MathDNN Homework 4

Department of Computer Science and Engineering
2021-16988 Jaewan Park

Problem 1

The parameters for convolution are $C_{\text{in}} = 1$, $C_{\text{out}} = 2$, $F = 3$, $S = 1$, $P = 1$. Then the relationship between X and Y becomes

$$Y_{l,i,j} = \sum_{\alpha=1}^{F}\sum_{\beta=1}^{F} w_{l,\alpha,\beta} X_{i+\alpha-1,j+\beta-1} + b_l$$

Therefore,

$$Y_{1,i,j} = \sum_{\alpha=1}^{3}\sum_{\beta=1}^{3} w_{1,\alpha,\beta} X_{i+\alpha-1,j+\beta-1} + b_1 = X_{i+1,j} - X_{i,j}$$

$$Y_{2,i,j} = \sum_{\alpha=1}^{3}\sum_{\beta=1}^{3} w_{2,\alpha,\beta} X_{i+\alpha-1,j+\beta-1} + b_2 = X_{i,j+1} - X_{i,j}$$

$$\therefore w_{1,\alpha,\beta} = \begin{cases} 1 & (\alpha = 2,\ \beta = 1) \\ -1 & (\alpha = 1,\ \beta = 1) \\ 0 & (\text{otherwise}) \end{cases},\quad w_{2,\alpha,\beta} = \begin{cases} 1 & (\alpha = 1,\ \beta = 2) \\ -1 & (\alpha = 1,\ \beta = 1) \\ 0 & (\text{otherwise}) \end{cases},\quad b = 0.$$

Problem 2

A common 2D convolution has the following relationship between X and Y. (Assume the batch size is 1.)

$$Y_{l,i,j} = \sum_{\gamma=1}^{C_{\text{in}}}\sum_{\alpha=1}^{f_1}\sum_{\beta=1}^{f_2} w_{l,\gamma,\alpha,\beta} X_{\gamma,S(i-1)+\alpha,S(j-1)+\beta} + b_l$$

In the case of the given average pooling operation, $C_{\text{in}} = C_{\text{out}} = C$, $f_1 = f_2 = S = k$, $b = 0$, $P = 0$, and

$$w_{l,\gamma,\alpha,\beta} = \begin{cases} \dfrac{1}{k^2} & (\gamma = l) \\ 0 & (\text{otherwise}) \end{cases}.$$

Substituting these to the above equation gives

$$\begin{aligned} Y_{l,i,j} &= \sum_{\gamma=1}^{C}\sum_{\alpha=1}^{k}\sum_{\beta=1}^{k} w_{l,\gamma,\alpha,\beta} X_{\gamma,k(i-1)+\alpha,k(j-1)+\beta} \\ &= \frac{1}{k^2}\sum_{\alpha=1}^{k}\sum_{\beta=1}^{k} X_{l,k(i-1)+\alpha,k(j-1)+\beta} \end{aligned}$$

which equals to the given relationship where indices α, β, l correspond to a, b, c. Also,

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} - f_1 + 2P}{S} + 1 \right\rfloor = \left\lfloor \frac{m - k + 0}{k} + 1 \right\rfloor = \frac{m}{k}$$

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} - f_2 + 2P}{S} + 1 \right\rfloor = \left\lfloor \frac{m - k + 0}{k} + 1 \right\rfloor = \frac{m}{k}$$

so the input and output dimensions match with the given conditions. In this way, average pooling can be represented as a convolution.

Problem 3

Consider an unbiased convolution and set parameters $C_{\text{in}} = 3$, $C_{\text{out}} = 1$, $F = 1$, $S = 1$, $P = 0$. Then substituting these parameters gives

$$\begin{aligned} Y_{i,j} &= \sum_{\gamma=1}^{C_{\text{in}}} \sum_{\alpha=1}^{F} \sum_{\beta=1}^{F} w_{\gamma,\alpha,\beta} X_{\gamma, S(i-1)+\alpha, S(j-1)+\beta} + b_l \\ &= w_{1,1,1} X_{1,i,j} + w_{2,1,1} X_{2,i,j} + w_{3,1,1} X_{3,i,j} \end{aligned}$$

Therefore $w_{1,1,1} = 0.299$, $w_{2,1,1} = 0.587$, $w_{3,1,1} = 0.114$.

Problem 4

Two functions σ and ρ are commute as the following.

$$\begin{aligned} [\sigma(\rho(X))]_{i,j} &= \sigma\Big([\rho(X)]_{i,j}\Big) \\ &= \sigma\left(\max_{1 \le p \le \frac{m}{k},\ 1 \le q \le \frac{n}{l}} X_{\frac{m}{k}(i-1)+p, \frac{n}{l}(j-1)+q} \right) \\ &= \max_{1 \le p \le \frac{m}{k},\ 1 \le q \le \frac{n}{l}} \sigma\Big(X_{\frac{m}{k}(i-1)+p, \frac{n}{l}(j-1)+q} \Big) \quad (\because \sigma \text{ is a nondecreasing function}) \\ &= \max_{1 \le p \le \frac{m}{k},\ 1 \le q \le \frac{n}{l}} [\sigma(X)]_{\frac{m}{k}(i-1)+p, \frac{n}{l}(j-1)+q} \\ &= [\rho(\sigma(X))]_{i,j} \end{aligned}$$

Problem 6

Notation For a matrix or vector X, the notation $[X]_{i,j}$ or $[X]$ refers to the element of X at that index, and $\{f(i,j)\}_{i,j}$ refers to a matrix of which element at (i,j) is $f(i,j)$.

(a) Since $y_L = A_L y_{L-1} + b_L$, it is trivial that

$$\frac{\partial y_L}{\partial b_L} = 1, \quad \frac{\partial y_L}{\partial y_{L-1}} = A_L.$$

For $y_\ell = \sigma(A_\ell y_{\ell-1} + b_\ell)$, we can obtain the following.

$$\begin{aligned}
\frac{\partial y_\ell}{\partial b_\ell} &= \frac{\partial}{\partial b_\ell}\sigma(A_\ell y_{\ell-1} + b_\ell) \\
&= \left\{ \frac{\partial}{\partial [b_\ell]_j} [\sigma(A_\ell y_{\ell-1} + b_\ell)]_i \right\}_{i,j} = \left\{ \frac{\partial}{\partial [b_\ell]_j} \sigma([A_\ell y_{\ell-1} + b_\ell]_i) \right\}_{i,j} \\
&= \left\{ \frac{\partial}{\partial [b_\ell]_j} \sigma([A_\ell y_{\ell-1}]_i + [b_\ell]_i) \right\}_{i,j} \\
&= \left\{ \sigma'([A_\ell y_{\ell-1}]_i + [b_\ell]_i) \cdot \frac{\partial}{\partial [b_\ell]_j} ([A_\ell y_{\ell-1}]_i + [b_\ell]_i) \right\}_{i,j} \\
&= \{\sigma'([A_\ell y_{\ell-1}]_i + [b_\ell]_i) \cdot \delta_{ij}\}_{i,j} \quad (\delta_{ij}\text{: Kronecker Delta}) \\
&= \operatorname{diag}\Big(\sigma'(A_\ell y_{\ell-1} + b_\ell)\Big)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial y_\ell}{\partial y_{\ell-1}} &= \frac{\partial}{\partial y_{\ell-1}}\sigma(A_\ell y_{\ell-1} + b_\ell) \\
&= \left\{ \sigma'([A_\ell y_{\ell-1}]_i + [b_\ell]_i) \cdot \frac{\partial}{\partial [y_{\ell-1}]_j} ([A_\ell y_{\ell-1}]_i + [b_\ell]_i) \right\}_{i,j} \\
&= \left\{ \sigma'([A_\ell y_{\ell-1}]_i + [b_\ell]_i) \cdot \frac{\partial}{\partial [y_{\ell-1}]_j} \left(\sum_{k=1}^{n_{\ell-1}} [A_\ell]_{i,k} [y_{\ell-1}]_k + [b_\ell]_i \right) \right\}_{i,j} \\
&= \Big\{ \sigma'([A_\ell y_{\ell-1}]_i + [b_\ell]_i) \cdot [A_\ell]_{i,j} \Big\}_{i,j} \\
&= \operatorname{diag}\Big(\sigma'(A_\ell y_{\ell-1} + b_\ell)\Big) A_\ell
\end{aligned}$$

(b) Since $y_L = A_L y_{L-1} + b_L$,

$$\begin{aligned}
\left[\frac{\partial y_L}{\partial A_L}\right]_{1,j} &= \frac{\partial y_L}{\partial [A_L]_{1,j}} \\
&= \frac{\partial}{\partial [A_L]_{1,j}} (A_L y_{L-1} + b_L) = \frac{\partial}{\partial [A_L]_{1,j}} \left(\sum_{k=1}^{n_{\ell-1}} [A_L]_{1,k} [y_{L-1}]_k \right) \\
&= [y_{L-1}]_j
\end{aligned}$$

so $\dfrac{\partial y_L}{\partial A_L} = y_{L-1}^\intercal$. For $\ell = 1, \cdots, L-1$, we can obtain the following.

$$\begin{aligned}
\left[\frac{\partial y_L}{\partial A_\ell}\right]_{i,j} &= \frac{\partial y_L}{\partial [A_\ell]_{i,j}} = \frac{\partial y_L}{\partial y_\ell}\frac{\partial y_\ell}{\partial [A_\ell]_{i,j}} \\
&= \frac{\partial y_L}{\partial y_\ell}\left\{\frac{\partial [y_\ell]_k}{\partial [A_\ell]_{i,j}}\right\}_k \\
&= \frac{\partial y_L}{\partial y_\ell}\left\{\frac{\partial}{\partial [A_\ell]_{i,j}}\sigma([A_\ell y_{\ell-1}]_k + [b_\ell]_k)\right\}_k \\
&= \frac{\partial y_L}{\partial y_\ell}\left\{\sigma'([A_\ell y_{\ell-1}]_k + [b_\ell]_k)\left(\frac{\partial}{\partial [A_\ell]_{i,j}}([A_\ell y_{\ell-1}]_k + [b_\ell]_k)\right)\right\}_k \\
&= \frac{\partial y_L}{\partial y_\ell}\begin{bmatrix} 0 \\ \vdots \\ \sigma'([A_\ell y_{\ell-1}]_i + [b_\ell]_i)[y_{\ell-1}]_j \\ \vdots \\ 0 \end{bmatrix} \quad \text{(All elements except the } i\text{-th element are 0.)} \\
&= \left[\frac{\partial y_L}{\partial y_\ell}\right]_i \sigma'([A_\ell y_{\ell-1}]_i + [b_\ell]_i)[y_{\ell-1}]_j \\
&= [\sigma'(A_\ell y_{\ell-1} + b_\ell)]_i \left[\frac{\partial y_L}{\partial y_\ell}\right]_i [y_{\ell-1}]_j
\end{aligned}$$

Therefore $\dfrac{\partial y_L}{\partial A_\ell} = \operatorname{diag}\left(\sigma'(A_\ell y_{\ell-1} + b_\ell)\right)\left(\dfrac{\partial y_L}{\partial y_\ell}\right)^\intercal y_{\ell-1}^\intercal$.