Authors:

- Mhd Jawad Al Rahwanji - 7038980 - mhal00002@stud.uni-saarland.de
- Christian Singer - 7039059 - chsi00002@stud.uni-saarland.de

Exercise 6.1 - Forward Pass (Inference) and Backward Propagation (Backprop) of Errors for a Fully-Connected Neural Network

a) We have:

$$X = [3,\ 1,\ -1,\ 2]$$

$$W_1 = [[-0.2, -0.1,\ 0.2,\ 0.2],\ [0.9,\ 0.3,\ 0.5, -0.5],\ [0.4,\ 0.4, -0.7,\ 0.5]]^T$$

$$W_2 = [[0.6, -0.1, -0.5],\ [-0.2,\ 0.8, -0.3]]^T$$

Now we begin the forward pass:

$$H = \sigma_{LeakyReLU}(XW_1,\ a = 0.01) = \sigma_{LeakyReLU}([-0.5,\ 1.5,\ 3.3],\ a = 0.01)$$
$$= [-0.005,\ 1.5,\ 3.3]$$

$$O = \sigma_{Softmax}(HW_2) = \sigma_{Softmax}([-1.803,\ 0.211]) =$$
$$[\frac{e^{-1.803}}{e^{-1.803}+e^{0.211}},\ \frac{e^{0.211}}{e^{-1.803}+e^{0.211}}] = [0.11774083035,\ 0.88225916964]$$

b) We have the following:

$$w = w - a\frac{\partial E}{\partial w} \qquad \text{over } W_1 \text{ and } W_2$$

$$E_{BCE} = -(y\,log(o) + (1-y)\,log(1-o)) \qquad \text{over O and } Y^T$$

$$E'_{BCE} = -(\frac{y}{o} - \frac{1-y}{1-o}) = \frac{o-y}{o(1-o)} \qquad \text{over O and } Y^T$$

$$\sigma'_{Softmax} = o(1-o) \qquad \text{over O}$$

$$\sigma'_{LeakyReLU} = \{0.01 : x < 0\ \&\ 1 : x \geq 0\} \qquad \text{over H}$$

We begin backpropagation for W_2:

$$\frac{\partial E}{\partial W_2} = \frac{\partial E}{\partial O}\frac{\partial O}{\partial W_2} = E'_{BCE}\frac{\partial O}{\partial W_2} = E'_{BCE}\frac{\partial \sigma_{Softmax}(HW_2)}{\partial W_2}$$
$$= [E'_{BCE} \circ \sigma'_{Softmax}]\frac{\partial (HW_2)}{\partial W_2} = H^T[E'_{BCE} \circ \sigma'_{Softmax}]$$

Dim = 3x2, same as W_2

Now W_1:

$$\frac{\partial E}{\partial W_1} = \frac{\partial E}{\partial O}\frac{\partial O}{\partial W_1} = E'_{BCE}\frac{\partial O}{\partial W_1} = E'_{BCE}\frac{\partial \sigma_{Softmax}(HW_1)}{\partial W_1}$$
$$= [E'_{BCE} \circ \sigma'_{Softmax}]\frac{\partial (HW_2)}{\partial W_1} = X^T[[[E'_{BCE} \circ \sigma'_{Softmax}]W_2^T] \circ \sigma'_{LeakyReLU}]$$

Dim = 4x3, same as W_1

c) We have the following:

$$w = w - a\frac{\partial E}{\partial w} \qquad \text{over } W_1 \text{ and } W_2$$

$$a = 0.1$$

The updated W_1 and W_2:

$$W_1 = [[-4.16802e-6, -1.38934e-6, 1.38934e-6, -2.77868e-6],$$
$$[0.124765, 0.0415884, -0.0415884, 0.0831768],$$
$$[0.274486, 0.0914952, -0.0914952, 0.18299]]^T$$

$$W_2 = [[-0.00059, 0.17661, 0.388545],$$
$$[0.00059, -0.17661, -0.388545]]^T$$