Quiz-06

Started: Oct 5 at 9:47pm

Quiz Instructions

Question 1 1 pts

A convolution layer of a 2D CNN operates on the following input (the input is two channels of 4x4, represented by the two 4x4 matrices, where the left matrix represents the first channel and the right matrix represents the second channel):

$$Y = \begin{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \end{bmatrix}$$

with the following two convolutional filters:

$$W_1 = \begin{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \end{bmatrix}$$

$$W_2 = \begin{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{bmatrix}$$

The filters operate on input Y to compute the output Z. The convolution uses a stride of 2.

What is the value at position (2,2) of the second channel of Z (i.e. $Z(2,2,2)$ where the first index represents the channel)? The indexing is 1-based (so indices start at 1).

Hint: Lecture 10, Slides 40-65

2

Question 2 1 pts

A convolution layer of a 2D CNN operates on the following input (the input is two channels of 4x4, represented by the two 4x4 matrices, where the left matrix represents the first channel and the right matrix represents the second channel):

$$Y = \left[\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}\right]$$

with the following two convolutional filters:

$$W_1 = \left[\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right]$$

$$W_2 = \left[\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right]$$

The filters operate on input Y to compute the output Z. The convolution uses a stride of 1. No zero padding is employed in the forward pass.

During the backward pass you compute dZ (the derivative of the divergence with respect to Z) as

$$dZ = \left[\begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix}\right]$$

During the backward pass you compute the derivative for the filters dW. What is the (2,2) element of the derivative for the second channel of the second filter (i.e. $dW_2(2, 2, 2)$ where the first index is the channel index)? The indexing is 1-based (so indices start at 1).

Hint: Lec 11, 166-199. For code refer to slides 202-207. Drawing the data flow diagram may be useful. Also, feel free to use Python for calculations.

Question 3 1 pts

A convolution layer of a 2D CNN operates on the following input (the input is two channels of 4x4, represented by the two 4x4 matrices, where the left matrix represents the first channel and the right matrix represents the second channel):

$$Y = \left[\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}\right]$$

with the following two convolutional filters:

$$W_1 = \left[\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right]$$

$$W_2 = \left[\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right]$$

The filters operate on input Y to compute the output Z. The convolution uses a stride of 2. No zero padding is employed in the forward pass.

During the backward pass you compute dZ (the derivative of the divergence with respect to Z) as

$$dZ = \left[\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\right]$$

During the backward pass you compute the derivative for the filters dW. What is the (2,2) element of the derivative for the first channel of the first filter (i.e. $dW_1(1, 2, 2)$, where the first index is the channel index)? The indexing is 1-based (so indices start at 1).

Hint: Lec 11, 166-198, but you will have to adapt the equations to 1-D convolution.

Question 4 1 pts

A convolution layer of a 1D CNN (a TDNN) operates on the following input:

$$Y = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

with the following three convolutional filters

$$W_1 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$W_3 = \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix}$$

The filters operate on input Y to compute the output Z. The convolution uses a stride of 1. No zero padding is employed in the forward pass.

During the backward pass you compute dZ (the derivative of the divergence with respect to Z) as

$$dZ = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 2 & 1 & 2 \end{bmatrix}$$

During the backward pass you compute the derivative for the input, dY. What is the first element of the derivative for the second channel of Y (i.e. $dY(2,1)$, where the first index represents channel)? The indexing is 1-based (so indices start at 1).

Hint: Lec 11, 57-160. Drawing the data flow diagram may be useful.

Question 5 1 pts

A convolution layer of a 2D CNN operates on the following input (the input is two channels of 4x4, represented by the two 4x4 matrices, where the left matrix represents the first channel and the right matrix represents the second channel):

$$Y = \begin{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} & \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \end{bmatrix}$$

with the following two convolutional filters:

$$W_1 = \begin{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} & \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \end{bmatrix}$$

$$W_2 = \begin{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{bmatrix}$$

The filters operate on input Y to compute the output Z. The convolution uses a stride of 2. No zero padding is employed in the forward pass.

During the backward pass you compute dZ (the derivative of the divergence with respect to Z) as

$$dZ = \left[\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\right]$$

During the backward pass you compute the derivative for the input, dY. What is the (3,3) element of the derivative for the first channel of Y (i.e. $dY(1,3,3)$, where the first index represents channel)? The indexing is 1-based (so indices start at 1).

Hint: Lec 11, 166-199. For code refer to slides 202-207. Drawing the data flow diagram may be useful.

Question 6 1 pts

A convolution layer of a 2D CNN operates on the following input (the input is two channels of 3x3, represented by the two 3x3 matrices, where the left matrix represents the first channel and the right matrix represents the second channel):

$$Y = \left[\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\right]$$

with the following two convolutional filters:

$$W_1 = \left[\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right]$$

$$W_2 = \left[\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right]$$

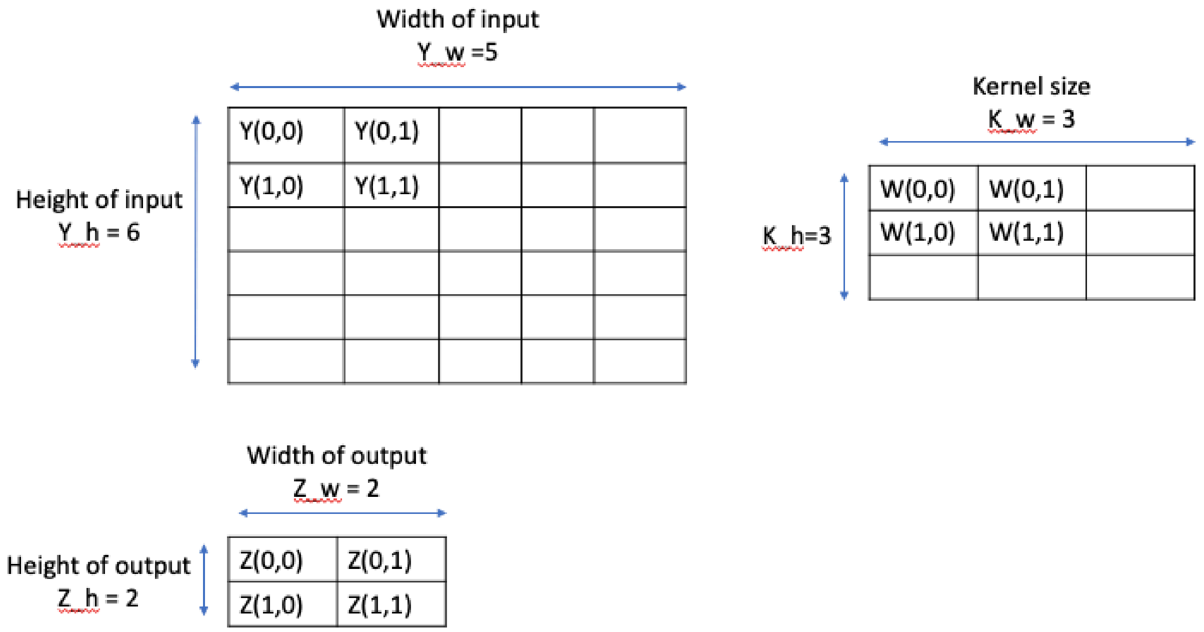
The filters operate on input Y to compute the output Z. The convolution uses a stride of 0.5 (i.e. an upsampling “transpose convolution” that results in an output twice the width and twice the height of the input). Assume the symmetric padding of input channels used in the lecture to compute this answer.

What is the value at position (2,2) of the second channel of Z (i.e. $Z(2,2,2)$ where the first index represents channel)? The indexing is 1-based (so indices start at 1).

Hint: A transpose convolution is an upsampling followed by convolution. The question is asking about the output of the convolution layer at a single position.

Question 7 1 pts

Consider a 2-D convolution setup. Let the input be Y with height (h) = 6, width (w) =5 and in-channels (m_i) = 1. Let there be only 1 filter W with height (K_h) = 3 and width = K_w = 3. Let the stride along the width (s_w) = 2 and stride along height (s_h) = 3. Let the output of this convolution operation be Z.



You are given the gradient of the loss wrt Z i.e dL/dZ. Construct a new filter F which when convolved with the input Y (stride 1) produces an output A of shape same as filter W and has A(i,j) = dL/dW(i,j)

NOTE: Height is the first dimension and width is the second dimension. We are using 0-based indexing.

The notation is used below is as follows:

Y(i,j) refers to the pixel at ith row and jth column. Similar convention for Z, W, A and F

- [ ] $Z(i,j) = \sum_{i'=0}^{K_h-1} \sum_{j'=0}^{K_w-1} W(i',j')Y(s_h * i + i', s_w * j + j')$

- [ ] $\frac{dL}{dW(1,1)} = \frac{dL}{dZ(0,0)}Y(1,1) + \frac{dL}{dZ(1,0)}Y(4,1) + \frac{dL}{dZ(0,1)}Y(1,3) + \frac{dL}{dZ(1,1)}Y(4,3)$

- [ ] Shape of proposed filter F is (4,3)

Question 8 1 pts

The input to a 2D convolutional layer is size 28x28, with 32 input channels. We wish to perform a convolution using kernels of size 5x5 on it to obtain 32 output channels. No bias is employed.

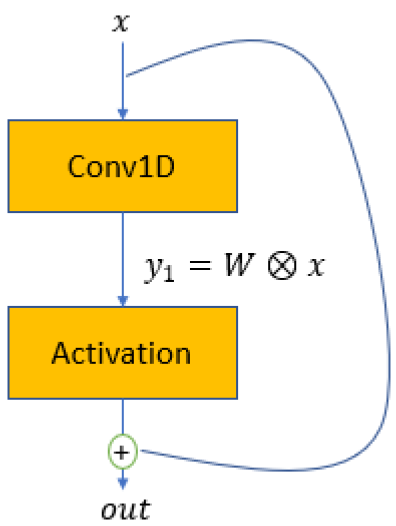
What is the difference in the number of parameters to be learned between a conventional convolutional layer, and a depthwise separable convolutional layer?

Hint: Lecture 12, Slides 83-92

- The regular convnet needs 1024 parameters while the depthwise separable convolution needs 800 parameters
- The regular convnet and the depthwise separable network both need 800 parameters, but they are distributed differently across filters and layers
- A regular convnet requires 25600 parameters while a depthwise separable convolution needs 1824
- The regular convnet and the depthwise separable convolution both need 1824 parameters

Question 9 1 pts

The figure below shows one residual block in a ResNet. Here W represents the filters of the convolution and the $\otimes$ symbol represents convolution.



Let the activation function be σ() and its derivative be σ'().

What would be the derivative of the divergence of the above residual unit with respect to the input x?

fud is short for flip up and down. flr is the flip left to right. In the answers below x_{shift} refers to the zeropadded version of x

$\frac{ddiv}{dx} = \frac{ddiv}{dout}(1 + \sigma'(W \otimes x) \cdot (fud(flr(W)) \otimes x_{shift}))$

$\frac{ddiv}{dx} = \frac{ddiv}{dout}(1 + \sigma'(fud(flr(W)) \otimes x_{shift}) \cdot (W \otimes x))$

$\frac{ddiv}{dx} = \frac{ddiv}{dout}(x + \sigma'(W \otimes x) \cdot (fud(flr(W)) \otimes x_{shift}))$

$\frac{ddiv}{dx} = \frac{ddiv}{dout}(\sigma'(W \otimes x) \cdot (fud(flr(W)) \otimes x_{shift}))$

Question 10 1 pts

While backpropagating through a mean pooling layer, the divergence derivative at the input location of the non maximum values is

Hint: Lecture 12, Slides 24-28

0

-1

Proportional to the value of the input at that location

Identical to the derivative at the input location of the maximum value

Quiz saved at 12:21pm Submit Quiz