

CPE 721 – RNs Feedforward

1ª Série de Exercícios – Otimização e 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.

Otimização.

1 - Considere a função

$$F(\vec{w}) = w_1^3 w_2^2 - 2w_1^3 w_2 - 3w_1 w_2^2 + 6w_1 w_2$$

1.1 – Calcule a expressão analítica do gradiente e da Hessiana.

1.2 – Calcule o valor da função, do gradiente e da Hessiana nos quatro pontos indicados à seguir e interprete os resultados em termos de singularidades e extremos da função. Pontos (w_1, w_2) : (0,0), (1,1), (-1, 1), (1, -1)

1.3 – Escreva a expressão analítica da curva de nível que passa pelo ponto (1, -1).

1.4 – Calcule a expressão da tangente a curva de nível no ponto (1, -1). Sugestão: lembre das propriedades do gradiente.

1.5 – Supondo que a aproximação quadrática é válida e partindo do ponto (1, -1), calcule o extremo usando Newton Raphson

$$\vec{w}_{extremo} = \vec{w}_0 - \vec{H}^{-1}(\vec{w}_0) \nabla(\vec{w}_0)$$

Pela análise do gradiente verifique se o ponto encontrado é mesmo um extremo. Comente.

1.6 - Repita o item 1.5 a partir da origem (0, 0). Comente.

1.7 – Novamente a partir do ponto (1, -1) calcule e aplique o passo ótimo para minimização (sem usar a Hessiana, usando otimização em linha) via gradiente descendente. Analise o ponto encontrado.

1.8 - Calcule o valor da aproximação da Hessiana proposta por Levenberg-Marquadt para estes os pontos do item 1.2 e compare com o resultado exato.

2 – Considerando que a aproximação quadrática é válida mostre que para que uma função objetivo diminua seu valor é necessário que o passo seja menor que o dobro do passo ótimo, $\alpha < 2\alpha_{\text{ótimo}}$.

3 - No método BP resiliente se a aproximação de segunda ordem da função objetivo for válida existe uma fórmula para calcular o valor de α_i que conduz ao minimante em um único passo se a derivada não mudou de sinal. Apresente a fórmula.

Treinamento como um processo de otimização

4 - 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 ajustado para minimizar o erro F

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

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

4.1 – Para o mapeador da questão 4 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: $([0, 0]^t, [0, 1]^t)$, $([1, 1]^t, [.83, 2.2]^t)$, $([0, 1]^t, [.83, 1]^t)$, $([1, 0]^t, [0, 2.2]^t)$. Considere $\alpha = .1$

5 - 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.

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

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

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.