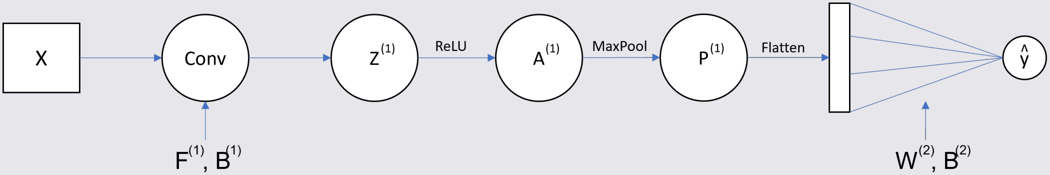
Task 3.1: CNNs - Forward and backward pass in a CNN

Given



- Input $X = \begin{bmatrix} 2 & 0 & 0 & 2 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 4 \\ 0 & 4 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 & 6 \end{bmatrix}$, filter $F^{(1)} = \begin{bmatrix} 0.5 & 1 \\ 0 & -0.5 \end{bmatrix}$ (stride 2, no padding), bias $B^{(1)} = -1$
- Weight matrix $W^{(2)} = \begin{bmatrix} 0.1 & 0.3 & -0.8 & 0.2 \end{bmatrix}$, bias $B^{(2)} = 0.7$, label $y = 1$
- Sigmoid activation function $f(z) = \frac{1}{1+e^{-z}}$ with derivation $f'(z) = f(z)(1 - f(z))$ in fully connected layer
- Cost function $C(w) = \frac{1}{N}\sum_{n=1}^{N} C^n$ where $C^n(w) = -(y^n \ln(\hat{y}^n) + (1 - y^n)\ln(1 - \hat{y}^n))$
- Learning rate $\alpha = 0.5$, MaxPool has stride 2 and no padding

Task 3.1: CNNs - Forward and backward pass in a CNN

Find

a) Compute output $\hat{y}$ (forward pass) and the corresponding loss.

b) Update weight matrix $W^{(2)}$, filter $F^{(1)}$, and biases $B^{(1)}$ and $B^{(2)}$ (backward pass). During backpropagation, also give the transposed convolutional matrix.

c) Assume that the input X was the output $P^{(0)}$ of a previous convolutional layer in a larger architecture, and compute $\frac{\partial C}{\partial P^{(0)}}$ to continue with the backpropagation.

Solution

a) The first step of the forward pass is the convolution of X using filter $F^{(2)}$ and bias $B^{(1)}$:

$$Z^{(1)} = \begin{bmatrix} 2 & 0 & 0 & 2 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 4 \\ 0 & 4 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 & 6 \end{bmatrix} \circledast \begin{bmatrix} 0.5 & 1 \\ 0 & -0.5 \end{bmatrix} - 1 = \begin{bmatrix} -2 & -1 & 1 & 0 \\ 2 & 1 & -1 & -3 \\ -1 & 0 & -1 & 3 \\ 3 & 1 & -1 & 2 \end{bmatrix}.$$

If the total task is expected to take longer than 10 minutes, this part of the forward pass can be given already. Convolution is also needed during backpropagation, so the operation will be performed again.

Task 3.1: CNNs - Forward and backward pass in a CNN

Solution (cont.)

As an example, the first entry of $Z^{(1)}$ is calculated as

$$Z_{11}^{(1)} = 2 \cdot 0.5 + 0 \cdot 1 + 0 \cdot 0 + 4 \cdot (-0.5) - 1 = -2.$$

Using ReLU as the activation function, all negative entries in $Z^{(1)}$ are mapped to zero

$$A^{(1)} = \mathrm{ReLU}(Z^{(1)}) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 \\ 3 & 1 & 0 & 2 \end{bmatrix}$$

and lastly, the MaxPool operation with stride 2 is applied without padding

$$P^{(1)} = \mathrm{MaxPool}(A^{(1)}) = \begin{bmatrix} 2 & 1 \\ 3 & 3 \end{bmatrix}$$

and $P^{(1)}$ is flattened to $S = \begin{bmatrix} 2 & 1 & 3 & 3 \end{bmatrix}^\top$ to be used as an input to the fully connected layer.

Task 3.1: CNNs - Forward and backward pass in a CNN

Solution (cont.)

Using the weight matrix W^2 and bias $B^{(2)}$ for the fully connnected layer, we calculate

$$W^{(2)}S + B^{(2)} = \begin{bmatrix} 0.1 & 0.3 & -0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 & 3 \end{bmatrix}^\top + 0.7 = -1.3$$

and we find the activated output of the fully connected layer, using the sigmoid activation function, to be

$$\hat{y} = f(-1.3) = 0.21.$$

Subsequently, the loss can be computed as

$$C(w) = -(y \ln(\hat{y}) + (1-y)\ln(1-\hat{y})) = -(1 \cdot \ln(0.21) + (1-1)\ln(1-0.21)) = 1.56.$$

b) We are now ready to begin with the backpropagation. First, we propagate the output $\hat{y}$ back through the fully connected layer to obtain

$$\frac{\partial C}{\partial W^{(2)}} = \frac{\partial C}{\partial Z^{(2)}} \frac{\partial Z^{(2)}}{\partial W^{(2)}} = (\hat{y} - y) \cdot S^\top = (0.21 - 1) \cdot \begin{bmatrix} 2 & 1 & 3 & 3 \end{bmatrix} = \begin{bmatrix} -1.58 & -0.79 & -2.37 & -2.37 \end{bmatrix} \text{ and}$$

Task 3.1: CNNs - Forward and backward pass in a CNN

Solution (cont.)

$$\frac{\partial C}{\partial B^{(2)}} = \frac{\partial C}{\partial Z^{(2)}} = (\hat{y} - y) = -0.79.$$

We can now update

$$W^{(2)} := W^{(2)} - \alpha \frac{\partial C}{\partial W^{(2)}} = W^{(2)} = \begin{bmatrix} 0.1 & 0.3 & -0.8 & 0.2 \end{bmatrix} - 0.5 \begin{bmatrix} -1.58 & -0.79 & -2.37 & -2.37 \end{bmatrix}$$

$$= \begin{bmatrix} 0.89 & 0.695 & 0.385 & 1.385 \end{bmatrix}$$

and

$$B^{(2)} := B^{(2)} - \alpha \frac{\partial C}{\partial B^{(2)}} = 0.7 - 0.5 \cdot (-0.79) = 1.095.$$

To continue the backpropagation through the preceding layer, we also need

$$\frac{\partial C}{\partial S} = {W^{(2)}}^{\top} \cdot \frac{\partial C}{\partial Z^{(2)}} = \begin{bmatrix} 0.1 & 0.3 & -0.8 & 0.2 \end{bmatrix}^{\top} \cdot 0.21 = \begin{bmatrix} 0.02 & 0.063 & -0.168 & 0.042 \end{bmatrix}^{\top}.$$

Task 3.1: CNNs - Forward and backward pass in a CNN

Solution (cont.)

In order to go backwards through the flatten operation and obtain $\frac{\partial C}{\partial P^{(1)}}$, only a reshape operation is needed:

$$\frac{\partial C}{\partial P^{(1)}} = \text{reshape}\left(\frac{\partial C}{\partial S}, \text{shape}(P^{(1)})\right) = \begin{bmatrix} 0.02 & 0.063 \\ -0.168 & 0.042 \end{bmatrix}.$$

Next, we compute

$$\frac{\partial C}{\partial A^{(1)}} = \begin{bmatrix} 0 & 0 & 0.063 & 0 \\ 0.02 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.168 \\ 0.042 & 0 & 0 & 0 \end{bmatrix}$$

by reversing the MaxPool operation, placing the values of $\frac{\partial C}{\partial P^{(1)}}$ where the maximum values in $A^{(1)}$ were located.

Task 3.1: CNNs - Forward and Backward pass in a CNN

Solution (cont.)

With the ReLU derivation

$$\frac{\partial A^{(1)}}{\partial Z^{(1)}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

(1 at all places where $Z^{(1)}$ has positive values), we can find

$$\frac{\partial C}{\partial Z^{(1)}} = \frac{\partial C}{\partial A^{(1)}} \frac{\partial A^{(1)}}{\partial Z^{(1)}} = \begin{bmatrix} 0 & 0 & 0.063 & 0 \\ 0.02 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.168 \\ 0.042 & 0 & 0 & 0 \end{bmatrix} \odot \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0.063 & 0 \\ 0.02 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.168 \\ 0.042 & 0 & 0 & 0 \end{bmatrix}.$$

The backpropagation through the convolutional layer can then be found as

$$\frac{\partial C}{\partial F^{(1)}} = X \circledast \frac{\partial C}{\partial Z^{(1)}} = \begin{bmatrix} 2 & 0 & 0 & 2 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 4 \\ 0 & 4 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 & 6 \end{bmatrix} \circledast \begin{bmatrix} 0 & 0 & 0.063 & 0 \\ 0.02 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.168 \\ 0.042 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -0.298 \\ 0.294 & 0.04 \end{bmatrix}.$$

Solution (cont.)

Alternatively, the same computation can be expressed using a “zero-padded” convolutional matrix (zero padding in a different way than adding a border of zeroes), which expresses the filter $\frac{\partial C}{\partial Z^{(1)}}$ as a 25×4-dimensional matrix:

$$\begin{bmatrix} 0 & 0 & 0.063 & 0 & 0 & 0.02 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.168 & 0 & 0.042 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.063 & 0 & 0 & 0.02 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.168 & 0 & 0.042 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.063 & 0 & 0 & 0.02 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.168 & 0 & 0.042 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.063 & 0 & 0 & 0.02 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.168 & 0 & 0.042 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Multiplied with the input matrix as a 25×1 vector,

$$\begin{bmatrix} 2 & 0 & 0 & 2 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 4 & 0 & 4 & 0 & 0 & 0 & 7 & 0 & 0 & 0 & 6 \end{bmatrix}^\top$$

this yields the same calculation as the convolution in the previous slide. The output is of size 16×1 and needs to be reshaped to 4×4 afterwards.

In the case where the filter is larger than the input and a transposed convolution needs to be computed to upsample the input, this matrix can be transposed (then called the **transposed convolution matrix**), and multiplication with the input then yields the transposed convolution operation.

Solution (cont.)

Finally, the updates of $F^{(1)}$ and $B^{(1)}$ are

$$F^{(1)} := F^{(1)} - \alpha \frac{\partial C}{\partial F^{(1)}} = \begin{bmatrix} 0.5 & 1 \\ 0 & -0.5 \end{bmatrix} - 0.5 \cdot \begin{bmatrix} 0 & -0.298 \\ 0.294 & 0.04 \end{bmatrix} = \begin{bmatrix} 0.5 & 1.149 \\ 0.147 & -0.52 \end{bmatrix}$$

$$B^{(1)} := B^{(1)} - \alpha \frac{\partial C}{\partial B^{(1)}} = \sum \frac{\partial C}{\partial Z^{(1)}} = -0.043.$$

c) Assuming that X was in fact the output $P^{(0)}$ of a previous convolutional layer in a larger architecture, we can compute

$$\frac{\partial C}{\partial P^{(0)}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & -1 & 1 & 0 & 0 \\ 0 & 2 & 1 & -1 & -3 & 0 \\ 0 & -1 & 0 & -1 & 3 & 0 \\ 0 & 3 & 1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \circledast \begin{bmatrix} -0.5 & 0 \\ 1 & 0.5 \end{bmatrix} = \begin{bmatrix} -1 & -2.5 & -0.5 & 1 & 0 \\ 1 & 3.5 & 1 & 3 & -3 \\ -0.5 & 2 & -1 & 1 & 4.5 \\ 1.5 & 4 & 0.5 & 0.5 & 0.5 \\ 0 & -1.5 & -0.5 & 0.5 & -1 \end{bmatrix}$$

as the convolution of the zero-padded $Z^{(1)}$ with $F^{(1)}$ flipped by 180 degrees.

Task 3.1: CNNs - Forward and backward pass in a CNN

Notes on Task 3.1

The workload in this task can be easily adapted to vary the length it will take to work on it. For example:

- Give the forward pass through the convolutional layer, as convolution will be performed later, and can therefore be tested there.
- Give the first three values in $\frac{\partial C}{\partial W^{(2)}}$ and the update of $W^{(2)}$ (but not the formulas to get there).
- Leave out the transposed convolutional matrix, or count as extra task.
- Give the update of $F^{(1)}$ expect for one value to be filled in.
- The computations of $\frac{\partial C}{\partial Z^1}$ and $\frac{\partial C}{\partial P^{(0)}}$ can not be given to shorten the task however, as giving the structure of the output would already give away the way to the solution.

References for this task were the lecture slides and videos from the Coding Lane Youtube channel.