

CPE 721 – 2017 - RNs Feedforward

1ª Série de Exercícios – Otimização

1 - Considere a função

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

Calcule a expressão analítica do gradiente e da Hessiana.

$$\nabla = \begin{bmatrix} \frac{\partial F}{\partial w_1} \\ \frac{\partial F}{\partial w_2} \end{bmatrix} = \begin{bmatrix} 3w_1^2 w_2^2 - 6w_1^2 w_2 - 3w_2^2 + 6w_2 \\ 2w_1^3 w_2 - 2w_1^3 - 6w_1 w_2 + 6w_1 \end{bmatrix}$$

$$H = \begin{bmatrix} \frac{\partial^2 F}{\partial w_1^2} & \frac{\partial^2 F}{\partial w_1 \partial w_2} \\ \frac{\partial^2 F}{\partial w_1 \partial w_2} & \frac{\partial^2 F}{\partial w_2^2} \end{bmatrix} = \begin{bmatrix} 6w_1 w_2^2 - 12w_1 w_2 & 6w_1^2 w_2 - 6w_1^2 - 6w_2 + 6 \\ 6w_1^2 w_2 - 6w_1^2 - 6w_2 + 6 & 2w_1^3 - 6w_1 \end{bmatrix}$$

1.1 – 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)

$$\nabla(0,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\nabla(1,1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\nabla(-1,1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\nabla(1,-1) = \begin{bmatrix} 0 \\ 8 \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 0 & 6 \\ 6 & 0 \end{bmatrix}$$

$$H(1,1) = \begin{bmatrix} -6 & 0 \\ 0 & -4 \end{bmatrix}$$

$$H(-1,1) = \begin{bmatrix} 6 & 0 \\ 0 & 4 \end{bmatrix}$$

$$H(1,-1) = \begin{bmatrix} 18 & 0 \\ 0 & -4 \end{bmatrix}$$

H não definida
Sela

H def negativa
Máximo

H def positiva
Mínimo

gradiente não nulo
não extremo

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

$$F(1,-1) = w_1^3 w_2^2 - 2w_1^3 w_2 - 3w_1 w_2^2 + 6w_1 w_2 = -6$$

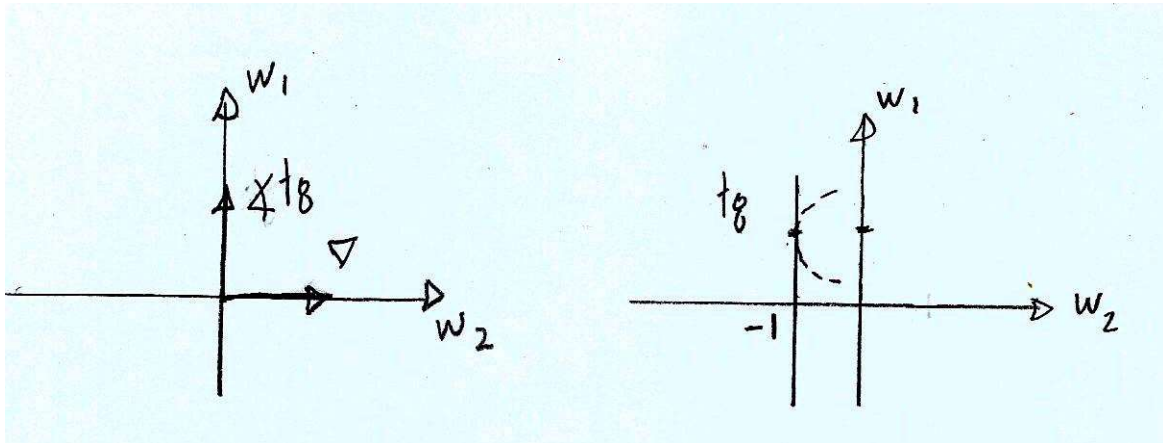
$$\text{Curva de nível: } w_1^3 w_2^2 - 2w_1^3 w_2 - 3w_1 w_2^2 + 6w_1 w_2 = -6$$

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

Tg perpendicular ao gradiente (relação entre os w)

$$\underline{w}' \underline{\nabla}(1, -1) = 0 \quad \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} 0 \\ 8 \end{bmatrix} = 0w_1 + 8w_2 = 0 \quad w_2 = 0 \quad w_1 \text{ qualquer}$$

Passando no ponto (1, -1) $w_2 = 0 \quad w_1 \text{ qualquer} \Rightarrow w_2 = -1 \quad \text{ou} \quad w_2 + 1 = 0$



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

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

$$\vec{w}_{extremo} = \vec{w}_0 - \vec{H}^{-1}(\vec{w}_0) \nabla(\vec{w}_0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} \frac{1}{18} & 0 \\ 0 & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} 0 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\nabla(1,1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad H(1,1) = \begin{bmatrix} -6 & 0 \\ 0 & -4 \end{bmatrix}$$

Gradiente nulo, Hessiana def negativa >>> (1,1) é um extremo, um máximo.

1.5 - Repita o item 1.4 a partir da origem (0, 0). Comente.

$$\vec{w}_{extremo} = \vec{w}_0 - \vec{H}^{-1}(\vec{w}_0) \nabla(\vec{w}_0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 & \frac{1}{6} \\ \frac{1}{6} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Já está em um extremo, uma solução, logo não se move

1.6 – 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.

arbitramos $\alpha_1 = .125$

$$\underline{w}_0 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \nabla(\underline{w}_0) = \begin{bmatrix} 0 \\ 8 \end{bmatrix} \quad \underline{w}_1 = \underline{w}_0 - \alpha_1 \nabla(\underline{w}_0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - .0125 \begin{bmatrix} 0 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \\ -1.1 \end{bmatrix}$$

$$F(\vec{w}) = w_1^3 w_2^2 - 2w_1^3 w_2 - 3w_1 w_2^2 + 6w_1 w_2 \quad F(1, -1) = -6 \quad F(1, -1.1) = -6.82$$

$$\alpha^* = \frac{\alpha_1^2 |\underline{\nabla}|^2}{2(F_1 - F_0 + \alpha_1 |\underline{\nabla}|^2)} = \frac{(.000156)(64)}{2[(-6.82) - (-6) + (.0125)(64)]} = -0.25 \quad !?$$

$$\underline{w}_{ot} = \underline{w}_0 - \alpha^* \nabla(\underline{w}_0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - (-.25) \begin{bmatrix} 0 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \Delta F_{esperado} = -\frac{1}{2} \alpha^* |\underline{\nabla}|^2 = +8 \quad !!!!$$

$$\nabla(1,1) = \begin{bmatrix} 3w_1^2 w_2^2 - 6w_1^2 w_2 - 3w_2^2 + 6w_2 \\ 2w_1^3 w_2 - 2w_1^3 - 6w_1 w_2 + 6w_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$H(1,1) = \begin{bmatrix} 6w_1 w_2^2 - 12w_1 w_2 & 6w_1^2 w_2 - 6w_1^2 - 6w_2 + 6 \\ 6w_1^2 w_2 - 6w_1^2 - 6w_2 + 6 & 2w_1^3 - 6w_1 \end{bmatrix} = \begin{bmatrix} -6 & 0 \\ 0 & -4 \end{bmatrix}$$

Gradiente nulo, H definida negativa >>>>> $\alpha < 0$ levou a um máximo !

- 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, $\alpha < 2\alpha_{\text{ótimo}}$.

$$F(\underline{w}_0 + \alpha \underline{d}) = F(\underline{w}_0) + \alpha \underline{d}' \nabla + \frac{1}{2} \alpha \underline{d}' \underline{H} \alpha \underline{d}$$

$$\Delta F(\alpha) = \alpha \underline{d}' \nabla + \frac{1}{2} \alpha^2 \underline{d}' \underline{H} \underline{d}$$

$\Delta F(\alpha)$ é uma parábola com raízes em

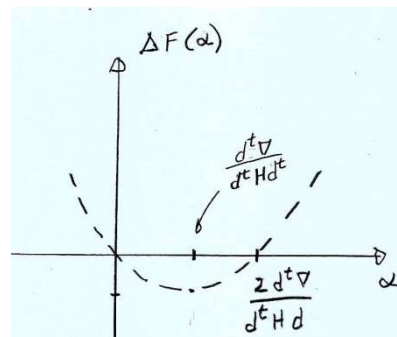
$$F(\alpha)$$

$$\alpha = 0 \text{ e}$$

$$0$$

$$\alpha_{\text{ótimo}}$$

e minimante em $\alpha_{\text{ótimo}} = -\underline{d}' \nabla / \underline{d}' \underline{H} \underline{d}$



- 3 - Considere a função $F(\vec{w}) = \left[(w_1^2 + w_2^2 - 1) w_1 w_2 \right]^2$. Calcule o valor exato e o da aproximação da Hessiana proposta por Levenberg-Marquardt para os pontos $\vec{w}_a = .7000 \ .7000]^t$ e $\vec{w}_b = 1.000 \ 1.000]^t$. Compare os erros e comente a diferença. Observe que $F(\underline{w})$ tem calhas (mínimos) nos eixos w_1 , w_2 e em uma circunferência de raio unitário centrada na origem.

$$F(\vec{w}) = \left[(w_1^2 + w_2^2 - 1) w_1 w_2 \right]^2 = [\varepsilon]^2$$

$$\varepsilon(\vec{w}) = (w_1^2 + w_2^2 - 1) w_1 w_2 = w_1^3 w_2 + w_1 w_2^3 - w_1 w_2$$

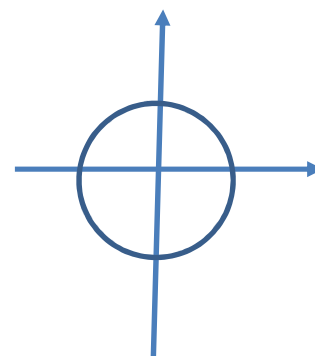
$$\frac{\partial \varepsilon}{\partial w_1} = 3w_1^2 w_2 + w_2^3 - w_2$$

$$\frac{\partial^2 \varepsilon}{\partial w_1^2} = 6w_1 w_2$$

$$\frac{\partial \varepsilon}{\partial w_2} = 3w_2^2 w_1 + w_1^3 - w_1$$

$$\frac{\partial^2 \varepsilon}{\partial w_2^2} = 6w_1 w_2$$

$$\frac{\partial^2 \varepsilon}{\partial w_1 \partial w_2} = 3w_1^2 + 3w_2^2 - 1$$



$$\frac{\partial F}{\partial w_1} = g_1 = 2\varepsilon \frac{\partial \varepsilon}{\partial w_1} \quad \frac{\partial F}{\partial w_2} = g_2 = 2\varepsilon \frac{\partial \varepsilon}{\partial w_2}$$

$$\frac{\partial^2 F}{\partial w_i \partial w_j} = h_{ij} = 2 \left[\frac{\partial \varepsilon}{\partial w_i} \frac{\partial \varepsilon}{\partial w_j} + \varepsilon \frac{\partial^2 \varepsilon}{\partial w_i \partial w_j} \right] \quad LM \quad \tilde{h}_{ij} = 2 \left[\frac{\partial \varepsilon}{\partial w_i} \frac{\partial \varepsilon}{\partial w_j} \right]$$

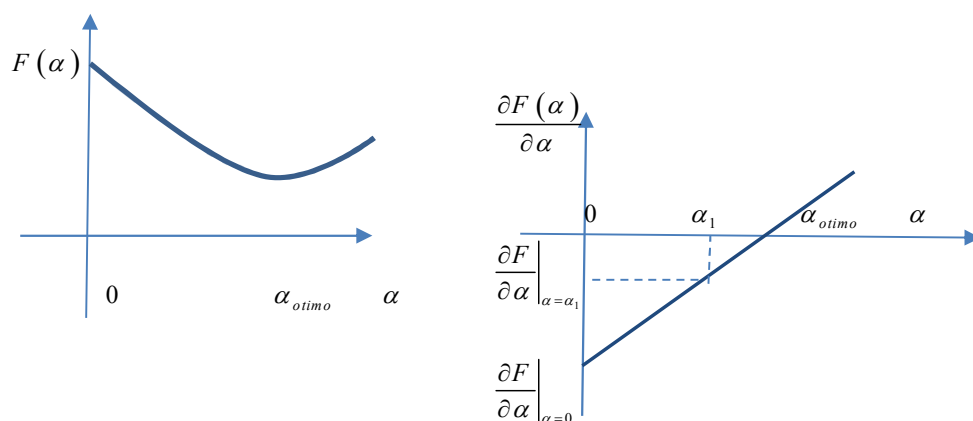
Ponto $\vec{w}_a = [.7000 \ .7000]^T$ próximo de um mínimo (calha) $\varepsilon \simeq .01$

$h_{11} = 962.10^{-3}$	$\tilde{h}_{11} = 904.10^{-3}$	$e\% = 6\%$
$h_{22} = 962.10^{-3}$	$\tilde{h}_{22} = 904.10^{-3}$	$e\% = 6\%$
$h_{12} = 942.10^{-3}$	$\tilde{h}_{12} = 904.10^{-3}$	$e\% = 4\%$

Ponto $\vec{w}_b = [1.000 \ 1.000]^T$ longe de um mínimo $\varepsilon = 1$

$h_{11} = 30$	$\tilde{h}_{11} = 18$	$e\% = 40\%$
$h_{22} = 30$	$\tilde{h}_{22} = 18$	$e\% = 40\%$
$h_{12} = 28$	$\tilde{h}_{12} = 18$	$e\% = 36\%$

4 – Considerando que a aproximação de segunda ordem para $F(w_i)$ é válida crie uma fórmula que permita calcular o passo $\alpha_{ótimo}$ para uma sinapse do algoritmo BP resiliente ou leve em um passo ao ponto de gradiente nulo.



$$\frac{\alpha_{ótimo}}{\frac{\partial F}{\partial \alpha} \Big|_{\alpha=0}} = \frac{\alpha_1}{\frac{\partial F}{\partial \alpha} \Big|_{\alpha=0} - \frac{\partial F}{\partial \alpha} \Big|_{\alpha=\alpha_1}}$$