Saarland University, Department of Computer Science
Neural Network Assignment 6

Deborah Dormah Kanubala (7025906) , Irem Begüm Gündüz (7026821), Anh Tuan Tran (7015463)

December 21, 2022

Exercise 6.3 - Reading exercise (Bonus)

Backpropagation requires forward pass computations to compute the derivatives correctly. This makes backpropagation application not possible when an unknown computation is introduced in a forward pass. Reinforcement learning may be useful in such cases, however, reinforcement learning comes with its own challenges such as bad scalability when averaging the noise by all permutations of variables. Besides that, it is not clear whether real neurons use backward passes to propagate derivatives therefore, it is not possible to model cortex learning using backpropagation. There are new methodologies being developed to overcome the limitations of backpropagation.

The forward-forward algorithm replaces forward and backward passes of the backpropagation algorithm with two forward-passing steps that compute forward-pass steps with different data, as one computes a negative pass while the other computes a positive pass. The sum of the squared neural activities and its negative are used to evaluate "goodness" in every hidden layer while adjusting the weights by increasing or decreasing with respectively. It is possible to model cortex learning using the FF since it doesn't use backward passes nor store derivatives to update the weights. Therefore, it offers an alternative to RIL in terms of efficiency to use hardware. However, it is not an alternative to backpropagation since it doesn't scale well in larger neural networks.

Exercise 6.1

6.1)

a) Denote

$$h = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}; \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ -1 \\ 2 \end{bmatrix}; \quad o = \begin{bmatrix} o_1 \\ o_2 \end{bmatrix}$$

the parameters of LeakyReLU is α

we have

$$h = f\left(W^{(1)T} x\right)$$

$$= f\left(\begin{bmatrix} -0.2 & -0.1 & 0.2 & 0.2 \\ 0.9 & 0.3 & 0.5 & -0.5 \\ 0.4 & 0.4 & -0.7 & 0.5 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ -1 \\ 2 \end{bmatrix} \right)$$

$$= f\left(\begin{bmatrix} -0.5 \\ 1.5 \\ 3.3 \end{bmatrix} \right) = \begin{bmatrix} \alpha \times -0.5 \\ 1.5 \\ 3.3 \end{bmatrix}$$

$$o = s\left(W^{(2)T} h\right)$$

$$= s\left(\begin{bmatrix} 0.6 & -0.1 & -0.5 \\ -0.2 & 0.8 & -0.3 \end{bmatrix} \begin{bmatrix} -0.5\alpha \\ 1.5 \\ 3.3 \end{bmatrix} \right)$$

$$= S\left(\begin{bmatrix} -0.3\alpha - 1.8 \\ 0.1\alpha + 0.21 \end{bmatrix}\right)$$

$$= \begin{bmatrix} e^{-0.3\alpha - 1.8} \Big/ \left(e^{-0.3\alpha-1.8} + e^{0.1\alpha+0.21}\right) \\ e^{0.1\alpha+0.21} \Big/ \left(e^{-0.3\alpha-1.8} + e^{0.1\alpha+0.21}\right) \end{bmatrix}$$

b) We have

$$Y = \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\text{Loss} = L = \frac{-1}{2} \sum_{i=1}^{2} y_i \cdot \log(O_i) + (1-y_i)\log(1-O_i)$$

$$= \log(1-O_0) + \log(O_1)$$

$$\partial L / \partial O_0 = \frac{-1}{2} \times \frac{1}{1-O_0} \qquad \partial L / \partial O_1 = \frac{-1}{2} \times \frac{1}{O_1}$$

$$\frac{\partial O_i}{\partial w^{(2)}_{ij}} = \frac{\partial\, e^{\left(\sum_{j=1}^{3} w^{(1)}_{ij} h_j\right)} \Big/ \left(e^{\sum_{j=1}^{3} w^{(1)}_{0j} h_j} + e^{\sum_{j=1}^{3} w_{1j} h_j}\right)}{\partial w^{(1)}_{ij}}$$

Denote $a_i = \left(\sum_{j=1}^{3} w^{(1)}_{ij} h_j\right) \Rightarrow \frac{\partial a_i}{\partial w^{(1)}_{ij}} = h_j$

$$\frac{\partial O_i}{\partial w^{(2)}_{ij}} = \frac{\partial\, e^{a_i}/(e^{a_0} + e^{a_1})}{\partial a_i} \cdot \frac{\partial a_i}{\partial w^{(1)}_{ij}}$$

$$\frac{\partial\, e^{a_i}/(e^{a_0} + e^{a_1})}{\partial a_i} = O_i \cdot (1 - O_i)$$

$$\Rightarrow \frac{\partial O_i}{\partial w^{(2)}_{ij}} = O_i (1 - O_i) \cdot h_j$$

We have, for $k \neq i$, $\frac{\partial O_k}{\partial w^{(1)}_{ij}} = \frac{\partial\, e^{a_k}/(e^{a_0} + e^{a_1})}{\partial a_i} \cdot \frac{\partial a_i}{\partial w^{(1)}_{ij}}$

$$= - O_k O_i \cdot h_j$$

$$\Rightarrow \frac{d O_0}{d W^{(1)}} = \begin{bmatrix} O_0(1-O_0)\, h_1 & -O_0 O_1 \cdot h_1 \\ O_0(1-O_0)\, h_2 & -O_0 O_1 \cdot h_2 \\ O_0(1-O_0)\, h_3 & -O_0 O_1 \cdot h_3 \end{bmatrix}$$

$$\frac{d O_1}{d w^{(1)}} = \begin{bmatrix} -O_0 O_1 \cdot h_1 & O_1 (1-O_1) h_1 \\ -O_0 O_1 \cdot h_2 & O_1 (1-O_1) h_2 \\ -O_0 O_1 \cdot h_3 & O_1 (1-O_1) h_3 \end{bmatrix}$$

$$\frac{d O}{d w^{(1)}} = \frac{d O_1}{d w^{(1)}} + \frac{d O_2}{d w^{(1)}}$$

$$\frac{d L}{d w^{(1)}} = \frac{\partial L}{\partial O_0} \cdot \frac{d O_0}{d w^{(1)}} + \frac{\partial L}{\partial O_1} \cdot \frac{d O_1}{d w^{(1)}}$$

$$\frac{d L}{d w^{(1)}} = \frac{d L}{d h} \cdot \frac{d h}{d w^{(2)}}$$