MathDNN HW 4

2018-13260 차재현

1 Problem 1

Pad $(3-1)/2 = 1$ with value zero, and give two convolutional filters F_1, F_2 as

$$F_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, F_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

When middle entry of F_1 matches to $X_{i,j}$ the result is $X_{i+1,j} - X_{i,j} = Y_{1,i,j}$, and for the case F_2 we get $X_{i,j+1} - X_{i,j} = Y_{1,i,j}$. Thus we can define a filter w so that $w[1] = F_1, w[2] = F_2$.

2 Problem 2

Without padding, define k^2 convolutional filters $F_{i,j} = \delta_{i,j}/k^2 (1 \leq i, j \leq k)$, where $\delta_{i,j}$ is $k \times k$ matrix whose (i,j)-th entry is 1, and 0 otherwise. For single $X \in \mathbb{R}^{C \times m \times n}$, apply $F_{i,j}$ and get single $Y \in \mathbb{R}^{C \times m/k \times n/k}$. Resulting Y becomes

$$\begin{aligned} Y_{c,a,b} &= \sum_{i,j} F_{i,j} \# \begin{pmatrix} X_{c,a(k-1)+1,b(k-1)+1} & \cdots & X_{c,a(k-1)+1,bk} \\ \cdots & & \cdots \\ X_{c,ak,b(k-1)+1} & \cdots & X_{c,ak,bk} \end{pmatrix} \\ &= \frac{1}{k^2} \sum_{i,j} X_{c,a(k-1)+i,b(k-1)+j}, \end{aligned}$$

where $\sum_{i,j}$ is summation over the range of i, j given above, and the operator # between matrix is composition of elementwise multiplication with summation of elements in matrix. The result is exactly same expression as average pooling.

3 Problem 3

The meaning of 1×1 convolutional network is just multiplicating by constant, because after convolutional network $w \in \mathbb{R}$ applied, each entry X_{ij} of input X

becomes wX_{ij}. Thus, taking $w_{1,1,1} = 0.299, w_{2,1,1} = 0.587, w_{3,1,1} = 0.114$, by the same way as Problem 2 with $k = 1$ we get

$$\begin{aligned} Y_{i,j} &= w_{1,1,1}X_{1,i,j} + w_{2,1,1}X_{2,i,j} + w_{3,1,1}X_{3,i,j} \\ &= 0.299X_{1,i,j} + 0.587X_{2,i,j} + 0.114X_{3,i,j}, \end{aligned}$$

as expected.

4 Problem 4

Let ρ be maxpool sending $A \in \mathscr{M}_{p,q}(\mathbb{R})$ to its maximal entry. Suppose we choose $A = (a_{ij})_{i,j}$, and $a_{i_0 j_0}$ be maximal entry. Then $\sigma(\rho(A)) = \sigma(a_{i_0 j_0})$. Now, since σ is nondecreasing, the condition $a_{i_0 j_0}$ being maximal implies $\sigma(a_{ij}) \le \sigma(a_{i_0 j_0})$. Therefore $\rho(\sigma(A)) = \rho((\sigma(a_{ij}))_{i,j}) = \sigma(a_{i_0 j_0})$.
This can be applied to general X, by dividing X into (p, q)-submatrices.

5 Problem 6

Denote A_{ij} as (i, j)-th entry of matrix A, and y_i as i-th element of vector y. Also, $\sum_k$ means summation over all k in its range.
(a) First of all,

$$\begin{aligned} \frac{\partial y_L}{\partial b_L} &= \frac{\partial A_L y_{L-1}}{\partial b_L} + 1 = 1, \\ \frac{\partial y_L}{\partial {y_{L-1}}_i} &= \frac{\partial \sum_i A_{Li1} {y_{L-1}}_i}{\partial {y_{L-1}}_i} = A_{Li1}, \end{aligned}$$

which gives first two equality. Next, from

$$y_{l_i} = \sigma \left(\sum_j A_{l_{ij}} {y_{l-1}}_j + b_{l_i} \right),$$

$\partial y_{l_i} / \partial b_{l_j}$ is nonzero only when $i = j$. In such case, we have

$$\frac{\partial y_{l_i}}{\partial b_{l_j}} = \sigma' \left(\sum_j A_{l_{ij}} {y_{l-1}}_j + b_{l_i} \right).$$

Denote the right side by Δ_i. Combining the result, we get

$$\frac{\partial y_l}{\partial b_l} = \begin{pmatrix} \Delta_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Delta_{n_l} \end{pmatrix} = \mathrm{diag}(\sigma'(A_l y_{l-1} + b_l)).$$

Same calculation can be applied to last case. This is obtained by

$$\frac{\partial {y_l}_i}{\partial {b_l}_j} = \sigma' \left(\sum_j {A_l}_{ij} {y_{l-1}}_j + {b_l}_i \right) {A_l}_{ij},$$

which becomes (i, j)-entry of the matrix $\partial y_l / \partial y_{l-1}$. Combining with above case, we get

$$\frac{\partial y_l}{\partial y_{l-1}} = \begin{pmatrix} \Delta_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Delta_{n_l} \end{pmatrix} A_l = \mathrm{diag}(\sigma'(A_l y_{l-1} + b_l)) A_l.$$

(b) First expression can be obtained from

$$\frac{\partial y_L}{\partial A_{Li1}} = \frac{\partial}{\partial A_{Li1}} \left(\sum_j A_{Lj1} {y_{L-1}}_j + b_L \right) = {y_{L-1}}_i. \tag{1}$$

By expressing as in problem, the result is of the form $\mathbb{R}^{1 \times n_{l-1}}$, i.e. a matrix with only one row, and whose $i-th$ element is (1).

Next, differentiation of composited function gives

$$\frac{\partial y_L}{\partial {A_l}_{ij}} = \sum_k \frac{\partial y_L}{\partial {y_l}_k} \frac{\partial {y_l}_k}{\partial {A_l}_{ij}}.$$

The ${y_l}_k$ can be written explicitly as

$${y_l}_k = \sigma \left(\sum_m {A_l}_{km} {y_{l-1}}_m + b_l \right),$$

so when $i \neq k$, summand of right side becomes zero. When $i = k$, we get

$$\frac{\partial {y_l}_k}{\partial {A_l}_{ij}} = \sigma' \left(\sum_m {A_l}_{im} {y_{l-1}}_m + b_l \right) {y_{l-1}}_m. \tag{2}$$

By the same way as first case, we can express the result as

$$\frac{\partial y_L}{\partial A_l} = \mathrm{diag}\sigma'(A_l y_{l-1} + b_l) \begin{pmatrix} \partial y_L / \partial y_{l_1} \\ \vdots \\ \partial y_L / \partial y_{l_{n_l}} \end{pmatrix} \begin{pmatrix} y_{l-1_1} & \cdots & y_{l-1_{n_{l-1}}} \end{pmatrix},$$

The term $\mathrm{diag}\sigma'()$ must appear in the first, because $\sigma'()$ in the right side of (2) is independent of j. This is the desired result.