Week 7 Exercise Sheet Solutions

The following exercises have different levels of difficulty indicated by (*), (**), (***). An exercise with (*) is a simple exercise requiring less time or effort to solve compared to an exercise with (***), which is a more complex exercise.

Neural Networks

1. (**) Consider a network with 1 hidden layer which can be described by the following equations:

$$\mathbf{h} = f\left(W^{(1)}\mathbf{x} + \mathbf{b}^{(1)}\right), \tag{1}$$
$$\mathbf{y} = f\left(W^{(2)}\mathbf{h} + \mathbf{b}^{(2)}\right), \tag{2}$$

where $W^{(1)}$ and $W^{(2)}$ are matrices. If the activation function is linear ($f(a) = a$), show that this is equivalent to a single layer $\mathbf{y} = W\mathbf{x} + \mathbf{b}$. Give expressions for W and $\mathbf{b}$ in terms of $W^{(1)}$, $W^{(2)}$, $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(2)}$.

Solution:
If the activation function is linear then the equations for each layer can be written as

$$\mathbf{h} = W^{(1)}\mathbf{x} + \mathbf{b}^{(1)}, \tag{3}$$
$$\mathbf{y} = W^{(2)}\mathbf{h} + \mathbf{b}^{(2)}. \tag{4}$$

Now, inserting the equation for the hidden layer (eq.3) into the equation for the output layer (eq.4) we get

$$\begin{aligned}\mathbf{y} &= W^{(2)}\left(W^{(1)}\mathbf{x} + \mathbf{b}^{(1)}\right) + \mathbf{b}^{(2)} && (5)\\ &= W^{(2)}W^{(1)}\mathbf{x} + W^{(2)}\mathbf{b}^{(1)} + \mathbf{b}^{(2)} && (6)\\ &= W\mathbf{x} + \mathbf{b}, && (7)\end{aligned}$$

where $W = W^{(2)}W^{(1)}$ and $\mathbf{b} = W^{(2)}\mathbf{b}^{(1)} + \mathbf{b}^{(2)}$. This shows that without the activation function then the network is simply performing the same operation as a single weight matrix.

---

2. (*) Consider the following input image:

$$x = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix} \tag{8}$$

If the convolutional filter is

$$W = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{9}$$

calculate the corresponding 3×3 output feature map. What feature does this filter detect?

Solution:
Applying the convolution operation to this input with this filer map gives

$$y = \begin{bmatrix} 3 & 3 & 3 \\ 1 & 3 & 2 \\ 2 & 1 & 2 \end{bmatrix}. \tag{10}$$

This filter map is detecting the presence of a diagonal line (with a width of 1 pixel). The top row contains 3 diagonal lines, hence the large output, while the other rows are weaker.

---

3. (***) If we have a 512×512 RGB colour image that we apply 100 5×5 filters with a stride of 7 and a padding of 2. What is the output volume size and how many parameters are in this layer?

Solution:
We can use the formula given in the lecture to work out the output size,

$$O = \frac{I + 2P - F}{S} + 1 \tag{11}$$

where O is the output size, I is the input size (for side of the image), F is the filter size and S is the stride. The padding is added symmetrically to the input size (where P is the padding size). So in this example $I = 512$, $F = 5$, $S = 7$ and

$P = 2$ so the output size will be

$$O = \frac{512 + 2 \times 2 - 5}{7} + 1 \tag{12}$$
$$= 74. \tag{13}$$

This is 1 dimension of the output array, so the full output tensor shape will be number of output channels by output size by output size which is

$$100 \times O \times O = 100 \times 74 \times 74. \tag{14}$$

The number of parameters will be (Filter width x filter height x number of channels in previous layer + 1) x number of output channels. The +1 comes from the bias. In this example, we have 3 input channels as it is a RGB image and 100 output channels, while the filter height and width is 5. So

$$\text{Number of parameters} = ((5 \times 5 \times 3) + 1) \times 100 \tag{15}$$
$$= 7,600. \tag{16}$$
