

CPE 721 – RNs Feedforward

2ª Série de Exercícios – Treinamento

Obs: O objetivo da série de exercícios é a fixação do aprendizado. A série pode ser feita em grupo, mas é importante que cada um tente achar as soluções individualmente antes do trabalho em grupo.

Treinamento como um processo de otimização

1 - Deseja-se ajustar o vetor de parâmetros \vec{w} de um mapeador para P pares entrada – saída (\vec{x}^p, \vec{y}^p) , $p = 1, \dots, P$. O mapeador tem a seguinte transferência:

$$\tilde{y}_1 = w_1 e^{-w_2 x_1^2} + w_3 \ln(x_2 + 1)$$

$$\tilde{y}_2 = w_3 x_1^2 + w_4$$

e os parâmetros w_i devem ser ajustados para minimizar o erro F

$$F(\vec{w}) = E \sum_{p=1}^P \left\{ 2 \left(\frac{y_1^p - \tilde{y}_1^p}{y_1^p + 1} \right)^2 + 7 (y_2^p - \tilde{y}_2^p)^2 \right\}$$

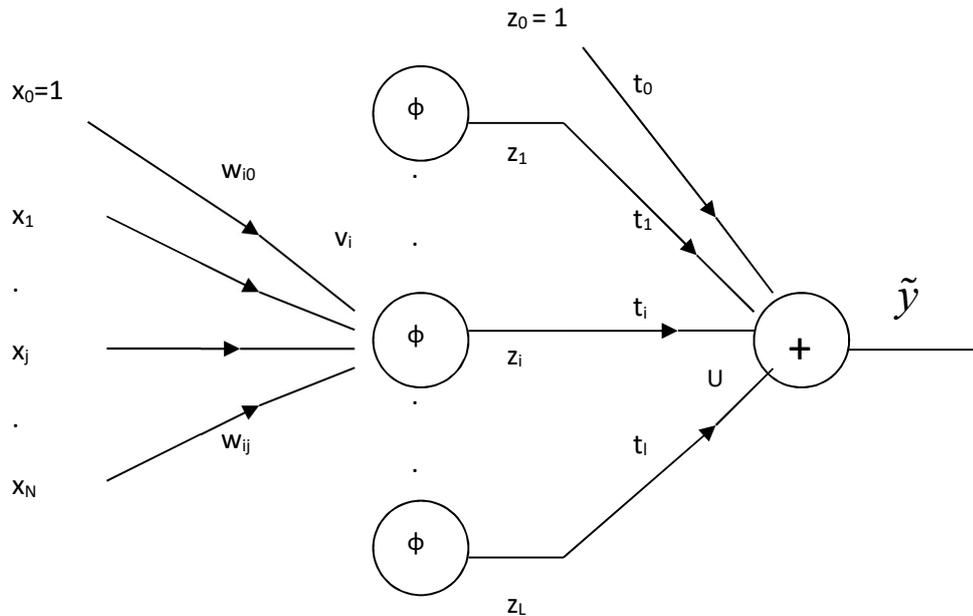
Calcule a expressão dos acréscimos Δw_i , $i = 1, \dots, 4$ que devem ser aplicados para executar a minimização por regra delta.

1.1 – Para o mapeador acima a partir do ponto $\underline{w} = [1, 1, 1, 1]^t$ calcule os deslocamentos Δw_2 e Δw_3 para o lote de 4 pares entrada saída $(\underline{x}, \underline{y})$ a seguir: $([0, 0]^t, [0, 1]^t)$, $([1, 1]^t, [0.83, 2.2]^t)$, $([0, 1]^t, [0.83, 1]^t)$, $([1, 0]^t, [0, 2.2]^t)$. Considere $\alpha = .1$

2 - Considere a rede de duas camadas abaixo. A camada de saída tem um neurônio linear e a camada intermediária com neurônios com a função de ativação bi-hiperbólica ϕ definida abaixo (*Learning and Non-linear Models*, Vol. I, pp. 276-282., Rio de Janeiro, Brasil)

$$\phi(v, \lambda, \tau_1, \tau_2) = \sqrt{\lambda^2 \left(v + \frac{1}{4\lambda} \right)^2 + \tau_1^2} - \sqrt{\lambda^2 \left(v - \frac{1}{4\lambda} \right)^2 + \tau_2^2} + \frac{1}{2} \quad \text{onde} \quad v = \sum_0^n w_i x_i$$

Estabeleça o algoritmo de treinamento para minimizar o emq na saída para todos os pares entrada – saída. Estabeleça os acréscimos a serem aplicados nos parâmetros $w_{ij}, t_i, v_i, \lambda_i, \tau_{1i}, \tau_{2i}$ de cada neurônio da camada intermediária e da saída.



3 - Considere a otimização via gradiente descendente utilizando um conjunto de P pares entrada-saída por batelada ou por regra delta. Escolhendo α *suficientemente pequeno* o ponto de operação após a aplicação dos P pares por regra delta ou por batelada é praticamente o mesmo. Mostre que é válido para a aplicação de dois pares consecutivos, e discuta o que significa *suficientemente* pequeno considerando os valores da Hessiana e do gradiente no ponto.

4 - Em um processo de aprendizagem por regra delta com momento propõe-se aplicar no passo n o acréscimo

$$\Delta w_i(n) = \frac{1}{n} \left\{ (n-1) \Delta w_i(n-1) + \Delta w_i^D(n) \right\} = \Delta w_i(n-1) + \frac{1}{n} \left\{ \Delta w_i^D(n) - \Delta w_i(n-1) \right\}$$

onde $\Delta w_i^D(n)$ é o acréscimo calculado por regra delta no passo n. Explique o que representa o acréscimo $\Delta w_i(n)$ que esta sendo aplicado no passo n em termos dos acréscimos regra delta dos passos anteriores.