Math Exercise - MLP 804861

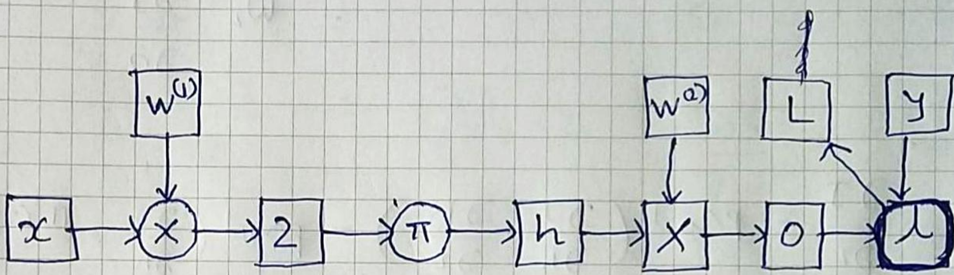
$$f(x) = W_2^T \pi (W_1^T x) \qquad \pi(x) = \frac{1}{1+e^{-x}}$$

$$\lambda = \frac{1}{2}(y - f(x))^2 \quad , \quad \underline{\hat{y} = f(x)}$$

$$\underline{\lambda = \frac{1}{2}(y-\hat{y})^2} \qquad \underline{z = W_1^T x} \quad , \quad \underline{h = \pi(z)}$$

$$\underline{o = W_2^T h}$$

a)



* computational graph *

b) gradient matrix G_1 of λ with respect to W_1.

$$G_1 = \frac{\partial L}{\partial W_1} \qquad J = L$$

$$\frac{\partial J}{\partial L} = 1$$

$$\frac{\partial J}{\partial o} = \text{prod}\left(\frac{\partial J}{\partial L}, \frac{\partial L}{\partial o}\right) = \frac{\partial L}{\partial o}$$

$$f(x) = o = \hat{y} \qquad \frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}} \frac{1}{2}(y-\hat{y})^2$$

$$= -(y-\hat{y})$$

$$\frac{\partial J}{\partial w_2} = \text{prod}\left(\frac{\partial J}{\partial o}, \frac{\partial o}{\partial w_2}\right)$$

$$\frac{\partial}{\partial w_2} w_2^T h = h^T$$

$$= \text{prod}\left(\frac{\partial J}{\partial o}, h^T\right) = \frac{\partial J}{\partial o} h^T$$

$$\boxed{G_2 = h^T \cdot -(y - \hat{y})}$$ — ⓒ

$$\frac{\partial J}{\partial h} = \text{prod}\left(\frac{\partial J}{\partial o}, \frac{\partial o}{\partial h}\right) = \text{prod}\left(\frac{\partial J}{\partial o}, w_2\right)$$

$$= w_2^T \frac{\partial J}{\partial o}$$

$$\frac{\partial J}{\partial z} = \text{prod}\left(\frac{\partial J}{\partial h}, \frac{\partial h}{\partial z}\right) = \text{prod}\left(\frac{\partial J}{\partial h}, \pi'(z)\right)$$

$$= \frac{\partial J}{\partial h} \odot \pi'(z)$$

$$\boxed{\pi'(z) = \frac{e^{-z}}{(1 + e^{-z})^2}}$$

$$\frac{\partial J}{\partial w_1} = \text{prod}\left(\frac{\partial J}{\partial z}, \frac{\partial z}{\partial w_1}\right) = \text{prod}\left(\frac{\partial J}{\partial z}, x\right)$$

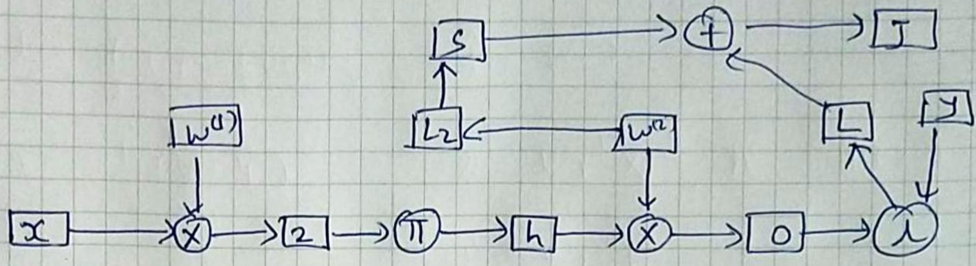
$$\boxed{G_1 = \frac{\partial J}{\partial z} x^T}$$

$$\boxed{G_1 = \frac{\partial J}{\partial w_1} = \left[w_2^T \left(-(y - \hat{y})\right) \odot \pi'(z)\right] x^T}$$ — ⓑ

FABER-CASTELL

D) gradient with L_2 regularization on w_2.

$$L = \lambda + \|w_2\|_2^2$$



* computational graph with L_2 *

$$J = S + L \qquad S = \frac{\lambda}{2}\left(\|w^2\|_2^2\right)$$

$$\tilde{g}_2 = \frac{\partial J}{\partial w_2} \qquad \frac{\partial J}{\partial L} = 1 \qquad \frac{\partial J}{\partial S} = 1 \qquad \frac{\partial S}{\partial w^{(2)}} = \lambda w^{(2)}$$

$$\tilde{g}_2 = g_2 + \frac{\partial S}{\partial w^{(2)}}$$

(prev. calculated g_2 grad)

$$\boxed{\tilde{g}_2 = -(y - \hat{y})h^T + \lambda w^{(2)}}$$