Very Deep Learning | **RPTU Kaiserslautern**
WS 2023/24 | Dr. Muhammad Zeshan Afzal

Exercise 2 - Convolutions and Loss Functions

Deadline: November 27, 2023 at 23:59 **Total Marks: 30**

Submission

- Submissions through OLAT. Only one group member needs to submit it.
- Your submission should contain a PDF with the solutions to the exercise questions (and any Python code files) zipped together in a single file.
- Include the group number along with the names and matriculation numbers of all group members on the PDF.
- For Jupyter notebooks, please save them with the outputs of your code displayed.

2.1. Convolutions [3+1+1+1+1+1=8]

Denote by X the following 5x5 input image with one channel:

8	2	9	8	3
3	4	6	2	7
5	9	3	9	7
6	1	3	6	8
5	1	6	8	1

Denoting a 3x3 filter by K. Assume we have the following 3x3 filters:

1	0	-1
2	0	-2
1	0	-1

1	2	1
0	0	0
-1	-2	-1

0	1	0
1	-4	1
0	1	0

We adopt the convention that for a 2D array X (or K), $X_{i,j}$ denotes the element in **row** i **and column** j**, counting from the top left corner**. In particular, $X_{3,4} = 9$ in our example. Furthermore, unless otherwise stated, we assume zero padding, i.e. $X_{i,j} \equiv 0$ if $i < 0$ or $i > \text{size}(X)$ (similarly for j).

Tasks:

i) Apply all three given filters (by cross-correlation) to the above dataset, i.e.:

$$Y_{i,j} = (K \star X)_{i,j} = \sum_{m=-1}^{1} \sum_{n=-1}^{1} K_{m+2,n+2} X_{i+m,j+n} \tag{1}$$

where Y is the output. Use 'same' padding and a stride of one.

ii) Look at the structure of the filters. What do they do?

iii) What is the difference between 'valid' and 'same' padding?

iv) Why do we prefer Convolutional Neural Networks (CNNs) over a Multi-Layer Perceptron (MLP) for image data?

v) Given an input image of size $H \times H$, a convolutional filter of size k, padding p and stride s. Write down the formula for calculating the dimensions of the output of the convolution operation.

vi) Given an input of size (H, W, C) convolved with N Conv2D filters of size k, what are the number of trainable parameters in this convolutional layer? Write down the formula. Assume both weights and biases are present.

2.2. Loss Functions and Optimization [3+4=7]

i) Given $\hat{y} = \text{softmax}(z)$ with

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{k=1}^{N} e^{z_k}},$$

where $\hat{y} \in \mathbb{R}^N$ and N is the number of classes of a classification problem. Calculate $\frac{\partial \hat{y}_i}{\partial z_j}$.

ii) Given $\hat{y} = \text{softmax}(z)$, a target vector $y \in \mathbb{R}^N$ and the cross-entropy loss function defined as

$$L(y, \hat{y}) = -\sum_{k=1}^{N} y_k \log \hat{y}_k. \tag{2}$$

Calculate $\frac{\partial L}{\partial z_i}$ and simplify your results as far as possible. *Hint: Make use of the chain rule, note that y_i are constants and $\sum_i y_i = 1$.*

2.3. Loss Functions with Regularization [5]

Consider the input dataset $X \in \mathbb{R}^{n \times d}$ with n samples of size d, a target vector $y \in \mathbb{R}^n$, a weight vector $w \in \mathbb{R}^d$ and a prediction $\hat{y} = Xw$. The **regularized** mean squared error

(MSE) is given by:

$$L(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 + \lambda \sum_{i=1}^{d} w_i^2 \tag{3}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 + \lambda ||w||_2^2, \tag{4}$$

where $||w||_2$ is the Euclidean norm of w and $\lambda > 0$ is some given regularization parameter.

Task: Determine in closed form the vector w that minimizes L.

Hint: You may find the Matrix Cookbook (available online) useful.

2.4. CIFAR Challenge [5]

Follow the instructions in the Jupyter notebook `Task_2.4.ipynb` to complete the CIFAR competition using PyTorch. Your task is to fill in the missing code annotated with `TODO` tags in the comments, and get an accuracy of at least 70%.

2.5. Depthwise Separable Convolutions [5]

In this task, you will explore depth-wise separable convolutions, which is a special type of convolution. Follow the instructions in the Jupyter notebook `Task_2.5.ipynb` and fill in the missing code annotated with `TODO` tags.