Deep Learning DS-GA 1008 Assignment2

Xiao Jing xj655

February 2019

1 Convolution

(a) What is the dimensionality of the output if we forward propagate the image over the given convolution kernel with no padding and stride of 1?

Answer: The dimension is 3x3

(b) Give a general formula of the output width O in terms of the input width I, kernel width K, stride S, and padding P (both in the beginning and in the end).

Answer:

$$O = \frac{I - K + 2P}{S} + 1$$

In the example of (a), I=5, K=3, P=0, S=1. We can get the same answer that the dimension O=3.

(c) Compute the output C of forward propagating the image over the given convolution kernel. Assume that the bias term of the convolution is zero.

Answer:

109	92	72
108	85	74
110	74	79

(d) Suppose the gradient backpropagated from the layers above this layer is a 3 3 matrix of all 1s. Write the value of the gradient (w.r.t. the input image) backpropagated out of this layer.

Answer:

Assume $a_{ij} \in A$,and $i, j \in \{1, 2, \cdots, 5\}$

$$\frac{\partial E}{\partial A} = \frac{\partial E}{\partial C}\frac{\partial C}{\partial A}$$

Since for each scalar

$$\frac{\partial E}{\partial c_{ij}} = 1$$

So $\frac{\partial E}{\partial C}$ is a 3x3 matrix

1	1	1
1	1	1
1	1	1

For each

$$\frac{\partial E}{\partial a_{\mathrm{ij}}} = 1 * \frac{\partial C}{\partial a_{\mathrm{ij}}}$$

For each $\frac{\partial C}{\partial a_{\mathrm{ij}}}$,is to figure out how many weights in the kernel apply on the a_{ij}

Since

$$\frac{\partial C}{\partial A} = \begin{bmatrix} \frac{\partial C}{\partial a_{11}} & \frac{\partial C}{\partial a_{12}} & \cdots & \frac{\partial C}{\partial a_{15}} \\ \frac{\partial C}{\partial a_{21}} & \frac{\partial C}{\partial a_{22}} & \cdots & \frac{\partial C}{\partial a_{25}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial C}{\partial a_{51}} & \frac{\partial C}{\partial a_{52}} & \cdots & \frac{\partial C}{\partial a_{55}} \end{bmatrix} = \begin{bmatrix} 4 & 7 & 10 & 6 & 3 \\ \\ 9 & 17 & 25 & 16 & 8 \\ 11 & 23 & 34 & 22 & 11 \\ 4 & 12 & 18 & 12 & 6 \\ 2 & 6 & 9 & 6 & 3 \end{bmatrix}$$

So, the backpropagation of this layer is

$$\begin{bmatrix} 4 & 7 & 10 & 6 & 3 \\ \\ 9 & 17 & 25 & 16 & 8 \\ 11 & 23 & 34 & 22 & 11 \\ 4 & 12 & 18 & 12 & 6 \\ 2 & 6 & 9 & 6 & 3 \end{bmatrix}$$

2 Pooling

(a)
1. torch.nn.MaxPool2d— will pick the maximum value over each kernel size area from the input. In regular max pooling, you downsize an input set by taking the maximum value of smaller N x N subsections of the set (often 2x2), and try to reduce the set by a factor of N, where N is an integer.

2. torch.nn.AvgPool2d—will take the average of value over each kernel size area from the input

3. torch.nn.LPPool2d—Learn-norm Pooling will apply a 2D power-average pooling over an input signal composed of several inputplanes.

4. torch.nn.AdaptiveAvgPool2d—Applies a adaptive average pooling over an input signal composed of several input planes.

5.torch.nn.FractionalMaxPool2d—Slightly differenct from maxpooling. The overall reduction ratio N does not have to be an integer.The sizes of the pooling regions are generated randomly but are fairly uniform.

(b)
1. Max Pooling 2d, when $S^k_{i,j}$ is the indices of X^k.

$$Y^k_{i,j} = \max(X^k_i n) = \max(X^k[S^k_{i,j}]) \tag{1}$$

2. Average Pooling 2d

$$Y^k_{i,j} = average(X^k_i n) = average(X^k[S^k_{i,j}]) \tag{2}$$

3. LP-Pooling 2d

$$Y^k_{i,j} = \sqrt[p]{\sum_{x \in X^k[S^k_{i,j}]k} x^p} \tag{3}$$

- x is the elements in sub-region X^k
- At $p = \infty, f(X)$ gets Max Pooling
- At p = 1, one gets Sum Pooling (which is proportional to Average Pooling)

(c)Answer:

109	92
110	85

(d)

In LP pooling
When power p = 1, one gets Sum Pooling (which is proportional to Average Pooling)

$$Y^k_{i,j} = \sqrt[p]{\sum_{x \in X^k[S^k_{i,j}]} x^p} = \sqrt[1]{\sum_{x \in X^k[S^k_{i,j}]} x} = \sum_{x \in X^k[S^k_{i,j}]} x = S^k_{i,j} count \cdot average(X^k[S^k_{i,j}]) = count\{S^k_{i,j}\} \cdot Y^k_{avg,ij} \tag{4}$$

When $p = \infty$,

$$Y^k_{i,j} = \sqrt[p]{\sum_{x \in X^k[S^k_{i,j}]} x^p} \approx \sqrt[p]{\max_{x \in X^k[S^k_{i,j}]} (x^p)} = \max(X^k[S^k_{i,j}]) \tag{5}$$