Machine Learning (DS4023) Assignment 4

Problem 1: Neural Networks (20 pts)

Consider a 3-layer fully connected neural network with the following architecture:

- **Input layer**: $n = 4$ neurons.
- **Hidden layer**: $m = 3$ neurons using a custom activation function $f(x) = \mathrm{ReLU}(x) + \sin(x)$.
- **Output layer**: $k = 2$ neurons using the softmax activation function $\sigma(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$.

The network parameters (**weights and biases**) are given as:

- $\boldsymbol{W}_1 \in \mathbb{R}^{3\times4}$ and $\boldsymbol{b}_1 \in \mathbb{R}^3$ for the hidden layer.
- $\boldsymbol{W}_2 \in \mathbb{R}^{2\times3}$ and $\boldsymbol{b}_2 \in \mathbb{R}^2$ for the output layer.

Given the input vector $\boldsymbol{x} \in \mathbb{R}^4$ and target output $\boldsymbol{y} \in \mathbb{R}^2$. Define the loss function as cross-entropy loss:

$$\mathrm{Loss} = -\sum_{i=1}^{k} y_i \log(\hat{y}_i),$$

where $\hat{y}$ is the output after the softmax activation.

Your tasks (rounding to 4 decimal points):

1. 1) Derive the equations for the forward pass through the network, including both the hidden and output layers. (3 pts)

   2) Calculate the outputs $\boldsymbol{Z}_1, \boldsymbol{H}, \boldsymbol{Z}_2$, and $\hat{\boldsymbol{y}}$ explicitly for a given input $\boldsymbol{x} = [1, -1, 0.5, 2]^T$ and the following initial weights and biases:

$$\boldsymbol{W}_1 = \begin{pmatrix} 0.1 & -0.2 & 0.3 & 0.4 \\ 0.5 & -0.3 & 0.1 & -0.2 \\ 0.4 & 0.2 & -0.5 & 0.3 \end{pmatrix}, \quad \boldsymbol{b}_1 = \begin{pmatrix} 0.1 \\ -0.1 \\ 0.05 \end{pmatrix}$$

$$\boldsymbol{W}_2 = \begin{pmatrix} -0.3 & 0.2 & 0.1 \\ 0.4 & -0.5 & 0.3 \end{pmatrix}, \quad \boldsymbol{b}_2 = \begin{pmatrix} 0.05 \\ -0.05 \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Note that $\boldsymbol{Z}_1$ is the net input to the hidden layer, $\boldsymbol{H}$ is the activation output of the hidden layer, and $\boldsymbol{Z}_2$ is the net input to the output layer. (3 pts)

2. Derive the gradient of the loss with respect to each parameter $(\boldsymbol{W}_1, \boldsymbol{b}_1, \boldsymbol{W}_2, \boldsymbol{b}_2)$ in the network and obtain the gradient values using results from the first question. Use matrix calculus to express the gradients. Hint: You can first calculate the error terms $\boldsymbol{\delta}_2$ and $\boldsymbol{\delta}_1$ for each layer and use them to express the gradients. (10 pts)

3. Suppose the learning rate $\alpha = 0.001$. Please calculate the updated parameter values after one back propagation process. (4 pts)

Problem 2: Programming (80 pts)

Complete the jupyter notebook attached on programming for CNN and RNN. Submit the completed file.

To submit:

1. A file containing the written answer to the Problems 1.

2. The Jupyter notebook with solutions to Problem 2.