Universidade de São Paulo
Escola de Engenharia de São Carlos
Departamento de Engenharia Elétrica

Alunos: Luisa Helena Bartocci Liboni
Rodrigo de Toledo Caropreso

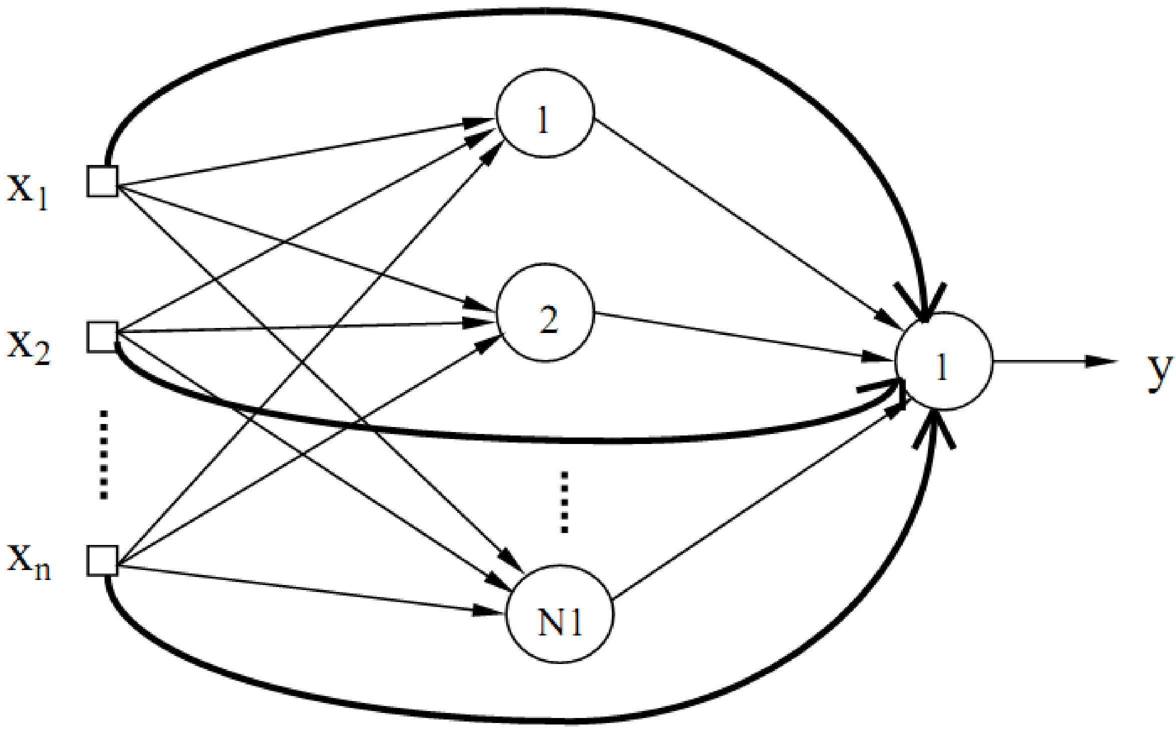
Data de Entrega: 07/05/2012

Redes Neurais Artificiais

(Prof. Ivan Nunes da Silva)

EPC-4

Visando uma aplicação específica, a equipe de engenheiros e cientistas de uma instituição projetou uma rede neural artificial do tipo perceptron multicamadas, constituída de três camadas, cuja topologia está ilustrada na figura seguinte.



As informações referentes à topologia da rede estão como se segue:

- Camada de Entrada → Constituída de "n" sinais de entrada (x_1, x_2, ..., x_n).
- Camada Neural Escondida → Constituída de "N1" neurônios.
- Camada Neural de Saída → Constituída de apenas 1 neurônio.
- Conjunto de Treinamento → Constituído de "p" amostras.

Demonstre de forma detalhada como será a seqüência e o ajuste das matrizes de pesos entre cada uma das camadas quando se utiliza o algoritmo de aprendizagem *Backpropagation*.

Utilize a seguinte convenção:

- Matriz W1 → Matriz de pesos entre a 1ª e 2ª camada.
- Matriz W2 → Matriz de pesos entre a 2ª e 3ª camada.
- Matriz W3 → Matriz de pesos entre a 1ª e 3ª camada.

OBSERVAÇÕES:

1. O EPC pode ser realizado em grupo de três pessoas. Se for o caso, entregar somente um EPC com o nome de todos integrantes.
2. As folhas contendo os resultados do EPC devem ser entregue em sequência e grampeadas (não use clips).
3. Em se tratando de EPC que contenha implementação computacional, anexe (de forma impressa) o programa fonte referente ao mesmo.

RESOLUÇÃO:

A resolução desta atividade faz referência às expressões e equações apresentadas no Capítulo 5 do livro:

SILVA, I. N., SPATTI, D. H., FLAUZINO, R.A. - **"Redes Neurais Artificiais para Engenharia e Ciências Aplicadas"** - 1ª Ed. 2010 - Editora Artliber, São Paulo- SP.

Cada um dos neurônios {j} pertencentes a uma das camadas {L} da totpologia mostrada na figura acima, pode ser configurado conforme a terminologia adotada na figura 5.4 (SILVA, et al), onde g(.) representa uma função de ativação.

Assim, temos:

- $W_{ji}^{(L)}$ 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).
- $W_{ji}^{(L,M)}$ 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 M.
- $I_{ji}^{(L)}$ são vetores cujos elementos denotam a entrada ponderada do j-ésimo neurônio da camada L.

- O termos $Y_j^{(L)}$ são vetores cujos elementos denotam a saída do j-éismo neurônio em relação a camada L.

O passo *Forward* pode ser descrito pelas seguintes equações, análogas às expressões apresentadas (SILVA, et al). Aqui, porém, a diferença na topologia fica evidenciada pelas conexões sinápticas entre as entradas e a camada de saída, o que irá compor um elemento a mais na descrição desta camada:

Camada (Oculta)
$$I_j^{(1)} = \sum_{i=0}^{n} W_{ji}^{(1)} \cdot x_i \tag{1}$$

Camada de Saída
$$I_j^{(2)} = \sum_{i=0}^{n_1} W_{ji}^{(2)} \cdot Y_i^{(1)} + \sum_{i=0}^{n} W_{ji}^{(2,0)} \cdot x_i \tag{2}$$

Saídas dos neurônios
$$Y_j^{(1)} = g\left(I_j^{(1)}\right) \tag{3}$$
$$Y_j^{(2)} = g\left(I_j^{(2)}\right) \tag{4}$$

O erro quadrático, utilizado no algoritmo Backpropagation para definir o erro de aproximação que mede os desvios entre as respostas produzidas pelos neurônios da rede em relação aos respectivos valores desejados (das amostras), pode ser dado por:

$$E(k) = \frac{1}{2} \cdot \sum_{j=1}^{n_2} \left(d_j(k) - Y_j^{(2)}\right)^2 \tag{5}$$

Como $n_2 = 1$, vem:

$$E(k) = \frac{\left(d_j(k) - Y_j^{(2)}\right)^2}{2} \tag{6}$$

Assim, define-se o <u>Erro Quadrático Médio</u> para p amostras, por:

$$E_M = \frac{1}{p} \cdot \sum_{k=1}^{p} E(k) \tag{7}$$

Para o passo *Backward*, deve-se dividir o cálculo em 2 situações:

Ajuste dos pesos da camada de saída:

A partir da utilização do gradiente e da regra de diferenciação em cadeia, apresentadas em (SILVA et al), temos:

$$\nabla E = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \quad (7)$$

onde o termo $\frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}}$ é dado por $\frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} + \frac{\partial I_j^{(2,0)}}{\partial W_{ji}^{(2,0)}}$

$$\frac{\partial E}{\partial Y_j^{(2)}} = -\left(d_j - Y_j^{(2)}\right) \quad (8)$$

$$\frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} = g'\left(I_j^{(2)}\right) \quad (9)$$

$$\frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} = Y_i^{(1)} \quad (10)$$

$$\frac{\partial I_j^{(2,0)}}{\partial W_{ji}^{(2,0)}} = x_i \quad (11)$$

$$\nabla E = -\left(d_j - Y_j^{(2)}\right) \cdot g'\left(I_j^{(2)}\right) \cdot \left(Y_i^{(1)} + x_i\right) \quad (12)$$

O ajuste deve ser feito em direção oposta ao gradiente. Portanto, o ajuste deve ser definido de acordo com as expressões abaixo:

$$W_{ji}^{(2)} \leftarrow W_{ji}^{(2)} + \eta \cdot \left(d_j - Y_j^{(2)}\right) \cdot g'\left(I_j^{(2)}\right) \cdot Y_i^{(1)} \quad (13)$$

$$W_{ji}^{(2,0)} \leftarrow W_{ji}^{(2,0)} + \eta \cdot \left(d_j - Y_j^{(2)}\right) \cdot g'\left(I_j^{(2)}\right) \cdot x_i \quad (14)$$

Ajuste dos pesos da camada oculta:

As expressões usadas são análogas às (5.29) a (5.39) de (SILVA et. al):

$$\frac{\partial E}{\partial W_{ji}^{(1)}} = -\left(\sum_{k=1}^{n_2} \delta_k^{(2)} \cdot W_{kj}^{(2)}\right) \cdot g'\left(I_j^{(1)}\right) \cdot x_i \quad \text{, onde} \quad (15)$$

$$\delta_k^{(2)} = \left(d_j - Y_j^{(2)}\right) \cdot g'\left(I_j^{(2)}\right) \quad (16)$$

O ajuste é dado por:

$$W_{ji}^{(1)} \leftarrow W_{ji}^{(1)} + \eta \cdot \delta_j^{(1)} \cdot x_i \quad \text{, onde} \quad (17)$$

$$\delta_j^{(1)} = \left(\sum_{k=1}^{n_2} \delta_k^{(2)} \cdot W_{kj}^{(2)}\right) \cdot g'\left(I_j^{(1)}\right) \quad (18)$$