Homework 01

Problem 1

Download the Docker container from Moodle and follow the instructions in the readme file to set it up.Verify that your setup works by running the Keras MNIST example1. You don't need to provide evidence to receive points for this task

Problem 2

$$
\begin{aligned}
\implies \sigma'(x) &= ((1 - exp(-x))^{-1})' \\
&= -(1 + exp(-x))^{-2}(-exp(-x)) \\
&= \frac{exp(-x)}{(1 + exp(-x))^2} \\
&= \frac{1}{1 + exp(-x)} \cdot \frac{exp(-x)}{1 + exp(-x)} \\
&= \frac{1}{1 + exp(-x)} \cdot \frac{(1 + exp(-x)) - 1}{1 + exp(-x)} \\
&= \frac{1}{1 + exp(-x)} \cdot \frac{(1 + exp(-x)) - 1}{1 + exp(-x)} \\
&= \frac{1}{1 + exp(-x)} \cdot (1 - \frac{1}{1 + exp(-x)}) \\
&= \sigma(x) \cdot (1 - \sigma(x))
\end{aligned}
$$

Problem 3

Answer to the problem goes here.

1. Training.
   Iteration 1:

$$w^{(1)} = w^{(0)} - \alpha * (\sigma(\boldsymbol{x_1} \cdot \boldsymbol{w^{(0)}}) - y_1) * \sigma'(\boldsymbol{x_1} \cdot \boldsymbol{w^{(0)}}) * \boldsymbol{x_1}$$

$$
\begin{aligned}
\sigma(\boldsymbol{x_1} \cdot \boldsymbol{w^{(0)}}) &= \sigma\left(\begin{bmatrix} -1,28 \\ 0,09 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) \\
&= \sigma(-1,28 * (-1) + 0,09 * 1) \\
&= \sigma(1,37) \\
&= \frac{1}{1 + exp(-1,37)} \\
&= 0,8
\end{aligned}
$$

$$\sigma^{'}(\boldsymbol{x_1} \cdot \boldsymbol{w^{(0)}}) = 0,8 * (1 - 0,8) = 0,16$$

$$\begin{aligned} w^{(1)} &= \begin{bmatrix} -1 \\ 1 \end{bmatrix} - 1 * (0,8 - 0) * 0,16 * \begin{bmatrix} -1,28 \\ 0,09 \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ 1 \end{bmatrix} - 0,128 * \begin{bmatrix} -1,28 \\ 0,09 \end{bmatrix} \\ &= \begin{bmatrix} -1 + 0,128 * 1,28 \\ 1 - 0,128 * 0,09 \end{bmatrix} \\ &= \begin{bmatrix} -0,84 \\ 0,99 \end{bmatrix} \end{aligned}$$

Iteration 2:

$$w^{(2)} = w^{(1)} - \alpha * (\sigma(\boldsymbol{x_2} \cdot \boldsymbol{w^{(1)}}) - y_2) * \sigma^{'}(\boldsymbol{x_2} \cdot \boldsymbol{w^{(1)}}) * \boldsymbol{x_2}$$

$$\begin{aligned} \sigma(\boldsymbol{x_2} \cdot \boldsymbol{w^{(1)}}) &= \sigma\left(\begin{bmatrix} 0,17 \\ 0,39 \end{bmatrix} \cdot \begin{bmatrix} -0,84 \\ 0,99 \end{bmatrix}\right) \\ &= \sigma(0,17 * (-0,84) + 0,39 * 0,99) \\ &= \sigma(0,24) \\ &= \frac{1}{1 + exp(-0,24)} \\ &= 0,56 \end{aligned}$$

$$\sigma^{'}(\boldsymbol{x_2} \cdot \boldsymbol{w^{(1)}}) = 0,56 * (1 - 0,56) = 0,25$$

$$\begin{aligned} w^{(2)} &= \begin{bmatrix} -0,84 \\ 0,99 \end{bmatrix} - 1 * (0,56 - 1) * 0,25 * \begin{bmatrix} 0,17 \\ 0,39 \end{bmatrix} \\ &= \begin{bmatrix} -0,84 \\ 0,99 \end{bmatrix} + 0,11 * \begin{bmatrix} 0,17 \\ 0,39 \end{bmatrix} \\ &= \begin{bmatrix} -0,84 + 0,11 * 0,17 \\ 0,99 + 0,11 * 0,29 \end{bmatrix} \\ &= \begin{bmatrix} -0,82 \\ 1,02 \end{bmatrix} \end{aligned}$$

Iteration 3:

$$w^{(3)} = w^{(2)} - \alpha * (\sigma(\boldsymbol{x_3} \cdot \boldsymbol{w^{(2)}}) - y_3) * \sigma^{'}(\boldsymbol{x_3} \cdot \boldsymbol{w^{(2)}}) * \boldsymbol{x_3}$$

$$\begin{aligned}
\sigma(\boldsymbol{x_3} \cdot \boldsymbol{w^{(2)}}) &= \sigma\left(\begin{bmatrix}1,36\\0,46\end{bmatrix} \cdot \begin{bmatrix}-0,82\\1,02\end{bmatrix}\right) \\
&= \sigma(1,36 * (-0,28) + 0,46 * 1,02) \\
&= \sigma(0,65) \\
&= \frac{1}{1 + exp(-0,65)} \\
&= 0,66
\end{aligned}$$

$$\sigma^{'}(\boldsymbol{x_3} \cdot \boldsymbol{w^{(2)}}) = 0,66 * (1 - 0,66) = 0,22$$

$$\begin{aligned}
w^{(3)} &= \begin{bmatrix}-0,82\\1,02\end{bmatrix} - 1 * (0,66 - 1) * 0,22 * \begin{bmatrix}1,36\\0,46\end{bmatrix} \\
&= \begin{bmatrix}-0,82\\1,02\end{bmatrix} + 0,07 * \begin{bmatrix}1,36\\0,46\end{bmatrix} \\
&= \begin{bmatrix}-0,82 + 0,07 * 1,36\\1,02 + 0,07 * 0,46\end{bmatrix} \\
&= \begin{bmatrix}-0,72\\1,05\end{bmatrix}
\end{aligned}$$

4. Iteration:

$$w^{(4)} = w^{(3)} - \alpha * (\sigma(\boldsymbol{x_4} \cdot \boldsymbol{w^{(3)}}) - y_4) * \sigma^{'}(\boldsymbol{x_4} \cdot \boldsymbol{w^{(3)}}) * \boldsymbol{x_4}$$

$$\begin{aligned}
\sigma(\boldsymbol{x_4} \cdot \boldsymbol{w^{(3)}}) &= \sigma\left(\begin{bmatrix}-0,51\\-0,32\end{bmatrix} \cdot \begin{bmatrix}-0,72\\1,05\end{bmatrix}\right) \\
&= \sigma(-0,51 * (-0,72) - 0,32 * 1,05) \\
&= \sigma(0,7) \\
&= \frac{1}{1 + exp(-0,7)} \\
&= 0,67
\end{aligned}$$

$$\sigma^{'}(\boldsymbol{x_4} \cdot \boldsymbol{w^{(3)}}) = 0,67 * (1 - 0,67) = 0,22$$

$$\begin{aligned} w^{(4)} &= \begin{bmatrix} -0,72 \\ 1,05 \end{bmatrix} - 1 * (0,67 - 0) * 0,22 * \begin{bmatrix} -0,51 \\ -0,32 \end{bmatrix} \\ &= \begin{bmatrix} -0,72 \\ 1,05 \end{bmatrix} - 1,45 * \begin{bmatrix} -0,51 \\ -0,32 \end{bmatrix} \\ &= \begin{bmatrix} -0,72 + 1,45 * 0,51 \\ 1,05 + 1,45 * 0,32 \end{bmatrix} \\ &= \begin{bmatrix} 0.02 \\ 1,51 \end{bmatrix} \end{aligned}$$

2. **Evaluation**

   Before training with $w^{(0)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

$$\begin{aligned} L &= (\sigma(\boldsymbol{x_1} \cdot \boldsymbol{w_0}) - y_1)^2 + (\sigma(\boldsymbol{x_2} \cdot \boldsymbol{w_0}) - y_2)^2 \\ &= (\sigma(-0,5 * (-1) - 1 * 1) - 0)^2 + (\sigma(0,75 * (-1) + 0,25 * 1) - 1)^2 \\ &= (\sigma(-0,5) - 0)^2 + (\sigma(-0,5) - 1)^2 \\ &= (0,38 - 0)^2 + (0,38 - 1)^2 \\ &= 0,38^2 + (-0,62)^2 = 0,1444 + 0,3844 = 0,5288 \end{aligned}$$

   After training with $w^{(4)} = \begin{bmatrix} 0,02 \\ 1,51 \end{bmatrix}$

$$\begin{aligned} L &= (\sigma(\boldsymbol{x_1} \cdot \boldsymbol{w_1}) - y_1)^2 + (\sigma(\boldsymbol{x_2} \cdot \boldsymbol{w_1}) - y_2)^2 \\ &= (\sigma(-0,5 * 0,02 - 1 * 1,51) - 0)^2 + (\sigma(0,75 * 0,02 + 0,25 * 1,51) - 1)^2 \\ &= (\sigma(-1,52) - 0)^2 + (\sigma(0,3) - 1)^2 \\ &= (0,179 - 0)^2 + (0,574 - 1)^2 \\ &= 0,179^2 + (-0,426)^2 = 0,032 + 0,181 = 0,213 \end{aligned}$$