MLAI Week 7 Exercise: Neural Networks

Note: An indicative mark is in front of each question. The total mark is 12. You may mark your own work when we release the solutions.

$\boxed{2}$ 1. Using the definitions for $\mathbf{o}$ and $\mathbf{h}$ on slide 10 of Lecture 7 to show that if the activation function is linear such that $g(a) = a$, then the one-hidden-layer on that slide encodes a linear relationship between the input $\mathbf{x}$ and output $\mathbf{o}$. Include all steps.

Solution:

$\boldsymbol{h} = g((W^{(1)})^T)\boldsymbol{x} + b^{(1)})$

$\boldsymbol{o} = g((W^{(2)})^T)\boldsymbol{h} + b^{(2)})$

$\boldsymbol{o} = g((W^{(2)})^T)g((W^{(1)})^T)\boldsymbol{x} + b^{(1)}) + b^{(2)})$

g is defined as $g(a) = a$

$\boldsymbol{o} = (W^{(2)})^T((W^{(1)})^T)\boldsymbol{x} + b^{(1)}) + b^{(2)})$

$\boldsymbol{o} = (W^{(2)})^T(W^{(1)})^T\boldsymbol{x} + (W^{(2)})^T b^{(1)} + b^{(2)}$

Substitute $W = (W^{(2)})^T(W^{(1)})^T; b = (W^{(2)})^T b^{(1)} + b^{(2)}$

$\boldsymbol{o} = W\boldsymbol{x} + b$

$\boxed{1}$ 2. In Slide 38: we change the 3×3 kernel to $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. What will be the 3×3 convolved features? What features can this kernel detect?

Solution:

$$\text{Image} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$\text{Kernel} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Apply kernel to image in the first position (the red indicated where the kernel is placed and the kernel values at their respective positions):

$$\text{Image} = \begin{bmatrix} 1_{\times 1} & 1_{\times 0} & 1_{\times 0} & 0 & 0 \\ 0_{\times 0} & 1_{\times 1} & 1_{\times 0} & 1 & 0 \\ 0_{\times 0} & 0_{\times 0} & 1_{\times 1} & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

The Convolved Feature after the first convolution, summing all the products between the image and the kernel values:

First Convolved Feature $= 1 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 1 = 3$

$$\text{Convolved Features} = \begin{bmatrix} 3 & & \\ & & \\ & & \end{bmatrix}$$

The next step is to shift the kernel and perform the same operation:

$$\text{Image} = \begin{bmatrix} 1 & 1_{\times 1} & 1_{\times 0} & 0_{\times 0} & 0 \\ 0 & 1_{\times 0} & 1_{\times 1} & 1_{\times 0} & 0 \\ 0 & 0_{\times 0} & 1_{\times 0} & 1_{\times 1} & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Calculate the new convolved feature:

$$\text{Convolved Features} = \begin{bmatrix} 3 & 3 & \\ & & \\ & & \end{bmatrix}$$

Perform the same operation for the entire image, resulting in the final Convolved Features:

$$\text{Convolved Features} = \begin{bmatrix} 3 & 3 & 3 \\ 1 & 3 & 2 \\ 1 & 1 & 2 \end{bmatrix}.$$

The kernel can detect diagonal edges in the image.

$\boxed{3}$ 3. We have a $512 \times 512 \times 3$ colour image. We apply 100 5×5 filters with stride 7, and pad 2 to obtain a convolution output. What is the output volume size? How many parameters are needed for such a layer?

Solution:

Size of output:

Size of output: (Image Length - Filter Size + 2$\times$ Padding) / Stride + 1

Image Length = 512
Filter Size = 5
Stride = 7
Padding = 2

After applying the first 5×5 filter:

Output Size After First Filter $= (512 - 5 + 2 \times 2)/7 + 1 = 74$

Final Output Shape = Number of Filters $\times$ Output Size $\times$ Output Size

Final Output Shape $= 100 \times 74 \times 74$

Number of parameters:

Number of parameters = (Filter Width $\times$ Filter Height $\times$ Filters in Previous Layer + 1) $\times$ Number of Filters

Filter Width = 5
Filter Height = 5
Filters in Previous Layer = 3
Number of Filters = 100

Number of parameters $= (5 \times 5 \times 3 + 1) \times 100 = 7600$

$\boxed{6}$ 4. For the AlexNet depicted in Slide 35 of Lecture 6, there are about 60 million learnable parameters. With the help of the illustration `https://static.packt-cdn.com/products/9781789956177/graphics/assets/ec08175c-5282-4be2-b6e7-6f2d99272166.png`, compute the exact number of learnable parameters in AlexNet, showing the steps.

Solution:

The AlexNet consists of convolutional layers, pooling layers and fully connected layers. The pooling layer does not have any learnable parameters.

The number of parameters in the convolutional layer is:

$$W_c = K^2 \times C \times N, \tag{1}$$

where the K is the size of the kernel, C is the number of channels in the input and N is the number of kernels. In addition to the weights, there are also N bias values. The final number of parameters is $P_c = N + W_c$.

There are also two types of fully connected (FC) layer: the first is where the last pooling layer is connected to a FC layer, and the other is where a FC layer is connected to another FC layer.

The number of parameters in the first case is:

$$W_{fc} = O^2 \times N \times F, \tag{2}$$

where the O is the size of the convolved output, N is the number of kernels in the previous convolutional layer and F is the number of neurons in the layer. The convolved output is flatted to a vector of length $O \times O \times N$. In addition to the weights, there are also F bias values. The total number of parameters in this layer is $P_c = F + W_{fc}$.

In the case where a fc layer is connected to another fc layer:

$$W_{fc} = F_{-1} \times F, \tag{3}$$

where, F_{-1} is the number of neurons in the previous layer and F is the number of neurons in the current layer. The total number of parameters in this layer is $P_c = F + W_{fc}$.

For example in the first layer:

$$P_1 = 11^2 \times 3 \times 96 + 96 = 34944. \tag{4}$$

The second layer:

$$P_2 = 5^2 \times 96 \times 256 + 256 = 614656. \tag{5}$$

After performing the appropriate operations at each layer the total number of parameters in AlexNet is: $62,378,344$.

Parameters in each layer:

- Conv Layer 1: 34944
- Conv Layer 2: 614656

- Conv Layer 3: 885120
- Conv Layer 4: 1327488
- Conv Layer 5: 884992
- FC layer 1: 37752832
- FC layer 2: 16781312
- FC layer 3: 4097000