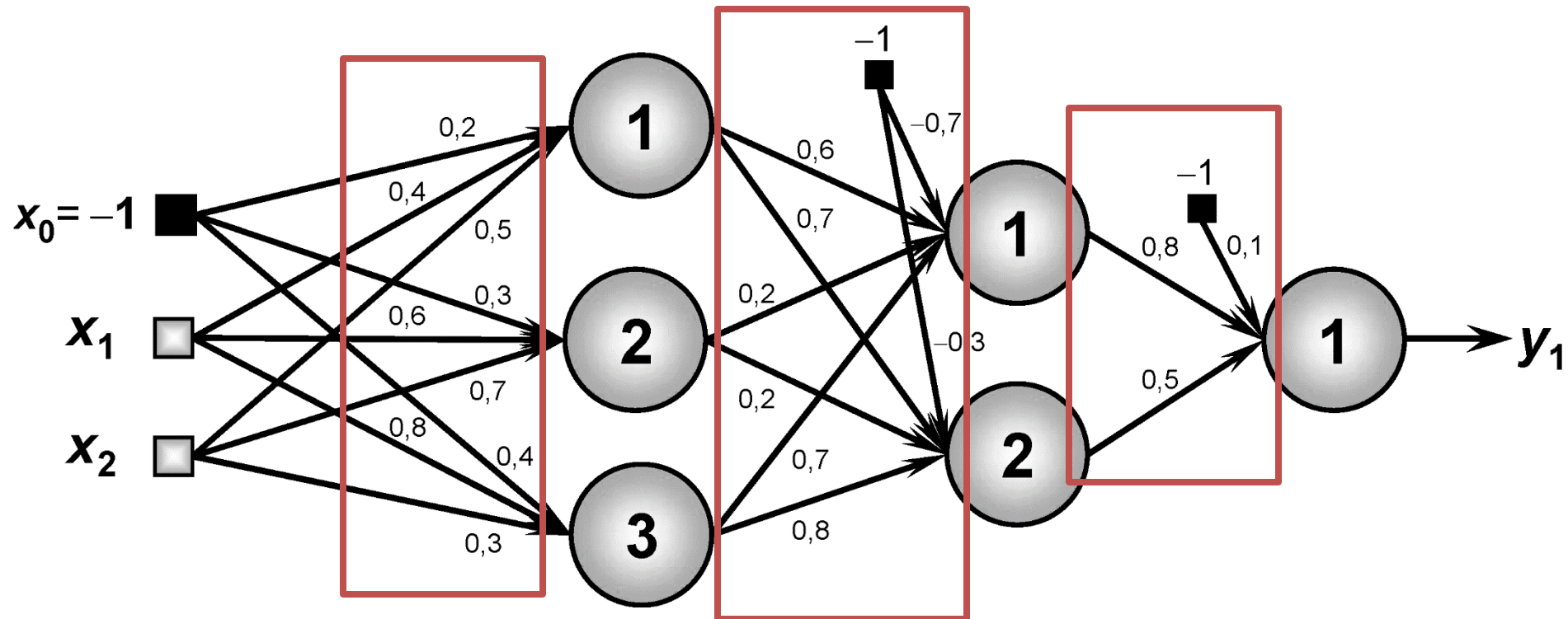


Um exemplo de topologia de uma MLP

- 2 variáveis (x_1 e x_2) + bias (x_0 : -1), ou seja, $n=2$.
- 3 neurônios na entrada/escondida n_1 , ou seja, $n_1=3$.
- 2 neurônios na escondida n_2 , ou seja, $n_2=2$.
- 1 neurônio na saída n_3 , ou seja, $n_3=1$.
- Considere que a função de ativação em todos os neurônios é a tangente hiperbólica (\tanh).

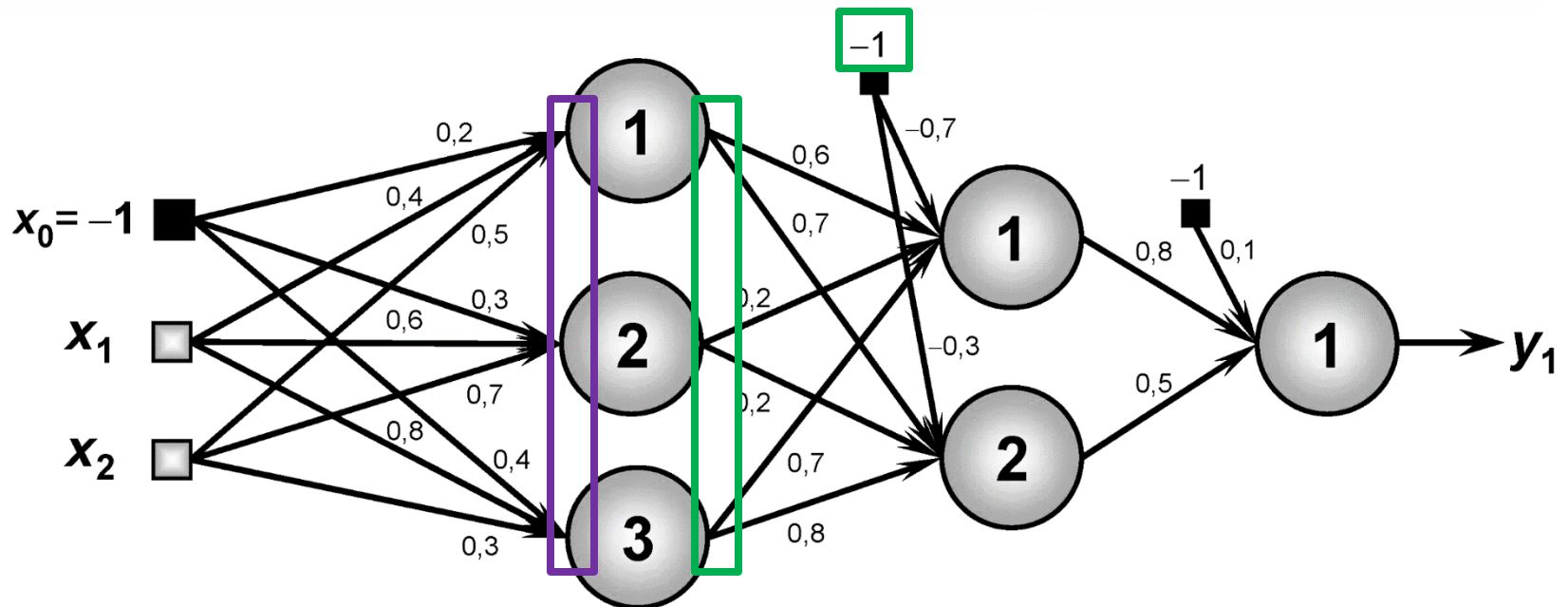


$$W_{ji}^{(1)} = \begin{bmatrix} 0,2 & 0,4 & 0,5 \\ 0,3 & 0,6 & 0,7 \\ 0,4 & 0,8 & 0,3 \end{bmatrix}; \quad W_{ji}^{(2)} = \begin{bmatrix} -0,7 & 0,6 & 0,2 & 0,7 \\ -0,3 & 0,7 & 0,2 & 0,8 \end{bmatrix}; \quad W_{ji}^{(3)} = [0,1 \quad 0,8 \quad 0,5]$$



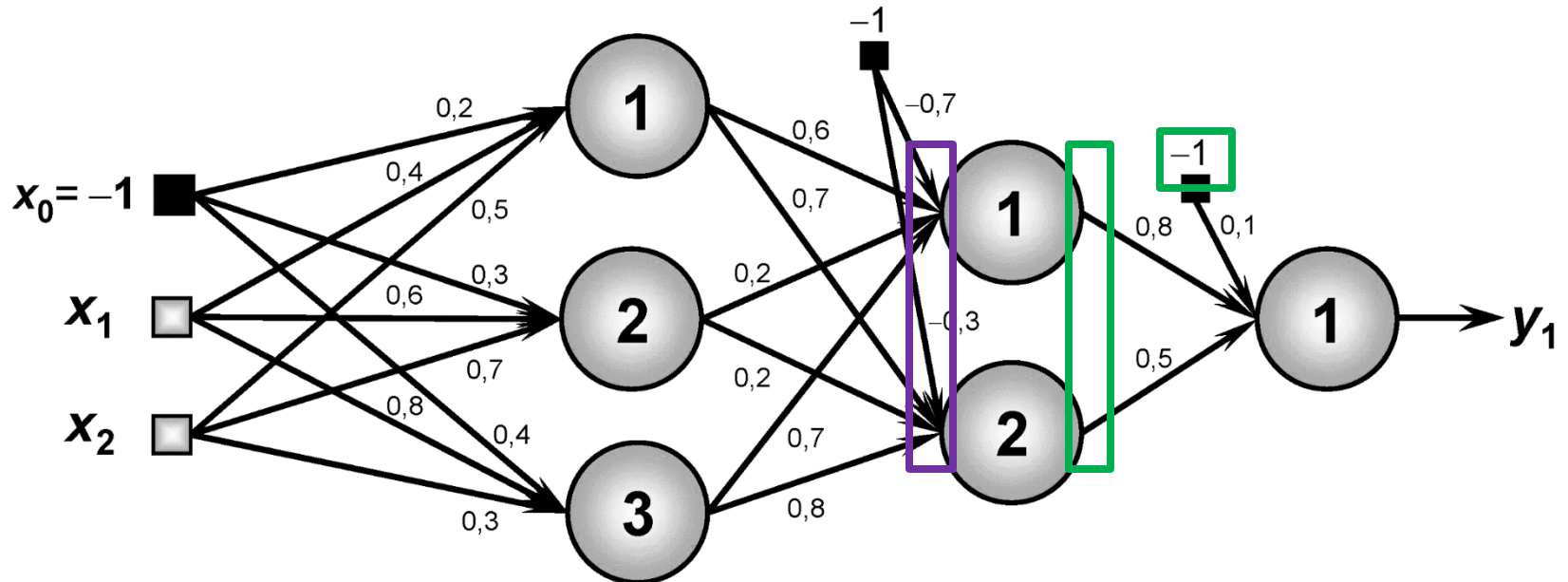
$$I_j^{(1)} = \begin{bmatrix} I_1^{(1)} \\ I_2^{(1)} \\ I_3^{(1)} \end{bmatrix} = \begin{bmatrix} W_{1,0}^{(1)} \cdot x_0 + W_{1,1}^{(1)} \cdot x_1 + W_{1,2}^{(1)} \cdot x_2 \\ W_{2,0}^{(1)} \cdot x_0 + W_{2,1}^{(1)} \cdot x_1 + W_{2,2}^{(1)} \cdot x_2 \\ W_{3,0}^{(1)} \cdot x_0 + W_{3,1}^{(1)} \cdot x_1 + W_{3,2}^{(1)} \cdot x_2 \end{bmatrix} = \begin{bmatrix} 0,2 \cdot (-1) + 0,4 \cdot 0,3 + 0,5 \cdot 0,7 \\ 0,3 \cdot (-1) + 0,6 \cdot 0,3 + 0,7 \cdot 0,7 \\ 0,4 \cdot (-1) + 0,8 \cdot 0,3 + 0,3 \cdot 0,7 \end{bmatrix} = \begin{bmatrix} 0,27 \\ 0,37 \\ 0,05 \end{bmatrix}$$

$$Y_j^{(1)} = \begin{bmatrix} Y_1^{(1)} \\ Y_2^{(1)} \\ Y_3^{(1)} \end{bmatrix} = \begin{bmatrix} g(I_1^{(1)}) \\ g(I_2^{(1)}) \\ g(I_3^{(1)}) \end{bmatrix} = \begin{bmatrix} \tanh(0,27) \\ \tanh(0,37) \\ \tanh(0,05) \end{bmatrix} = \begin{bmatrix} 0,26 \\ 0,35 \\ 0,05 \end{bmatrix} \xrightarrow{Y_0^{(1)} = -1} Y_j^{(1)} = \begin{bmatrix} Y_0^{(1)} \\ Y_1^{(1)} \\ Y_2^{(1)} \\ Y_3^{(1)} \end{bmatrix} = \begin{bmatrix} -1 \\ 0,26 \\ 0,35 \\ 0,05 \end{bmatrix}$$



$$I_j^{(2)} = \begin{bmatrix} I_1^{(2)} \\ I_2^{(2)} \end{bmatrix} = \begin{bmatrix} W_{1,0}^{(2)} \cdot Y_0^{(1)} + W_{1,1}^{(2)} \cdot Y_1^{(1)} + W_{1,2}^{(2)} \cdot Y_2^{(1)} + W_{1,3}^{(2)} \cdot Y_3^{(1)} \\ W_{2,0}^{(2)} \cdot Y_0^{(1)} + W_{2,1}^{(2)} \cdot Y_1^{(1)} + W_{2,2}^{(2)} \cdot Y_2^{(1)} + W_{2,3}^{(2)} \cdot Y_3^{(1)} \end{bmatrix} = \begin{bmatrix} 0,96 \\ 0,59 \end{bmatrix}$$

$$Y_j^{(2)} = \begin{bmatrix} Y_1^{(2)} \\ Y_2^{(2)} \end{bmatrix} = \begin{bmatrix} g(I_1^{(2)}) \\ g(I_2^{(2)}) \end{bmatrix} = \begin{bmatrix} \tanh(0,96) \\ \tanh(0,59) \end{bmatrix} = \begin{bmatrix} 0,74 \\ 0,53 \end{bmatrix} \xrightarrow{Y_0^{(2)} = -1} Y_j^{(2)} = \begin{bmatrix} Y_0^{(2)} \\ Y_1^{(2)} \\ Y_2^{(2)} \end{bmatrix} = \begin{bmatrix} -1 \\ 0,74 \\ 0,53 \end{bmatrix}$$



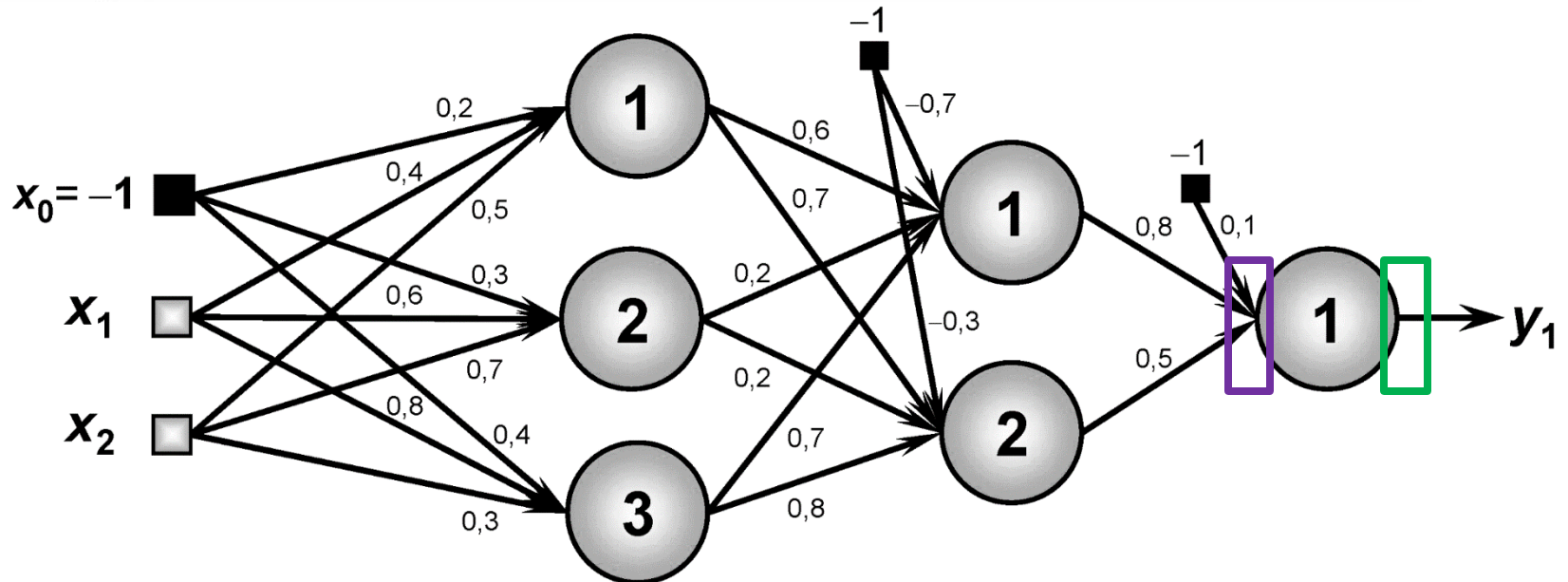
$$I_j^{(3)} = [I_1^{(3)}] = [W_{1,0}^{(3)} \cdot Y_0^{(2)} + W_{1,1}^{(3)} \cdot Y_1^{(2)} + W_{1,2}^{(3)} \cdot Y_2^{(2)}] = [0,76]$$

$$Y_j^{(3)} = [0,76] = [g(I_1^{(3)})] = [\tanh(0,76)] = [0,64]$$

Contabilização do EQM:

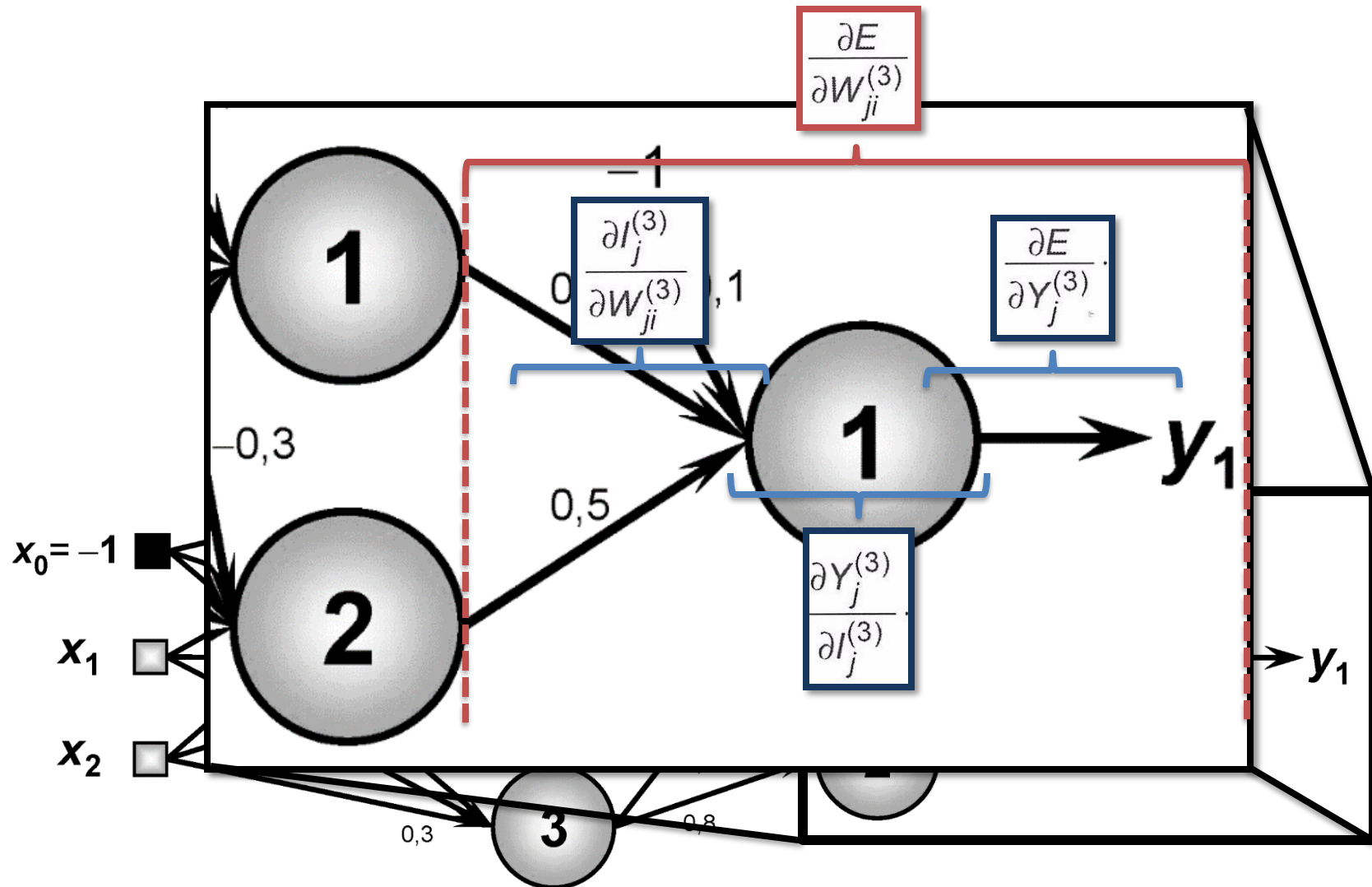
$$E(k) = \frac{1}{2} \sum_{j=1}^{n_3} (d_j(k) - Y_j^{(3)}(k))^2 \quad (5.7)$$

$$E_M = \frac{1}{p} \sum_{k=1}^p E(k) \quad (5.8)$$



Backward 1 – Ajuste dos pesos sinápticos da camada de saída

$$\nabla E^{(3)} = \frac{\partial E}{\partial W_{ji}^{(3)}} = \frac{\partial E}{\partial Y_j^{(3)}} \cdot \frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} \cdot \frac{\partial I_j^{(3)}}{\partial W_{ji}^{(3)}} \quad (5.9)$$



Backward 1 – Ajuste dos pesos sinápticos da camada de saída

$$\nabla E^{(3)} = \frac{\partial E}{\partial W_{ji}^{(3)}} = \boxed{\frac{\partial E}{\partial Y_j^{(3)}}} \cdot \frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} \cdot \frac{\partial I_j^{(3)}}{\partial W_{ji}^{(3)}} \quad (5.9)$$

Gradiente descendente

$$\frac{\partial E}{\partial Y_j^{(3)}} = -(d_j - Y_j^{(3)}) \quad \{\text{Obtido a partir de (5.7)}\} \quad (5.12)$$


$$E(k) = \frac{1}{2} \sum_{j=1}^{n_3} (d_j(k) - Y_j^{(3)}(k))^2 \quad (5.7)$$

Backward 1 – Ajuste dos pesos sinápticos da camada de saída

$$\nabla E^{(3)} = \frac{\partial E}{\partial W_{ji}^{(3)}} = \frac{\partial E}{\partial Y_j^{(3)}} \cdot \frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} \cdot \frac{\partial I_j^{(3)}}{\partial W_{ji}^{(3)}} \quad (5.9)$$

Gradiente descendente

$$\frac{\partial E}{\partial Y_j^{(3)}} = -(d_j - Y_j^{(3)}) \quad \{\text{Obtido a partir de (5.7)}\} \quad (5.12)$$

$$\hookrightarrow E(k) = \frac{1}{2} \sum_{j=1}^{n_3} (d_j(k) - Y_j^{(3)}(k))^2 \quad (5.7)$$

Derivada parcial de primeira ordem da função de ativação (função inversa)

$$\frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} = g'(I_j^{(3)}) \quad \{\text{Obtido a partir de (5.6)}\} \quad (5.11)$$

$$\hookrightarrow Y_j^{(3)} = g(I_j^{(3)}) \quad (5.6)$$

Backward 1 – Ajuste dos pesos sinápticos da camada de saída

$$\nabla E^{(3)} = \frac{\partial E}{\partial W_{ji}^{(3)}} = \frac{\partial E}{\partial Y_j^{(3)}} \cdot \frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} \boxed{\frac{\partial I_j^{(3)}}{\partial W_{ji}^{(3)}}} \quad (5.9)$$

Gradiente descendente

$$\frac{\partial E}{\partial Y_j^{(3)}} = -(d_j - Y_j^{(3)}) \quad \{\text{Obtido a partir de (5.7)}\} \quad (5.12)$$

$$\hookrightarrow E(k) = \frac{1}{2} \sum_{j=1}^{n_3} (d_j(k) - Y_j^{(3)}(k))^2 \quad (5.7)$$

Derivada parcial de primeira ordem da função de ativação (função inversa)

$$\frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} = g'(I_j^{(3)}) \quad \{\text{Obtido a partir de (5.6)}\} \quad (5.11)$$

$$\hookrightarrow Y_j^{(3)} = g(I_j^{(3)}) \quad (5.6)$$

$$\frac{\partial I_j^{(3)}}{\partial W_{ji}^{(3)}} = Y_i^{(2)} \quad \{\text{Obtido a partir de (5.3)}\} \quad (5.10)$$

$$\hookrightarrow I_j^{(3)} = \sum_{i=0}^{n_2} W_{ji}^{(3)} \cdot Y_i^{(2)} \Leftrightarrow I_j^{(3)} = W_{j,0}^{(3)} \cdot Y_0^{(2)} + W_{j,1}^{(3)} \cdot Y_1^{(2)} + \dots + W_{j,n_2}^{(3)} \cdot Y_{n_2}^{(2)} \quad (5.3)$$

Backward 1 – Ajuste dos pesos sinápticos da camada de saída

$$\nabla E^{(3)} = \frac{\partial E}{\partial W_{ji}^{(3)}} = \frac{\partial E}{\partial Y_j^{(3)}} \cdot \frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} \cdot \frac{\partial I_j^{(3)}}{\partial W_{ji}^{(3)}} \quad (5.9)$$

Calculadas as derivadas parciais basta então substituir na equação principal:

$$\frac{\partial E}{\partial W_{ji}^{(3)}} = -(d_j - Y_j^{(3)}) \cdot g'(I_j^{(3)}) \cdot Y_i^{(2)} \quad (5.13)$$

Adicionando a expressão a fórmula clássica de ajuste dos pesos sinápticos, temos que o cálculo do chamado “Gradiente Local”:

$$\Delta W_{ji}^{(3)} = -\eta \cdot \frac{\partial E}{\partial W_{ji}^{(3)}} \Leftrightarrow \Delta W_{ji}^{(3)} = \eta \cdot \delta_j^{(3)} \cdot Y_i^{(2)} \quad (5.14)$$

$$\delta_j^{(3)} = (d_j - Y_j^{(3)}) \cdot g'(I_j^{(3)}) \quad \text{DA FUNÇÃO DE ATIVAÇÃO} \quad (5.15)$$

Backward 1 – Ajuste dos pesos sinápticos da camada de saída

Representação matemática da equação de ajuste de pesos da camada de saída:

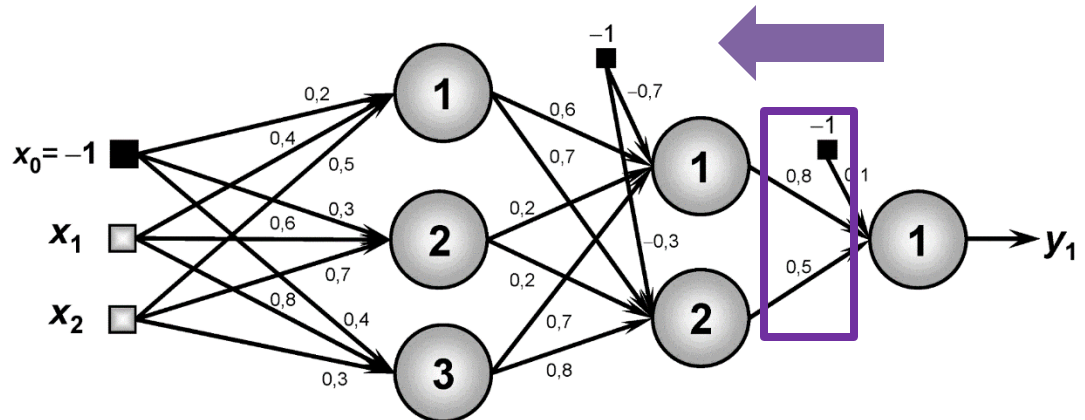
$$W_{ji}^{(3)}(t+1) = W_{ji}^{(3)}(t) + \eta \cdot \delta_j^{(3)} \cdot Y_i^{(2)} \quad (5.16)$$

Notação algorítmica da equação de ajuste de pesos da camada de saída:

$$W_{ji}^{(3)} \leftarrow W_{ji}^{(3)} + \eta \cdot \delta_j^{(3)} \cdot Y_i^{(2)} \quad (5.17)$$

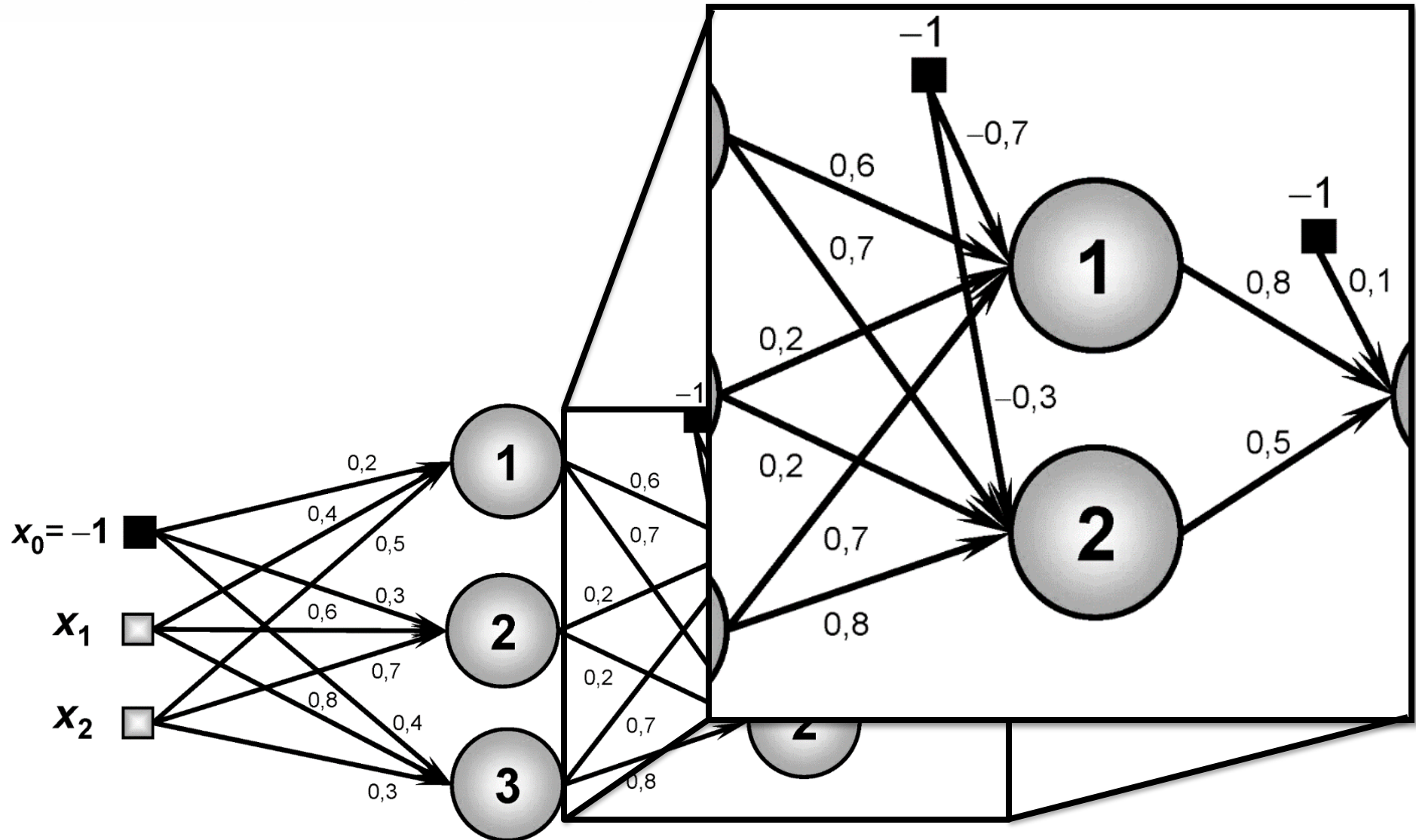
$$\delta_j^{(3)} = (d_j - Y_j^{(3)}) \cdot g'(I_j^{(3)})$$

Gradiente Local n2-n3



Backward 2 – Ajuste dos pesos sinápticos da camada intermediária n2

$$\nabla E^{(2)} = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \quad (5.18)$$



Backward 2 – Ajuste dos pesos sinápticos da camada intermediária n2

$$\nabla E^{(2)} = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \quad (5.18)$$

$$\frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} = Y_i^{(1)} \quad \{\text{Obtido a partir de (5.2)}\} \quad (5.19)$$

Backward 2 – Ajuste dos pesos sinápticos da camada intermediária n2

$$\nabla E^{(2)} = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \quad (5.18)$$

$$\frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} = Y_i^{(1)} \quad \{\text{Obtido a partir de (5.2)}\} \quad (5.19)$$

$$\frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} = g'(I_j^{(2)}) \quad \{\text{Obtido a partir de (5.5)}\} \quad (5.20)$$

Backward 2 – Ajuste dos pesos sinápticos da camada intermediária n2

$$\nabla E^{(2)} = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \quad (5.18)$$

$$\frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} = Y_i^{(1)} \quad \{\text{Obtido a partir de (5.2)}\} \quad (5.19)$$

$$\frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} = g'(I_j^{(2)}) \quad \{\text{Obtido a partir de (5.5)}\} \quad (5.20)$$

$$\frac{\partial E}{\partial Y_j^{(2)}} = \sum_{k=1}^{n_3} \frac{\partial E}{\partial I_k^{(3)}} \cdot \frac{\partial I_k^{(3)}}{\partial Y_j^{(2)}} \quad (5.21)$$

Backward 2 – Ajuste dos pesos sinápticos da camada intermediária n2

$$\nabla E^{(2)} = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \quad (5.18)$$

$$\frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} = Y_i^{(1)} \quad \{\text{Obtido a partir de (5.2)}\} \quad (5.19)$$

$$\frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} = g'(I_j^{(2)}) \quad \{\text{Obtido a partir de (5.5)}\} \quad (5.20)$$

$$\frac{\partial E}{\partial Y_j^{(2)}} = \sum_{k=1}^{n_3} \frac{\partial E}{\partial I_k^{(3)}} \cdot \frac{\partial I_k^{(3)}}{\partial Y_j^{(2)}} = \sum_{k=1}^{n_3} \underbrace{\frac{\partial E}{\partial I_k^{(3)}}}_{\text{parcela (i)}} \cdot \underbrace{\frac{\partial (\sum_{k=1}^{n_3} W_{kj}^{(3)} \cdot Y_j^{(2)})}{\partial Y_j^{(2)}}}_{\text{parcela (ii)}} \quad (5.21)$$

$$\delta_j^{(3)} = (d_j - Y_j^{(3)}) \cdot g'(I_j^{(3)})$$

$$\frac{\partial E}{\partial Y_j^{(2)}} = \sum_{k=1}^{n_3} \underbrace{\frac{\partial E}{\partial I_k^{(3)}}}_{\text{parcela (i)}} \cdot \underbrace{W_{kj}^{(3)}}_{\text{parcela (ii)}} \quad (5.22)$$

Pesos já ajustados.

Backward 2 – Ajuste dos pesos sinápticos da camada intermediária n2

$$\nabla E^{(2)} = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \quad (5.18)$$

$$\frac{\partial E}{\partial Y_j^{(2)}} = - \sum_{k=1}^{n_3} \delta_k^{(3)} \cdot W_{kj}^{(3)} \quad (5.23)$$

Backward 2 – Ajuste dos pesos sinápticos da camada intermediária n2

$$\nabla E^{(2)} = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \quad (5.18)$$

$$\frac{\partial E}{\partial Y_j^{(2)}} = - \sum_{k=1}^{n_3} \delta_k^{(3)} \cdot W_{kj}^{(3)} \quad (5.23)$$

Substituindo-se:

$$\frac{\partial E}{\partial W_{ji}^{(2)}} = - \left(\sum_{k=1}^{n_3} \delta_k^{(3)} \cdot W_{kj}^{(3)} \right) \cdot g'(I_j^{(2)}) \cdot Y_i^{(1)} \quad (5.24)$$

Backward 2 – Ajuste dos pesos sinápticos da camada intermediária n2

$$\nabla E^{(2)} = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \quad (5.18)$$

$$\frac{\partial E}{\partial Y_j^{(2)}} = - \sum_{k=1}^{n_3} \delta_k^{(3)} \cdot W_{kj}^{(3)} \quad (5.23)$$

Substituindo-se:

$$\frac{\partial E}{\partial W_{ji}^{(2)}} = - \left(\sum_{k=1}^{n_3} \delta_k^{(3)} \cdot W_{kj}^{(3)} \right) \cdot g'(I_j^{(2)}) \cdot Y_i^{(1)} \quad (5.24)$$

Na expressão de correção dos pesos:

$$\Delta W_{ji}^{(2)} = -\eta \cdot \frac{\partial E}{\partial W_{ji}^{(2)}} \Leftrightarrow \Delta W_{ji}^{(2)} = \eta \cdot \delta_j^{(2)} \cdot Y_i^{(1)} \quad (5.25)$$

Backward 2 – Ajuste dos pesos sinápticos da camada intermediária n2

Representação matemática da equação de ajuste de pesos da camada intermediária n2:

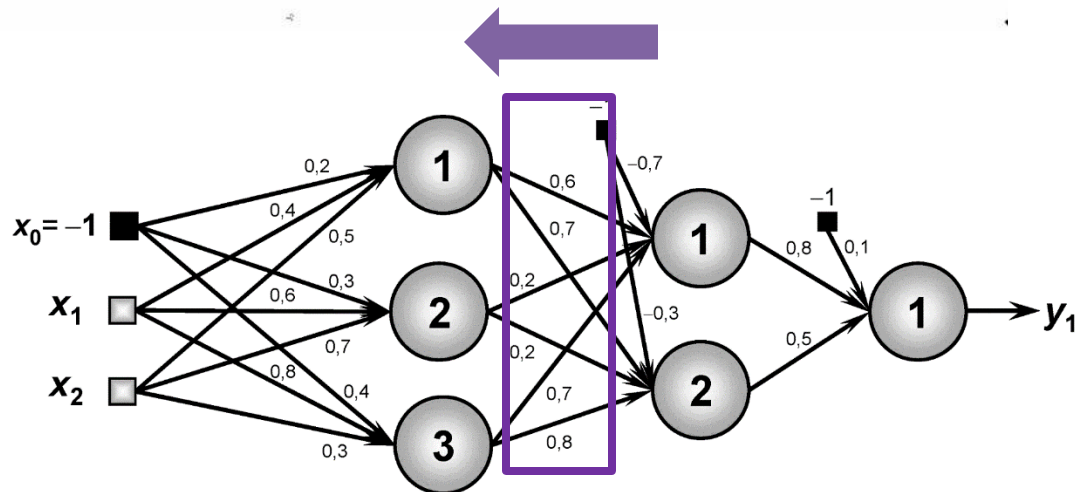
$$W_{ji}^{(2)}(t+1) = W_{ji}^{(2)}(t) + \eta \cdot \delta_j^{(2)} \cdot Y_i^{(1)} \quad (5.27)$$

Notação algorítmica da equação de ajuste de pesos da camada intermediária n2:

$$W_{ji}^{(2)} \leftarrow W_{ji}^{(2)} + \eta \cdot \delta_j^{(2)} \cdot Y_i^{(1)} \quad (5.28)$$

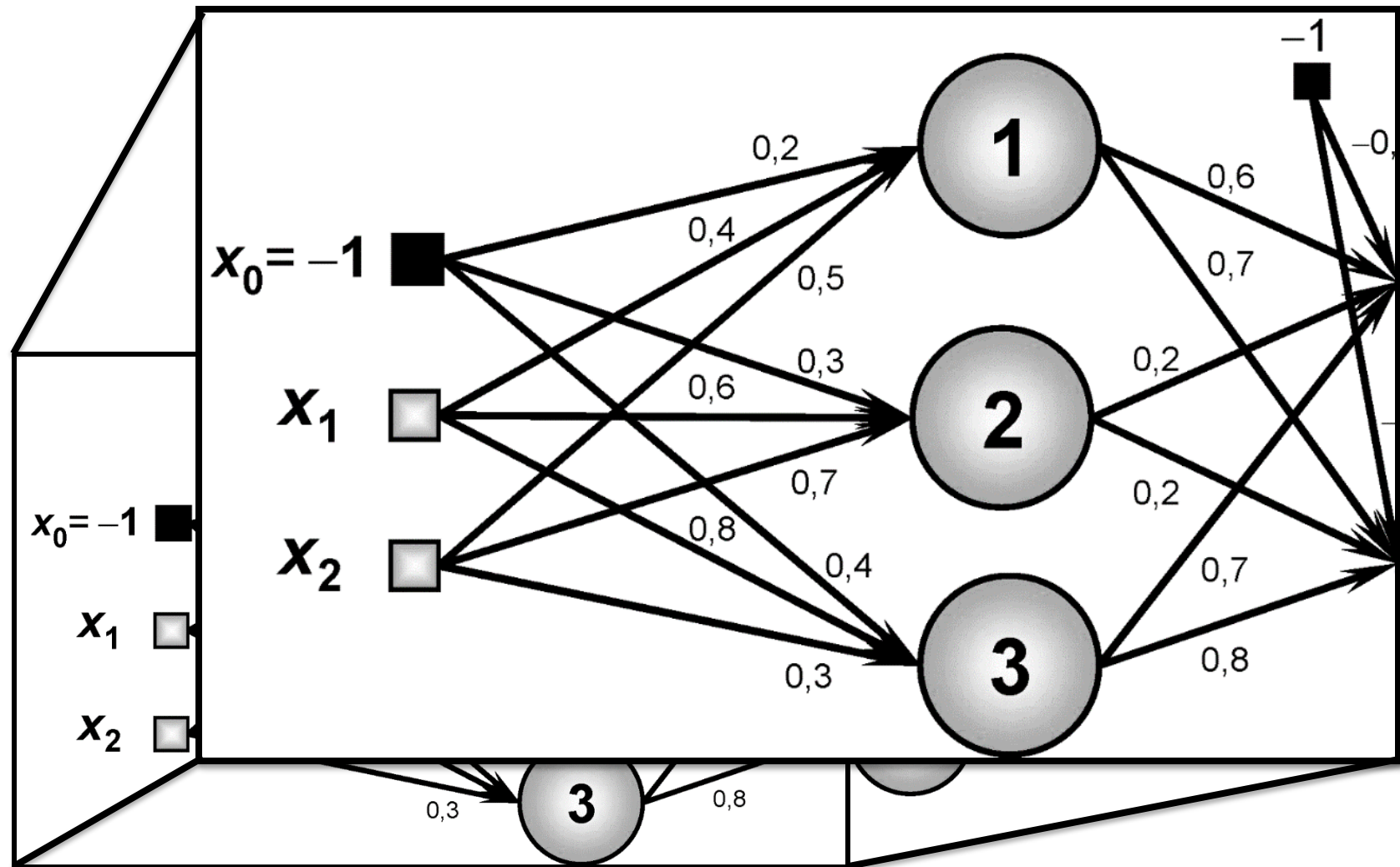
$$\delta_j^{(2)} = \left(\sum_{k=1}^{n_3} \delta_k^{(3)} \cdot W_{kj}^{(3)} \right) \cdot g'(I_j^{(2)})$$

Gradiente Local n1-n2



Backward 3 – Ajuste dos pesos sinápticos da camada intermediária n1

$$\nabla E^{(1)} = \frac{\partial E}{\partial W_{ji}^{(1)}} = \frac{\partial E}{\partial Y_j^{(1)}} \cdot \frac{\partial Y_j^{(1)}}{\partial I_j^{(1)}} \cdot \frac{\partial I_j^{(1)}}{\partial W_{ji}^{(1)}} \quad (5.29)$$



Backward 3 – Ajuste dos pesos sinápticos da camada intermediária n1

$$\frac{\partial I_j^{(1)}}{\partial W_{ji}^{(1)}} = x_i \quad \{\text{Obtido a partir de (5.1)}\} \quad (5.30)$$

$$\frac{\partial Y_j^{(1)}}{\partial I_j^{(1)}} = g'(I_j^{(1)}) \quad \{\text{Obtido a partir de (5.4)}\} \quad (5.31)$$

$$\frac{\partial E}{\partial Y_j^{(1)}} = \sum_{k=1}^{n_2} \underbrace{\frac{\partial E}{\partial I_k^{(2)}}}_{\text{parcela (i)}} \cdot \frac{\partial I_k^{(2)}}{\partial Y_j^{(1)}} = \sum_{k=1}^{n_2} \underbrace{\frac{\partial E}{\partial I_k^{(2)}}}_{\text{parcela (i)}} \cdot \underbrace{\frac{\partial (\sum_{k=1}^{n_2} W_{kj}^{(2)} \cdot Y_j^{(1)})}{\partial Y_j^{(1)}}}_{\text{parcela (ii)}} \quad (5.32)$$

$$\frac{\partial E}{\partial Y_j^{(1)}} = \sum_{k=1}^{n_2} \underbrace{\frac{\partial E}{\partial I_k^{(2)}}}_{\text{parcela (i)}} \cdot \underbrace{W_{kj}^{(2)}}_{\text{parcela (ii)}} \quad (5.33)$$

Backward 3 – Ajuste dos pesos sinápticos da camada intermediária n1

$$\frac{\partial E}{\partial Y_j^{(1)}} = - \sum_{k=1}^{n_2} \delta_k^{(2)} \cdot W_{kj}^{(2)} \quad (5.34)$$

$$\frac{\partial E}{\partial W_{ji}^{(1)}} = - \left(\sum_{k=1}^{n_2} \delta_k^{(2)} \cdot W_{kj}^{(2)} \right) \cdot g'(I_j^{(1)}) \cdot x_i \quad (5.35)$$

$$\Delta W_{ji}^{(1)} = -\eta \cdot \frac{\partial E}{\partial W_{ji}^{(1)}} \Leftrightarrow \Delta W_{ji}^{(1)} = \eta \cdot \delta_j^{(1)} \cdot x_i \quad (5.36)$$

Backward 3 – Ajuste dos pesos sinápticos da camada intermediária n1

Representação matemática da equação de ajuste de pesos da camada intermediária n1:

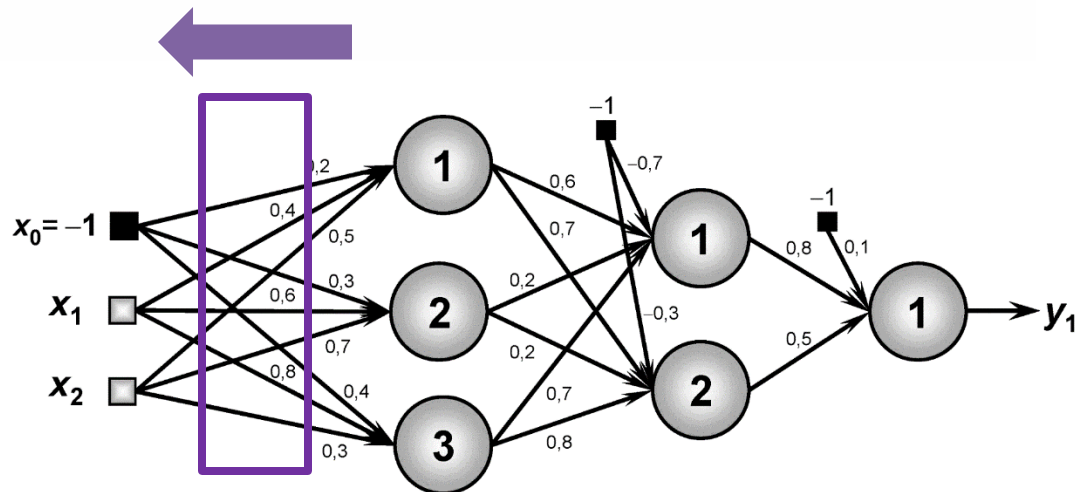
$$W_{ji}^{(1)}(t+1) = W_{ji}^{(1)}(t) + \eta \cdot \delta_j^{(1)} \cdot x_i \quad (5.38)$$

Notação algorítmica da equação de ajuste de pesos da camada intermediária n1:

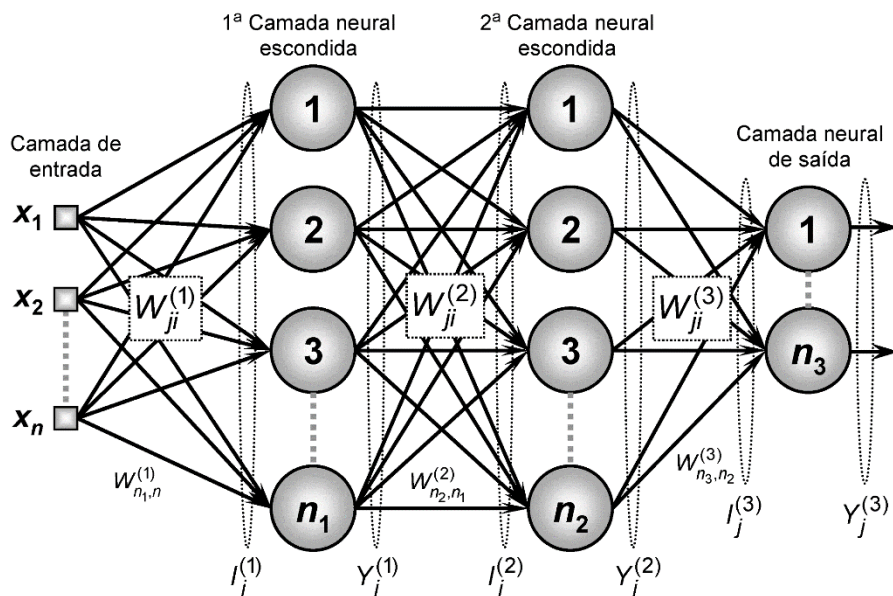
$$W_{ji}^{(1)} \leftarrow W_{ji}^{(1)} + \eta \cdot \delta_j^{(1)} \cdot x_i \quad (5.39)$$

$$\delta_j^{(1)} = \left(\sum_{k=1}^{n_2} \delta_k^{(2)} \cdot W_{kj}^{(2)} \right) \cdot g'(I_j^{(1)})$$

Gradiente Local x-n1



Algoritmo de treinamento da MLP



Início {Algoritmo PMC – Fase de Treinamento}

- <1> Obter o conjunto de amostras de treinamento $\{x^{(k)}\}$;
 - <2> Associar o vetor de saída desejada $\{d^{(k)}\}$ para cada amostra;
 - <3> Iniciar $W_{ji}^{(1)}$, $W_{ji}^{(2)}$ e $W_{ji}^{(3)}$ com valores aleatórios pequenos;
 - <4> Especificar taxa de aprendizagem $\{\eta\}$ e precisão requerida $\{\varepsilon\}$;
 - <5> Iniciar o contador de número de épocas $\{\text{época} \leftarrow 0\}$;
 - <6> Repetir as instruções:
 - <6.1> $E_M^{\text{anterior}} \leftarrow E_M$; {conforme (5.8)}
 - <6.2> Para todas as amostras de treinamento $\{x^{(k)}, d^{(k)}\}$, fazer:
 - <6.2.1> Obter $I_j^{(1)}$ e $Y_j^{(1)}$; {conforme (5.1) e (5.4)}
 - <6.2.2> Obter $I_j^{(2)}$ e $Y_j^{(2)}$; {conforme (5.2) e (5.5)}
 - <6.2.3> Obter $I_j^{(3)}$ e $Y_j^{(3)}$; {conforme (5.3) e (5.6)}
 - <6.2.4> Determinar $\delta_j^{(3)}$; {conforme (5.15)}
 - <6.2.5> Ajustar $W_{ji}^{(3)}$; {conforme (5.17)}
 - <6.2.6> Determinar $\delta_j^{(2)}$; {conforme (5.26)}
 - <6.2.7> Ajustar $W_{ji}^{(2)}$; {conforme (5.28)}
 - <6.2.8> Determinar $\delta_j^{(1)}$; {conforme (5.37)}
 - <6.2.9> Ajustar $W_{ji}^{(1)}$; {conforme (5.39)}
 - <6.3> Obter $Y_j^{(3)}$ ajustado; {conforme <6.2.1>, <6.2.2> e <6.2.3>}
 - <6.4> $E_M^{\text{atual}} \leftarrow E_M$; {conforme (5.8)}
 - <6.5> $\text{época} \leftarrow \text{época} + 1$;
- Até que: $|E_M^{\text{atual}} - E_M^{\text{anterior}}| \leq \varepsilon$

Fim {Algoritmo PMC – Fase de Treinamento}

Algoritmo de previsão da MLP

Início {Algoritmo PMC – Fase de Operação}

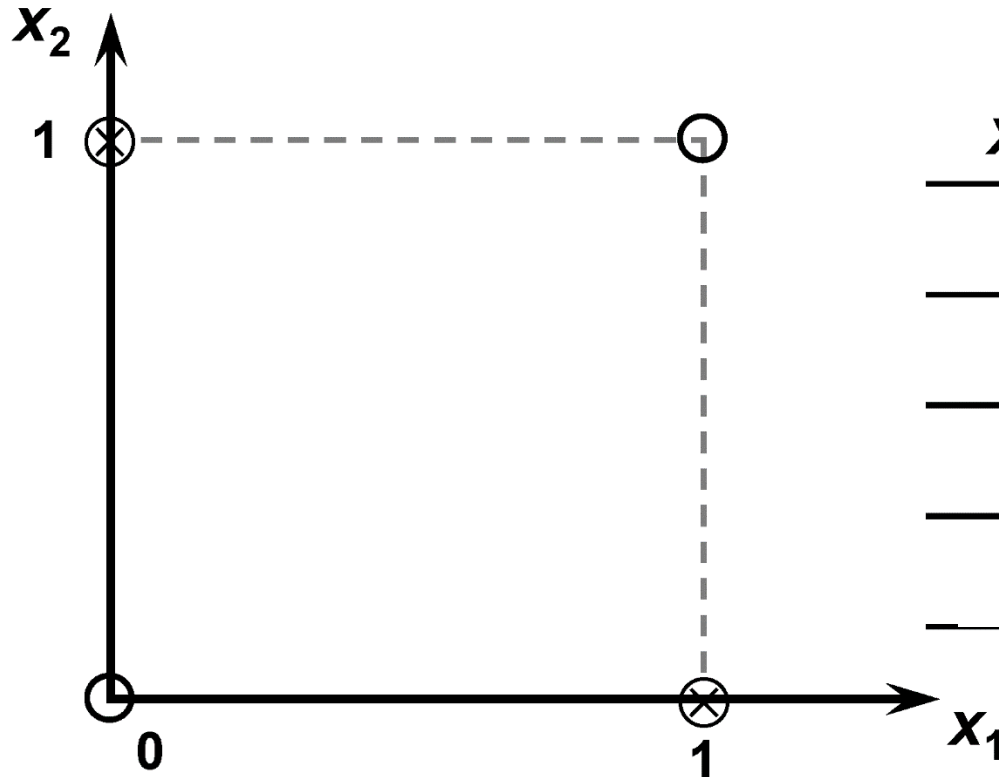
- <1> Obter uma amostra $\{ \mathbf{x} \}$;
- <2> Assumir $W_{ji}^{(1)}$, $W_{ji}^{(2)}$ e $W_{ji}^{(3)}$ já ajustadas no treinamento;
- <3> Execute as seguintes instruções:
 - <3.1> Obter $I_j^{(1)}$ e $Y_j^{(1)}$; {conforme (5.1) e (5.4)}
 - <3.2> Obter $I_j^{(2)}$ e $Y_j^{(2)}$; {conforme (5.2) e (5.5)}
 - <3.3> Obter $I_j^{(3)}$ e $Y_j^{(3)}$; {conforme (5.3) e (5.6)}
- <4> Disponibilizar as saídas da rede, as quais são dadas pelos elementos contidos em $Y_j^{(3)}$

Passo
Forward

Fim {Algoritmo PMC – Fase de Operação}

Exemplo

MLP aplicada ao problema de separação XOR



x_1	x_2	X_{OR}
0	0	$0 \rightarrow O$
0	1	$1 \rightarrow X$
1	0	$1 \rightarrow X$
1	1	$0 \rightarrow O$

Atividade

Topologia de MLP a ser implementada

