**Week 7 Exercise: Neural Networks**

**Note**: An indicative mark is in front of each question. The total mark is 10. You may mark your own work when we release the solutions.

Please answer the four questions BEFORE reading the solutions.

$\boxed{2}$ 1. Using the definitions for **o** and **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 **x** and output **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 $W = (W^{(2)})^T(W^{(1)})^T; b = (W^{(2)})^T b^{(1)} + b^{(2)}$
$\boldsymbol{o} = W\boldsymbol{x} + b$

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

**Solution:** $\begin{bmatrix} 2 & 3 & \overset{2}{\cancel{3}} \\ 1 & 3 & 3 \\ 1 & 3 & 2 \end{bmatrix}$.
This kernel can detect vertical lines.

$\boxed{2}$ 3. For the kernel in Question 2 above, a) show the convolved features with pad=1; and b) show the convolved features with stride=2.

**Solution:**
a)
$$\begin{bmatrix} 1 & 2 & 2 & 1 & 0 \\ 1 & 2 & 3 & 2 & 1 \\ 0 & 1 & 3 & 3 & 1 \\ 0 & 1 & 3 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \end{bmatrix}.$$

b)
$$\begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix}.$$

$\boxed{4}$ 4. We have a $256 \times 256$ colour image. We apply 80 $7 \times 7$ filters with stride 3, and pad 3 to obtain a convolution output. What is the output volume size? How many parameters are needed for such a layer?

**Solution:** Because it's a colour image, our image is of size $256 \times 256 \times 3$. For a single filter, we have the formula $\frac{N-F}{Stride}+1$ where $N$ is the height/length of the input dimension and $F$ is the size of the kernel.
$N = 256 + (2 \times 3) = 262$ as we add a padding of size 3 to both ends.

For a single filter we get a convolution output of size $(\frac{262-7}{3} + 1 = 86) \times 86$
For all 80 filters, we get a convolution output of size $86 \times 86 \times 80$.

The total number of parameters are
(Size of each kernel+1)$\times$Number of kernels
We add 1 for the bias for each kernel
$(7 \times 7 \times 3 + 1) \times 80 = \mathbf{11840}$