Introduction to Machine Learning (CSCI-UA.473): Homework 4

Instructor: Sumit Chopra

November 4^{th}, 2021

Submission Instructions

You must typeset the answers using LaTeXand compile them into a single PDF file. Name the pdf file as: ⟨Your-NetID⟩_hw4.pdf. For the programming part of the assignment, complete the Jupyter notebook named HW4.ipynb. Create a ZIP file containing both the PDF file and the completed Jupyter notebook. Name it ⟨Your-NetID⟩_hw4.zip. Submit the ZIP file on Brightspace. The due date is **November** 18^{th}**, 2021,** $11:59$ **PM**.

Theory

Question T1: Back propagation of a 2D Convolution Operation (15 points)

Let the input be an 2D gray scale image of size $m \times n$, denoted by the matrix $X \in \Re^{m \times n}$. Let the parameters of the $p \times p$ convolution kernel be denoted by $[W, b]$, where $W \in \Re^{p \times p}$ are the weights of the kernel and b is the bias associated with the kernel. Let us denote by L the loss function of your model and by δ the gradient of the loss with respect to the output of the convolution operation. Write the expression for the following:

1. (5 points) Gradient of the loss function L with respect to the inputs X: $\frac{dL}{dX}$
2. (5 points) Gradient of the loss function L with respect to the weights W: $\frac{dL}{dW}$
3. (5 points) Gradient of the loss function L with respect to the bias b: $\frac{dL}{db}$

Please write all the steps that led you to the final expression. **No points will be given if only the final expression is provided without the steps**.

Question T2: Back propagation of other functions (15 points)

Compute the back propagation expression (the gradient of the loss function L with respect to the input x, where $x \in \Re^d$ is the 1D input vector of size d), for the following functions:

1. (5 points) Tanh: $f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
2. (5 points) Max pooling: $f(x) = \max_{i \in \{1,\ldots,d\}} x_i$
3. (5 points) Average pooling: $f(x) = \frac{1}{d} \sum_{i=1}^{d} x_i$

Here again, assume that you know the gradient of the loss L with respect to the output of each function and denote it by δ. Please write all the steps that led you to the final expression. **No points will be given if only the final expression is provided without the steps**.

Practicum

See the accompanying Python notebook.

Question P1: Long-Short Term Memory Networks for sequence modeling (35 points)

Question P2: Ensemble of neural networks for multi-class classification (35 points)