SimplePyML Convolutional Layer

Vikram Rangarajan

January 2024

1 Definitions

$\phi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is the activation function. This must be differentiable.
You must define the number of filters in this layer to determine the output, bias, and kernel sizes.
You must also define the filter shape to both determine the kernel, bias, and output shapes, as well as the dimension of the layer.
X is the input of the layer. It has the shape (channels, ...) where ... is the image/other data itself.
K is the kernel array with shape (filters, channels, ...).

2 Foward Propagation

$\phi(\text{Input} \underset{valid}{\star} \text{Kernels} + \text{Biases}) = \text{Output}$

Input: X (2 Channels, 3x3 shape $\Rightarrow$ 2x3x3 array)

$$\text{Channel 1} \begin{bmatrix} X_{0,0,0} & X_{0,0,1} & X_{0,0,2} \\ X_{0,1,0} & X_{0,1,1} & X_{0,1,2} \\ X_{0,2,0} & X_{0,2,1} & X_{0,2,2} \end{bmatrix}$$
$$\text{Channel 2} \begin{bmatrix} X_{1,0,0} & X_{1,0,1} & X_{1,0,2} \\ X_{1,1,0} & X_{1,1,1} & X_{1,1,2} \\ X_{1,2,0} & X_{1,2,1} & X_{1,2,2} \end{bmatrix}$$

Kernels: K (2 Filters, 2 Channels, 2x2 shape $\Rightarrow$ 2x2x2x2 array)

$$\begin{array}{ccc} & \text{Filter 0} & \text{Filter 1} \\ \text{Channel 1} & \begin{bmatrix} K_{0,0,0,0} & K_{0,0,0,1} \\ K_{0,0,1,0} & K_{0,0,1,1} \end{bmatrix} & \begin{bmatrix} K_{1,0,0,0} & K_{1,0,0,1} \\ K_{1,0,1,0} & K_{1,0,1,1} \end{bmatrix} \\ \text{Channel 2} & \begin{bmatrix} K_{0,1,0,0} & K_{0,1,0,1} \\ K_{0,1,1,0} & K_{0,1,1,1} \end{bmatrix} & \begin{bmatrix} K_{1,1,0,0} & K_{1,1,0,1} \\ K_{1,1,1,0} & K_{1,1,1,1} \end{bmatrix} \end{array}$$

Biases: b (2 Filters, 2x2 output shape $\Rightarrow$ 2x2x2 array)

$$\begin{array}{cc} \text{Filter 0} & \text{Filter 1} \\ \begin{bmatrix} b_{0,0,0} & b_{0,0,1} \\ b_{0,1,0} & b_{0,1,1} \end{bmatrix} & \begin{bmatrix} b_{1,0,0} & b_{1,0,1} \\ b_{1,1,0} & b_{1,1,1} \end{bmatrix} \end{array}$$

Output: Y (2 Filters, 2x2 output shape $\Rightarrow$ 2x2x2 array)

$$\begin{array}{cc} \text{Filter 0} & \text{Filter 1} \\ \begin{bmatrix} Y_{0,0,0} & Y_{0,0,1} \\ Y_{0,1,0} & Y_{0,1,1} \end{bmatrix} & \begin{bmatrix} Y_{1,0,0} & Y_{1,0,1} \\ Y_{1,1,0} & Y_{1,1,1} \end{bmatrix} \end{array}$$

3 Expanding Output Formula

$$z[i] = X \underset{valid}{\star} K[i] + b[i], \text{ for each filter } i$$

$$Y = \phi(z)$$

$$z_{0,0,0} = b_{0,0,0} + \sum_{c=0}^{1} X_{c,0,0}K_{0,c,0,0} + X_{c,0,1}K_{0,c,0,1} + X_{c,1,0}K_{0,c,1,0} + X_{c,1,1}K_{0,c,1,1}$$

$$z_{0,0,1} = b_{0,0,1} + \sum_{c=0}^{1} X_{c,0,1}K_{0,c,0,0} + X_{c,0,2}K_{0,c,0,1} + X_{c,1,1}K_{0,c,1,0} + X_{c,1,2}K_{0,c,1,1}$$

$$z_{0,1,0} = b_{0,1,0} + \sum_{c=0}^{1} X_{c,1,0}K_{0,c,0,0} + X_{c,1,1}K_{0,c,0,1} + X_{c,2,0}K_{0,c,1,0} + X_{c,2,1}K_{0,c,1,1}$$

$$z_{0,1,1} = b_{0,1,1} + \sum_{c=0}^{1} X_{c,1,1}K_{0,c,0,0} + X_{c,1,2}K_{0,c,0,1} + X_{c,2,1}K_{0,c,1,0} + X_{c,2,2}K_{0,c,1,1}$$

$$z_{1,0,0} = b_{1,0,0} + \sum_{c=0}^{1} X_{c,0,0}K_{1,c,0,0} + X_{c,0,1}K_{1,c,0,1} + X_{c,1,0}K_{1,c,1,0} + X_{c,1,1}K_{1,c,1,1}$$

$$z_{1,0,1} = b_{1,0,1} + \sum_{c=0}^{1} X_{c,0,1}K_{1,c,0,0} + X_{c,0,2}K_{1,c,0,1} + X_{c,1,1}K_{1,c,1,0} + X_{c,1,2}K_{1,c,1,1}$$

$$z_{1,1,0} = b_{1,1,0} + \sum_{c=0}^{1} X_{c,1,0}K_{1,c,0,0} + X_{c,1,1}K_{1,c,0,1} + X_{c,2,0}K_{1,c,1,0} + X_{c,2,1}K_{1,c,1,1}$$

$$z_{1,1,1} = b_{1,1,1} + \sum_{c=0}^{1} X_{c,1,1}K_{1,c,0,0} + X_{c,1,2}K_{1,c,0,1} + X_{c,2,1}K_{1,c,1,0} + X_{c,2,2}K_{1,c,1,1}$$

4 Backpropagation Calculus

4.1 Bias Gradient

Observe:

$$\frac{\partial Y_{i,j,k}}{\partial b_{i,j,k}} = \frac{\partial Y_{i,j,k}}{\partial z_{i,j,k}}\frac{\partial z_{i,j,k}}{\partial b_{i,j,k}} = \phi'(z_{i,j,k})(1) = \phi'(z_{i,j,k})$$

If we are given $\frac{\partial L}{\partial Y}$, we can solve for $\frac{\partial L}{\partial b}$:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial Y} \odot \frac{\partial Y}{\partial b} = \frac{\partial L}{\partial Y} \odot \phi'(z)$$

4.2 Kernel Gradient

We must calculate $\frac{\partial L}{\partial K}$ given $\frac{\partial L}{\partial Y}$
Observe that for any filter i and channel c:

$$\frac{\partial L}{\partial K_{i,c,0,0}} = \sum_{(x,y)\in[0,1]\times[0,1]} \frac{\partial L}{\partial Y_{i,x,y}}\frac{\partial Y_{i,x,y}}{\partial z_{i,x,y}}\frac{\partial z_{i,x,y}}{\partial K_{i,c,0,0}} = \sum_{(x,y)\in[0,1]\times[0,1]} \frac{\partial L}{\partial Y_{i,x,y}}\phi'(z_{i,x,y})X_{c,x,y} = \sum_{(x,y)\in[0,1]\times[0,1]} \frac{\partial L}{\partial b_{i,x,y}}X_{c,x,y}$$

With this same logic,

$$\frac{\partial L}{\partial K_{i,c,0,1}} = \sum_{(x,y)\in[0,1]\times[0,1]} \frac{\partial L}{\partial b_{i,x,y}}X_{c,x,y+1}$$

$$\frac{\partial L}{\partial K_{i,c,1,0}} = \sum_{(x,y)\in[0,1]\times[0,1]} \frac{\partial L}{\partial b_{i,x,y}}X_{c,x+1,y}$$

$$\frac{\partial L}{\partial K_{i,c,1,1}} = \sum_{(x,y)\in[0,1]\times[0,1]} \frac{\partial L}{\partial b_{i,x,y}}X_{c,x+1,y+1}$$

This is the pattern for a valid cross-correlation between X and $\frac{\partial L}{\partial b}$:

$$\frac{\partial L}{\partial K_{i,c,:,:}} = X_{c,:,:} \underset{valid}{\star} \frac{\partial L}{\partial b_{i,:,:}}$$

4.3 Input Gradient

We must calculate $\frac{\partial L}{\partial X}$ given $\frac{\partial L}{\partial Y}$
Observe that for any channel c:

$$\frac{\partial L}{\partial X_{c,0,0}} = \sum_{i=0}^{1}\sum_{(x,y)\in[0,1]\times[0,1]} \frac{\partial L}{\partial Y_{i,x,y}}\frac{\partial Y_{i,x,y}}{\partial z_{i,x,y}}\frac{\partial z_{i,x,y}}{\partial X_{c,0,0}} = \sum_{i=0}^{1}\sum_{(x,y)\in[0,1]\times[0,1]} \frac{\partial L}{\partial b_{i,x,y}}\frac{\partial z_{i,x,y}}{\partial X_{c,0,0}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,0,0}}K_{i,c,0,0}$$

Repeating this differentiation for all elements of X,

$$\frac{\partial L}{\partial X_{c,0,0}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,0,0}} K_{i,c,0,0}$$

$$\frac{\partial L}{\partial X_{c,0,1}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,0,0}} K_{i,c,0,1} + \frac{\partial L}{\partial b_{i,0,1}} K_{i,c,0,0}$$

$$\frac{\partial L}{\partial X_{c,0,2}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,0,1}} K_{i,c,0,1}$$

$$\frac{\partial L}{\partial X_{c,1,0}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,0,0}} K_{i,c,1,0} + \frac{\partial L}{\partial b_{i,1,0}} K_{i,c,0,0}$$

$$\frac{\partial L}{\partial X_{c,1,1}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,0,0}} K_{i,c,1,1} + \frac{\partial L}{\partial b_{i,0,1}} K_{i,c,1,0} + \frac{\partial L}{\partial b_{i,1,0}} K_{i,c,0,1} + \frac{\partial L}{\partial b_{i,1,1}} K_{i,c,0,0}$$

$$\frac{\partial L}{\partial X_{c,1,2}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,0,1}} K_{i,c,1,1} + \frac{\partial L}{\partial b_{i,1,1}} K_{i,c,0,1}$$

$$\frac{\partial L}{\partial X_{c,2,0}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,1,0}} K_{i,c,1,0}$$

$$\frac{\partial L}{\partial X_{c,2,1}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,1,0}} K_{i,c,1,1} + \frac{\partial L}{\partial b_{i,1,1}} K_{i,c,1,0}$$

$$\frac{\partial L}{\partial X_{c,2,2}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,1,1}} K_{i,c,1,1}$$

This is exactly the pattern of a full convolution between the bias gradient and the kernel array, but summed over all the filters. This following summation indicates matrix addition:

$$\Rightarrow \frac{\partial L}{\partial X_{c,:,:}} = \sum_{i=0}^{1} \frac{\partial L}{\partial b_{i,:,:}} \underset{full}{*} K_{i,c,:,:}$$