Deep Learning

Theoretical Exercises – Week 12 – Chapter 9

Exercises on the book "Deep Learning" written by Ian Goodfellow,
Yoshua Bengio, and Aaron Courville.
Exercises and solutions by T. Méndez and G. Schuster

FS 2024

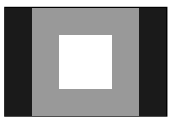
1 Exercises on Convolutional Networks

1. Convolution and Correlation:
   (a) What is the difference between convolution and correlation?
   (b) Which of the two operations is normally used in convolutional networks?
   (c) Why does it not matter whether convolution or correlation is used in convolutional networks?

2. What are the three stages of a convolutional layer? Describe each stage and explain what it is used for.

3. Given is the image

$$\boldsymbol{X} = \begin{bmatrix} 1 & 6 & 6 & 6 & 6 & 1 \\ 1 & 6 & 10 & 10 & 6 & 1 \\ 1 & 6 & 10 & 10 & 6 & 1 \\ 1 & 6 & 6 & 6 & 6 & 1 \end{bmatrix},$$



which is passed through the convolution stage with the kernel

$$\boldsymbol{K} = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

that detects vertical edges. The origin of the kernel is in the middle of the kernel.

   (a) Calculate the output of the convolution stage for all three cases of zero-padding (*valid*, *same* and *full*).
   (b) For the zero-padding case *same* also calculate the output of the detector stage and the output of the pooling stage. For the detector stage use the rectified linear activation function and for the pooling stage use max-pooling, which takes the maximum output within a (non-overlapping) neighborhood of 2×2.

4. Exercise 13.3 from the book "Digital Image Processing" Rafael C. Gonzalez and Richard E. Woods:
   Consider a CNN whose inputs are RGB color images of size 512×512 pixels. The network has two convolutional layers. Using this information, answer the following:

   (a) You are told that the spatial dimensions of the feature maps in the first layer are 504×504, and that there are 12 feature maps in the first layer. Assuming that no zero-padding is used, and that the kernels used are square, and of an odd size, what are the spatial dimensions of these kernels?

   (b) If subsampling is done using neighborhoods of size 2×2, what are the spatial dimensions of the pooled feature maps in the first layer?

   (c) What is the depth (number) of the pooled feature maps in the first layer?

   (d) The spatial dimensions of the convolution kernels in the second layer are 3×3. Assuming no zero-padding, what are the spatial dimensions of the feature maps in the second layer?

   (e) You are told that the number of feature maps in the second layer is 6, and that the size of the pooling neighborhoods is again 2×2. What is the length of the vector that results from flattening the last layer of the CNN?