

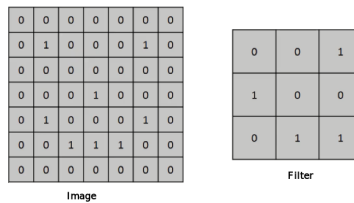
# Large Scale Machine Learning and Deep Learning

## Review Questions 6

1. Consider a CNN composed of three convolutional layers, each with  $3 \times 3$  filters, a stride of 2, and SAME padding. The lowest layer outputs 100 feature maps, the middle one outputs 200, and the top one outputs 400. The input images are RGB images of  $200 \times 300$  pixels. What is the total number of parameters  $w$  in the CNN?

**Answer:** Since its first convolutional layer has  $3 \times 3$  filters, and the input has three channels (red, green, and blue), then each feature map has  $3 \times 3 \times 3$  weights, plus a bias term. That's 28 parameters per feature map. Since this first convolutional layer has 100 feature maps, it has a total of  $28 \times 100 = 2800$  parameters. The second convolutional layer has  $3 \times 3$  kernels, and its input is the set of 100 feature maps of the previous layer, so each feature map has  $3 \times 3 \times 100 = 900$  weights, plus a bias term. Since it has 200 feature maps, this layer has  $901 \times 200 = 180200$  parameters. Finally, the third and last convolutional layer also has  $3 \times 3$  kernels, and its input is the set of 200 feature maps of the previous layers, so each feature map has  $3 \times 3 \times 200 = 1800$  weights, plus a bias term. Since it has 400 feature maps, this layer has a total of  $1801 \times 400 = 720400$  parameters. All in all, the CNN has  $2800 + 180200 + 720400 = 903400$  parameters.

2. Consider a CNN with one convolutional layer, in which it has a  $3 \times 3$  filter (as shown below) and a stride of 2. Please write the output of this layer for the given input image (the left image in the following figure)?



**Answer:** I borrowed the following figure from the submitted answers by Sumsam U. Khan and Vibhor Sharma. The figure shows how the kernel scrolls over the figure.

0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	1	0	0	0	1	0	0
0	0	1	1	1	0	0	0
0	0	0	0	0	0	0	0

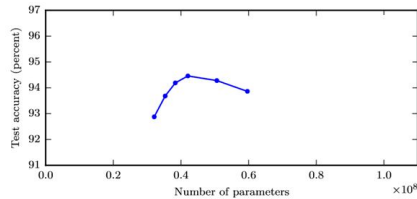
0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	1	0	0	0	1	0	0
0	0	1	1	1	0	0	0
0	0	0	0	0	0	0	0

0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0
0	1	0	0	0	1	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	1	0	0	0	1	0	0
0	0	1	1	1	0	0	0
0	0	0	0	0	0	0	0

As a result we have:

0	0	0
1	0	1
0	1	1

3. The below graph shows the accuracy of a trained 3-layer convolutional neural network vs. the number of parameters (i.e., the number of kernels). The trend suggests that as you increase the width of a neural network, the accuracy increases till a certain threshold value, and then starts decreasing. What could be the possible reason for this decrease?



- (a) Even if number of kernels increase, only few of them are used for prediction.
- (b) As the number of kernels increase, the predictive power of neural network decrease.
- (c) As the number of kernels increase, they start to correlate with each other which in turn helps overfitting.
- (d) None of these.

**Answer:** c