HW 4 Solution

1 Dimensions (8 points)

$$
\begin{aligned}
N_r^C &= N_c^C = \frac{7-3}{1} + 1 = 5 \ , \qquad \text{(Stride of 1 and } 3 \times 3 \text{ filter)} \\
N_r^A &= N_c^A = 5 \ , \qquad \text{(No change from convolution to activation layer)} \\
N_r^P &= N_c^P = \frac{5-3}{2} + 1 = 2 \ , \qquad \text{(Stride of 2 and } 3 \times 3 \text{ neighborhood)} \\
N^{\vec{P}} &= 2 \times 2 = 4 \ , \qquad \text{(Flattening a } 2 \times 2 \text{ matrix)} \ .
\end{aligned}
$$

Correctly calculating each dimension has **2 points**.

2 Forward Propagation (10 points)

$$
C[r][c] = \sum_{i=1}^{3} \sum_{j=1}^{3} X[r-1+i][c-1+j] W^x[i][j] + W_0^x[r][c] \ .
$$

By using the above formula, we obtain

$$
C = \begin{bmatrix} 0 & 0.8 & 0.7 & 0.6 & 0 \\ 0 & 0.8 & 0.7 & 0.6 & 0 \\ 0 & 0.8 & 0.7 & 0.6 & 0 \\ 0 & 0.8 & 0.7 & 0.6 & 0 \\ 0 & 0.8 & 0.7 & 0.6 & 0 \end{bmatrix} .
$$

Since activation layer only sets the values of negative elements of convolution layer to zero, it will not change C, i.e.,

$$
A = \begin{bmatrix} 0 & 0.8 & 0.7 & 0.6 & 0 \\ 0 & 0.8 & 0.7 & 0.6 & 0 \\ 0 & 0.8 & 0.7 & 0.6 & 0 \\ 0 & 0.8 & 0.7 & 0.6 & 0 \\ 0 & 0.8 & 0.7 & 0.6 & 0 \end{bmatrix} .
$$

For the pooling layer we have

$$
P[r][c] = \max_{\substack{1 \le i \le 3 \\ 1 \le j \le 3}} A[(r-1) \times 2 + i][(c-1) \times 2 + j]
$$

which gives

$$
P = \begin{bmatrix} 0.8 & 0.7 \\ 0.8 & 0.7 \end{bmatrix} .
$$

The fully connected layer converts the output of the pooling layer to a vector:

$$\vec{P} = \begin{bmatrix} 0.8 \\ 0.7 \\ 0.8 \\ 0.7 \end{bmatrix} .$$

Then, the output $\hat{y}$ can be obtain as follows

$$\hat{y} = \sigma((W^o)^T \vec{P} + W_0^o) = \sigma(0.55) = 0.6341 .$$

Forward propagation has **10 points**, **2 points** for each stage of the network.

3 Back-Propagation (20 points)

For squared loss we have

$$\Delta y = \hat{y} - y = -0.3659 .$$

Since we have sigmoid function at the output, for the back-propagation to the fully-connected layer we have

$$\begin{aligned} \nabla W_0^o &= \hat{y}(1-\hat{y})\Delta y = -0.0849 , \\ \nabla W^o &= \hat{y}(1-\hat{y})\vec{P}\Delta y = [-0.00679 \quad -0.00594 \quad -0.00679 \quad -0.00594]^T , \\ \nabla \vec{P} &= \hat{y}(1-\hat{y}) W^o \Delta y = [-0.01698 \quad -0.00849 \quad -0.02547 \quad -0.03396]^T . \end{aligned}$$

For back-propagation to the pooling layer, we only need to convert the vector $\nabla \vec{P}$ to a 2×2 matrix. Hence,

$$\nabla P = \begin{bmatrix} -0.01698 & -0.00849 \\ -0.02547 & -0.03396 \end{bmatrix} .$$

For the activation layer, by assuming that when multiple entries have the max value we select the central pixel, we get

$$\nabla A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -0.01698 & -0.00849 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -0.02547 & -0.03396 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} .$$

For the convolution layer, we use the property of the ReLU activation function, for which we have

$$\nabla C[r][c] = \begin{cases} \nabla A[r][c] & \text{if } C[r][c] > 0 \\ 0 & \text{else} \end{cases}$$

Since all the elements of C are positive, we get $\nabla C = \nabla A$. For the bias term of the input layer, we know that $\nabla W_0^x = \nabla C$, i.e.,

$$\nabla W_0^x = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -0.01698 & -0.00849 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -0.02547 & -0.03396 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} .$$

For the weight filter of the input layer, we have

$$\nabla W^x[i][j] = \sum_{r=1}^{5}\sum_{c=1}^{5} X[r+i-1][c+j-1]\nabla C[r][c] \ .$$

By calculating the above formula for any $i \in \{1,2,3\}$ and $j \in \{1,2,3\}$ we obtain

$$\nabla W^x = \begin{bmatrix} 0 & -0.04245 & -0.04245 \\ 0 & -0.04245 & -0.04245 \\ 0 & -0.04245 & -0.04245 \end{bmatrix} \ .$$

Each stage of the network has **4 points**.

4 Update Weights (4 points)

Update rule is

$$\text{new weight} = \text{old weight} - 0.5 \times \text{gradient of weight}$$

Therefore, we will get

$$W^o_{\text{new}} = \begin{bmatrix} 0.2034 \\ 0.1030 \\ 0.3034 \\ 0.4030 \end{bmatrix} \ .$$

$$W^o_{0,\text{new}} = -0.1576 \ .$$

$$W^x_{\text{new}} = \begin{bmatrix} 0.2 & 0.1212 & 0.3212 \\ 0.1 & 0.2212 & 0.4212 \\ 0.3 & 0.4212 & 0.1212 \end{bmatrix} \ .$$

$$W^x_{0,\text{new}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0.00849 & 0.00424 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0.01273 & 0.01698 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \ .$$

Each update has **1 point**.

5 New Output (8 points)

$$C[r][c] = \sum_{i=1}^{3}\sum_{j=1}^{3} X[r-1+i][c-1+j]W^x[i][j] + W^x_0[r][c] \ .$$

By using the above formula, we obtain

$$C = \begin{bmatrix} 0 & 0.8636 & 0.7636 & 0.6 & 0 \\ 0 & 0.8636 & 0.7636 & 0.6 & 0 \\ 0 & 0.8636 & 0.7636 & 0.6 & 0 \\ 0 & 0.8636 & 0.7636 & 0.6 & 0 \\ 0 & 0.8636 & 0.7636 & 0.6 & 0 \end{bmatrix} \ .$$

Since activation layer only sets the values of negative elements of convolution layer to zero, it will not change C, i.e.,

$$A = \begin{bmatrix} 0 & 0.8636 & 0.7636 & 0.6 & 0 \\ 0 & 0.8636 & 0.7636 & 0.6 & 0 \\ 0 & 0.8636 & 0.7636 & 0.6 & 0 \\ 0 & 0.8636 & 0.7636 & 0.6 & 0 \\ 0 & 0.8636 & 0.7636 & 0.6 & 0 \end{bmatrix} .$$

For the pooling layer we have

$$P[r][c] = \max_{\substack{1\leq i\leq 3 \\ 1\leq j\leq 3}} A[(r-1)\times 2+i][(c-1)\times 2+j]$$

which gives

$$P = \begin{bmatrix} 0.8636 & 0.7636 \\ 0.8636 & 0.7636 \end{bmatrix} .$$

The fully connected layer converts the output of the pooling layer to a vector:

$$\vec{P} = \begin{bmatrix} 0.8636 \\ 0.7636 \\ 0.8636 \\ 0.7636 \end{bmatrix} .$$

Then, the output $\hat{y}$ can be obtain as follows

$$\hat{\hat{y}} = \sigma((W^o)^T\vec{P} + W_0^o) = \sigma(0.8242) = 0.6951 .$$

The value of $\hat{\hat{y}}$ is closer to 1 compared to $\hat{y}$. Also, we can calculate the loss function values:

$$\ell(y,\hat{y}) = \frac{1}{2}(y-\hat{y})^2 = 0.0669$$
$$\ell(y,\hat{\hat{y}}) = \frac{1}{2}(y-\hat{\hat{y}})^2 = 0.0465$$

Calculating the new output has **4 points**, and calculating the old and new loss function values has **2 points** each.