DS-GA 1008: Deep Learning, Spring 2020
Homework Assignment 2

Due: Friday, March 6th, 2020 at 11:59pm

`The search for truth is more precious than its possession.`
`Albert Einstein (1879 - 1955)`

1. Fundamentals

1.1. Convolution

Table 1 depicts two matrices. The one on the left represents an 5×5 single-channel image $\boldsymbol{A}$. The one on the right represents a 3×3 convolution kernel $\boldsymbol{B}$.

(a) What is the dimensionality of the output if we forward propagate the image over the given convolution kernel with no padding and stride of 1?
Output Width $= \frac{A_w - B_w + 2P}{S_w} + 1 = 5 - 3 + 1 = 3$
Output height $= \frac{A_h - B_h + 2P}{S_h} + 1 = 3$
Therefore, the output is 3x3.

(b) Give a general formula of the output width O in terms of the input width I, kernel width K, stride S, and padding P (both in the beginning and in the end). Note that the same formula holds for the height. Make sure that your answer in part (a) is consistent with your formula.
$O = \frac{I - K + 2P}{S} + 1$

(c) Compute the output $\boldsymbol{C}$ of forward propagating the image over the given convolution kernel. Assume that the bias term of the convolution is zero.
$C_{11} = 4 * 4 + 3 * (5 + 2) + 5 * (3 + 3 + 2) + 2 * 4 + 4 * 3 + 3 * 4 = 109$
$C_{12} = 4 * 5 + 3 * (2 + 2) + 5 * (3 + 2 + 2) + 2 * 3 + 4 * 4 + 3 * 1 = 92$
$C_{13} = 4 * 2 + 3 * (2 + 1) + 5 * (2 + 2 + 4) + 2 * 4 + 4 * 1 + 3 * 1 = 72$
$C_{21} = 4 * 3 + 3 * (3 + 2) + 5 * (4 + 3 + 4) + 2 * 5 + 4 * 1 + 3 * 4 = 108$
$C_{22} = 4 * 3 + 3 * (2 + 2) + 5 * (3 + 4 + 1) + 2 * 1 + 4 * 4 + 3 * 1 = 85$
$C_{23} = 4 * 2 + 3 * (2 + 4) + 5 * (4 + 1 + 1) + 2 * 4 + 4 * 1 + 3 * 2 = 74$
$C_{31} = 4 * 4 + 3 * (3 + 4) + 5 * (5 + 1 + 4) + 2 * 5 + 4 * 1 + 3 * 3 = 110$
$C_{32} = 4 * 3 + 3 * (4 + 1) + 5 * (1 + 4 + 1) + 2 * 1 + 4 * 3 + 3 * 1 = 74$
$C_{33} = 4 * 4 + 3 * (1 + 1) + 5 * (4 + 1 + 2) + 2 * 3 + 4 * 1 + 3 * 4 = 79$

Therefore, $C =$

109	92	72
108	85	74
110	74	79

(d) Suppose the gradient backpropagated from the layers above this layer is a 3×3 matrix of all 1s. Write the value of the gradient with respect to the input image backpropagated out of this layer. That is, you are given that $\frac{\partial E}{\partial \boldsymbol{C}_{ij}} = 1$ for some scalar error E and $i, j \in \{1, 2, 3\}$. You need to compute $\frac{\partial E}{\partial \boldsymbol{A}_{ij}}$ for $i, j \in \{1, \ldots, 5\}$. The chain rule should help!

DS-GA 1008: Deep Learning, Spring 2020
Homework Assignment 2

Due: Friday, March 6th, 2020 at 11:59pm

Since it is laborious to type the chain rule for each cell in latex, I write the generalized chain rule formula and give a few examples for some i,j, calculate the rest result by hand, and put them in a gradient matrix.

$\frac{\partial E}{\partial A_{ij}} = \frac{\partial E}{\partial C_{11}}\frac{\partial C_{11}}{\partial A_{ij}} + \frac{\partial E}{\partial C_{12}}\frac{\partial C_{12}}{\partial A_{ij}} + \frac{\partial E}{\partial C_{13}}\frac{\partial C_{13}}{\partial A_{ij}} + \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{ij}} + \frac{\partial E}{\partial C_{22}}\frac{\partial C_{22}}{\partial A_{ij}} + \frac{\partial E}{\partial C_{23}}\frac{\partial C_{23}}{\partial A_{ij}} + \frac{\partial E}{\partial C_{31}}\frac{\partial C_{31}}{\partial A_{ij}} + \frac{\partial E}{\partial C_{32}}\frac{\partial C_{32}}{\partial A_{ij}} + \frac{\partial E}{\partial C_{33}}\frac{\partial C_{33}}{\partial A_{ij}} = \sum_{r=1}^{3}\sum_{q=1}^{3}\frac{\partial E}{\partial C_{rq}}\frac{\partial C_{rq}}{\partial A_{ij}} = \sum_{r=1}^{3}\sum_{q=1}^{3}\frac{\partial C_{rq}}{\partial A_{ij}}$ as $\frac{\partial E}{\partial C_{rq}} = 1$.

We know that $\frac{\partial C_{11}}{\partial A_{11}} = B_{11}$. Therefore, $\frac{\partial E}{\partial A_{11}} = B_{11} + 0 * 8 = B_{11} = 4$.
Similarly, $\frac{\partial E}{\partial A_{12}} = B_{12} + B_{11} + 0 * 7 = 4 + 3 = 7$, $\frac{\partial E}{\partial A_{13}} = B_{13} + B_{12} + B_{11} + 0 * 6 = 4 + 3 + 3 = 10$, $\frac{\partial E}{\partial A_{14}}$ ......

The result matrix is

4	7	10	6	3
9	17	25	16	8
11	23	34	23	11
7	16	24	17	8
2	6	9	7	3

$\boldsymbol{A} =$

4	5	2	2	1
3	3	2	2	4
4	3	4	1	1
5	1	4	1	2
5	1	3	1	4

$\boldsymbol{B} =$

4	3	3
5	5	5
2	4	3

Table 1: Image Matrix (5×5) and a convolution kernel (3×3).

DS-GA 1008: Deep Learning, Spring 2020
Homework Assignment 2

Due: Friday, March 6th, 2020 at 11:59pm

1.2. Pooling

Pooling is a technique for sub-sampling and comes in different flavors, for example max-pooling, average pooling, LP-pooling.

(a) List the `torch.nn` modules for the 2D versions of these pooling techniques and read what they do.
MaxPool2d, AvgPool2d, FractionalMaxPool2d, LPPool2d, AdaptiveMaxPool2d, AdaptiveAvgPool2d.

(b) Denote the k-th input feature maps to a pooling module as $\boldsymbol{X}^k \in \mathbb{R}^{H_{\text{in}} \times W_{\text{in}}}$ where H_{in} and W_{in} represent the input height and width, respectively. Let $\boldsymbol{Y}^k \in \mathbb{R}^{H_{\text{out}} \times W_{\text{out}}}$ denote the k-th output feature map of the module where H_{out} and W_{out} represent the output height and width, respectively. Let $S^k_{i,j}$ be a list of the indexes of elements in the sub-region of X^k used for generating $\boldsymbol{Y}^k_{i,j}$, the (i,j)-th entry of $\boldsymbol{Y}^k$. Using this notation, give formulas for $\boldsymbol{Y}^k_{i,j}$ from three pooling modules.
Max Pooling: $\boldsymbol{Y}^k_{i,j} = max(X^k_{l,m} | l, m \in S^k_{i,j})$
Average Pooling: $\boldsymbol{Y}^k_{i,j} = \frac{1}{|S^k_{i,j}|} \sum_{l,m \in S^k_{i,j}} X^k_{l,m}$
L^P pooling: $\boldsymbol{Y}^k_{i,j} = (\sum_{l,m \in S^k_{i,j}} (X^k_{l,m})^p)^{\frac{1}{p}}$

(c) Write out the result of applying a max-pooling module with kernel size of 2 and stride of 1 to $\boldsymbol{C}$ from Part 1.1.

$C =$

109	92	72
108	85	74
110	74	79

Output $=$

109	92
110	85

(d) Show how max-pooling and average pooling can be expressed in terms of LP-pooling.
Max Pooling: $\boldsymbol{Y}^k_{i,j} = (\sum_{l,m \in S^k_{i,j}} (X^k_{l,m})^p)^{\frac{1}{p}}$ where $p = \infty$.
Average Pooling: $\boldsymbol{Y}^k_{i,j} = \frac{1}{|S^k_{i,j}|} (\sum_{l,m \in S^k_{i,j}} (X^k_{l,m})^p)^{\frac{1}{p}}$ where $p = 1$.
In conclusion, max pooling is L^∞ pooling and average pooling is proportional to L^1 pooling, i.e. average pooling output $\boldsymbol{Y}^k_{i,j} = \frac{L^1}{|S^k_{i,j}|}$.

DS-GA 1008: Deep Learning, Spring 2020
Homework Assignment 2

Due: Friday, March 6th, 2020 at 11:59pm

2. PyTorch

If you haven't already, install most recent versions of Python (3.6 or higher), PyTorch (we recommend using conda for the installation), and Jupyter.

Complete the programming exercises provided in the homework 2 directory of the course Google Drive folder (link).

3. Evaluation

Homework is worth a total of 100 points.

- Part 1 - 50 points
- Part 2 - 50 points

4. Submission

You are required to write up your solutions to Part 1 using markdown or LaTeX.
Please submit the homework on the NYU classes assignment page. Please upload the following:

- `First-name_Last-name_netID_A2.tex` (or `.md`) file for Part 1
- `First-name_Last-name_netID_A2.pdf` file for Part 1
- `First-name_Last-name_netID_A2_code.ipynb` file for Part 2
- `First-name_Last-name_netID_A2_code.pdf` file for Part 2
  (you can use Jupyter Notebook's "File → Download as → PDF" feature)

5. Disclaimers

You are allowed to discuss problems with other students in the class but have to write up your solutions on your own.
As feedback might be provided during the first days, the current homework assignment might be undergoing some minor changes. We'll notify you if this happens.