Due Date : February 4th (11pm), 2020

Instructions

- *For all questions, show your work!*
- *Use LaTeX and the template we provide when writing your answers. You may reuse most of the notation shorthands, equations and/or tables. See the assignment policy on the course website for more details.*
- *Submit your answers electronically via Gradescope.*

Question 1 (4-4-4)**.** Using the following definition of the derivative and the definition of the Heaviside step function :

$$\frac{d}{dx}f(x) = \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon} \qquad H(x) = \begin{cases} 1 & \text{if } x > 0 \\ \frac{1}{2} & \text{if } x = 0 \\ 0 & \text{if } x < 0 \end{cases}$$

1. Show that the derivative of the rectified linear unit $g(x) = \max\{0, x\}$, **wherever it exists**, is equal to the Heaviside step function.
2. Give two alternative definitions of $g(x)$ using $H(x)$.
3. Show that $H(x)$ can be well approximated by the sigmoid function $\sigma(x) = \frac{1}{1+e^{-kx}}$ asymptotically (i.e for large k), where k is a parameter.

Answer 1.

1.
$$\frac{d}{dx}g(x) = \lim_{\epsilon \to 0} \frac{g(x+\epsilon) - g(x)}{\epsilon} = \lim_{\epsilon \to 0} \frac{\max\{0, x+\epsilon\} - \max\{0, x\}}{\epsilon}$$

Further,

$$\lim_{\epsilon \to 0} \frac{\max\{0, x+\epsilon\} - \max\{0, x\}}{\epsilon} = \begin{cases} x > 0, \lim_{\epsilon \to 0} \frac{x+\epsilon-x}{\epsilon} = 1 \\ x = 0, \lim_{\epsilon \to 0} \frac{\max\{0,\epsilon\} - \max\{0,0\}}{\epsilon}, \text{undefined} \\ x < 0, \lim_{\epsilon \to 0} \frac{0-0}{\epsilon} = 0 \end{cases}$$

This is the same as the definition of Heaviside step function

2. From Part 1),

$$g(x) = xH(x), x \neq 0$$

Alternatively,

$$g(x) = max(x-1, H(x-1)), x \neq 1$$

3.
$$\lim_{k \to \infty} \frac{1}{1+e^{-kx}} = \frac{1}{1+0} = 1$$
$$\lim_{k \to -\infty} \frac{1}{1+e^{kx}} = \frac{1}{1+\infty} = 0$$

Asymptoticallty this is good enough for approximation

Question 2 (3-3-3-3)**.** Recall the definition of the softmax function : $S(\boldsymbol{x})_i = e^{\boldsymbol{x}_i} / \sum_j e^{\boldsymbol{x}_j}$.

1. Show that softmax is translation-invariant, that is : $S(\boldsymbol{x}+c) = S(\boldsymbol{x})$, where c is a scalar constant.
2. Show that softmax is not invariant under scalar multiplication. Let $S_c(\boldsymbol{x}) = S(c\boldsymbol{x})$ where $c \geq 0$. What are the effects of taking c to be 0 and arbitrarily large ?
3. Let $\boldsymbol{x}$ be a 2-dimensional vector. One can represent a 2-class categorical probability using softmax $S(\boldsymbol{x})$. Show that $S(\boldsymbol{x})$ can be reparameterized using sigmoid function, i.e. $S(\boldsymbol{x}) = [\sigma(z), 1-\sigma(z)]^\top$ where z is a scalar function of $\boldsymbol{x}$.
4. Let $\boldsymbol{x}$ be a K-dimensional vector ($K \geq 2$). Show that $S(\boldsymbol{x})$ can be represented using $K-1$ parameters, i.e. $S(\boldsymbol{x}) = S([0, y_1, y_2, ..., y_{K-1}]^\top)$ where y_i is a scalar function of $\boldsymbol{x}$ for $i \in \{1, ..., K-1\}$.

Answer 2.

1.

$$S(\boldsymbol{x}+c) = \frac{e^{\boldsymbol{x}_i+c}}{\sum_j e^{\boldsymbol{x}_j+c}} = \frac{e^{\boldsymbol{x}_i}e^c}{\sum_j e^{\boldsymbol{x}_j}e^c} =$$

$$\frac{e^c(e^{\boldsymbol{x}_i})}{e^c(\sum_j e^{\boldsymbol{x}_j})} = \frac{e^{\boldsymbol{x}_i}}{\sum_j e^{\boldsymbol{x}_j}} = S(\boldsymbol{x})$$

2.

$$S_c(\boldsymbol{x}) = S(c\boldsymbol{x}) = \frac{e^{c\boldsymbol{x}_i}}{\sum_j e^{c\boldsymbol{x}_j}} =$$

$$\frac{(e^{\boldsymbol{x}_i})^c}{\sum_j (e^{\boldsymbol{x}_j})^c} = \frac{(e^{\boldsymbol{x}_i})^c}{(e^{\boldsymbol{x}_1})^c + (e^{\boldsymbol{x}_2})^c + ... + (e^{\boldsymbol{x}_j})^c} =$$

$$\frac{e^{\boldsymbol{x}_i}}{((e^{\boldsymbol{x}_1})^c + (e^{\boldsymbol{x}_2})^c + ... + (e^{\boldsymbol{x}_j})^c)^{\frac{1}{c}}} \neq \frac{e^{\boldsymbol{x}_i}}{((e^{\boldsymbol{x}_1} + e^{\boldsymbol{x}_2} + ... + e^{\boldsymbol{x}_j})^c)^{\frac{1}{c}}}$$

$$\frac{e^{\boldsymbol{x}_i}}{((\sum_j e^{\boldsymbol{x}_j})^c)^{\frac{1}{c}}} = \frac{e^{\boldsymbol{x}_i}}{\sum_j e^{\boldsymbol{x}_j}} = S(\boldsymbol{x})$$

Since $((e^{\boldsymbol{x}_1})^c + (e^{\boldsymbol{x}_2})^c + ... + (e^{\boldsymbol{x}_j})^c)^{\frac{1}{c}} \neq ((e^{\boldsymbol{x}_1} + e^{\boldsymbol{x}_2} + ... + e^{\boldsymbol{x}_j})^c)^{\frac{1}{c}}$, from above, $S_c(\boldsymbol{x}) \neq S(\boldsymbol{x})$

$$S_0(\boldsymbol{x}) = \frac{e^0}{\sum_j e^0} = \frac{1}{\sum_j 1} = \frac{1}{j}$$

$$S_\infty(\boldsymbol{x}) = \lim_{c\to\infty} \frac{e^{c\boldsymbol{x}_i}}{\sum_j e^{c\boldsymbol{x}_j}} = \lim_{c\to\infty} \frac{1}{\sum_j e^{c\boldsymbol{x}_j - c\boldsymbol{x}_i}} = \lim_{c\to\infty} \frac{1}{\sum_j e^{c(\boldsymbol{x}_j-\boldsymbol{x}_i)}} = \lim_{c\to\infty} \frac{1}{1+\sum_j e^{\infty}(j \neq i)} = \lim_{c\to\infty} \frac{1}{\infty} = 0$$

$$S_{-\infty}(\boldsymbol{x}) = \lim_{c\to-\infty} \frac{1}{1+\sum_j e^{-\infty}(j \neq i)} = \lim_{c\to-\infty} \frac{1}{1} = 1$$

3. For a class categorical probability whose distribution is softmax

$$P(\boldsymbol{x}_1) = S(\boldsymbol{x}_1) = \frac{e^{\boldsymbol{x}_1}}{e^{\boldsymbol{x}_1} + e^{\boldsymbol{x}_2}} = \frac{1}{1+e^{\boldsymbol{x}_2-\boldsymbol{x}_1}}$$

$$P(\boldsymbol{x}_2) = S(\boldsymbol{x}_2) = \frac{e^{\boldsymbol{x}_2}}{e^{\boldsymbol{x}_1} + e^{\boldsymbol{x}_2}} = \frac{e^{\boldsymbol{x}_2-\boldsymbol{x}_1}}{1+e^{\boldsymbol{x}_2-\boldsymbol{x}_1}}$$

If we denote z as a scalar function of $\boldsymbol{x}$, i.e. $z = \boldsymbol{x}_2 - \boldsymbol{x}_1$, then

$$S(\boldsymbol{x}_1) = \frac{1}{1+e^z} = \sigma(z)$$

$$S(\boldsymbol{x}_2) = \frac{e^z}{1+e^z} = 1 - \sigma(z)$$

Hence $S(\boldsymbol{x}) = [\sigma(z), 1-\sigma(z)]^\top$, where $z = \boldsymbol{x}_2 - \boldsymbol{x}_1$

4.

$$S(\boldsymbol{x}_1) = \frac{e^{\boldsymbol{x}_1}}{e^{\boldsymbol{x}_1} + e^{\boldsymbol{x}_2} + ... + e^{\boldsymbol{x}_j}} = \frac{e^0}{e^0 + e^{\boldsymbol{x}_2 - \boldsymbol{x}_1} ... + e^{\boldsymbol{x}_j - \boldsymbol{x}_1}}$$

$$S(\boldsymbol{x}_2) = \frac{e^{\boldsymbol{x}_2}}{e^{\boldsymbol{x}_1} + e^{\boldsymbol{x}_2} + ... + e^{\boldsymbol{x}_j}} = \frac{e^{\boldsymbol{x}_2 - \boldsymbol{x}_1}}{e^0 + e^{\boldsymbol{x}_2 - \boldsymbol{x}_1} ... + e^{\boldsymbol{x}_j - \boldsymbol{x}_1}}$$

$$...$$

$$S(\boldsymbol{x}_j) = \frac{e^{\boldsymbol{x}_j}}{e^{\boldsymbol{x}_1} + e^{\boldsymbol{x}_2} + ... + e^{\boldsymbol{x}_j}} = \frac{e^{\boldsymbol{x}_j - \boldsymbol{x}_1}}{e^0 + e^{\boldsymbol{x}_2 - \boldsymbol{x}_1} ... + e^{\boldsymbol{x}_j - \boldsymbol{x}_1}}$$

If we define $y_i = \boldsymbol{x}_j - \boldsymbol{x}_1$, since j is in the range of 2 to K, we have $K-2+1 = K-1$ parameters, therefore $i \in \{1, ..., K-1\}$. Similar to last question, we have $S(\boldsymbol{x}) = S([0, y_1, y_2, ..., y_{K-1}]^\top)$

Question 3 (16)**.** Consider a 2-layer neural network $y : \mathbb{R}^D \to \mathbb{R}^K$ of the form :

$$y(x, \Theta, \sigma)_k = \sum_{j=1}^{M} \omega_{kj}^{(2)} \sigma \left(\sum_{i=1}^{D} \omega_{ji}^{(1)} x_i + \omega_{j0}^{(1)} \right) + \omega_{k0}^{(2)}$$

for $1 \leq k \leq K$, with parameters $\Theta = (\omega^{(1)}, \omega^{(2)})$ and logistic sigmoid activation function σ. Show that there exists an equivalent network of the same form, with parameters $\Theta' = (\tilde{\omega}^{(1)}, \tilde{\omega}^{(2)})$ and tanh activation function, such that $y(x, \Theta', \tanh) = y(x, \Theta, \sigma)$ for all $x \in \mathbb{R}^D$, and express Θ' as a function of Θ.

Answer 3. We can show that,

$$tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} = \frac{e^{2x} + 1 - 2}{e^{2x} + 1} = 1 - 2 \times \frac{1}{1 + e^{2x}} = 1 - 2\sigma(-2x)$$

We also have,

$$\sigma(x) = \frac{1}{1+e^{-x}} = 1 - \frac{e^{-x}}{1+e^{-x}} = 1 - \frac{1}{e^x + 1} = 1 - \sigma(-x)$$

Therefore,

$$tanh(x) = 1 - 2\sigma(-2x) = 1 - 2(1 - \sigma(2x)) = 2\sigma(2x) - 1$$

$$\sigma(x) = \frac{1}{2} \times (tanh(\frac{x}{2}) + 1)$$

Using this, we can re-write y as :

$$y(x, \Theta, \tanh) = \sum_{j=1}^{M} \frac{1}{2} \omega_{kj}^{(2)} tanh \left(\frac{1}{2} \sum_{i=1}^{D} \omega_{ji}^{(1)} x_i + \frac{1}{2} \omega_{j0}^{(1)} + 1 \right) + \omega_{k0}^{(2)}$$

$$y(x, \Theta', \tanh) = \sum_{j=1}^{M} \tilde{\omega}_{kj}^{(2)} tanh \left(\sum_{i=1}^{D} \tilde{\omega}_{ji}^{(1)} x_i + \tilde{\omega}_{j0}^{(1)} \right) + \tilde{\omega}_{k0}^{(2)}$$

Whereby,

$$\tilde{\omega}_{kj}^{(2)} = \frac{1}{2} \omega_{kj}^{(2)}$$

TABLE 1 – Forward AD example, with $y = f(x_1, x_2) = \ln(x_1) + x_1x_2 - \sin(x_2)$ at $(x_1, x_2) = (2, 5)$ and setting $\dot{x}_1 = 1$ to compute $\partial y/\partial x_1$.

	Forward evaluation trace		
	v_{-1}	$= x_1$	$= 2$
	v_0	$= x_2$	$= 5$
	v_1	$= \ln(v_1)$	$= \ln(2)$
	v_2	$= v_{-1} \times v_0$	$= 2 \times 5$
$\Downarrow$	v_3	$= \sin(v_0)$	$\sin(5)$
	v_4	$= v_1 + v_2$	$= 0.6931 + 10$
	v_5	$= v_4 - v_3$	$= 10.6931 + 0.9589$
	y	$= v_5$	$= 11.6521$

	Forward derivative trace		
	$= \dot{v}_{-1}$	$\dot{x}_1$	$= 1$
	$= \dot{v}_0$	$\dot{x}_2$	$= 0$
	$\dot{v}_1$	$= \dot{v}_{-1}/v_{-1}$	$= 1/2$
	$\dot{v}_2$	$= \dot{v}_{-1} \times v_0 + v_{-1} \times \dot{v}_0$	$= 1 \times 5 + 2 \times 0$
$\Downarrow$	$\dot{v}_3$	$= \cos v_0 \times \dot{v}_0$	$= \cos(5) \times 0$
	$\dot{v}_4$	$= \dot{v}_1 + \dot{v}_2$	$= 0.5 + 5$
	$\dot{v}_5$	$= \dot{v}_4 - \dot{v}_3$	$= 5.5 - 0$
	$= \dot{y}$	$\dot{v}_5$	$= 5.5$

TABLE 2 – Reverse AD example, with $y = f(x_1, x_2) = \ln(x_1) + x_1x_2 - \sin(x_2)$ at $(x_1, x_2) = (2, 5)$. Setting $\bar{y} = 1$, $\partial y/\partial x_1$ and $\partial y/\partial x_2$ are computed in one reverse sweep.

	Forward evaluation trace		
	v_{-1}	$= x_1$	$= 2$
	v_0	$= x_2$	$= 5$
	v_1	$= \ln(v_1)$	$= \ln(2)$
	v_2	$= v_{-1} \times v_0$	$= 2 \times 5$
$\Downarrow$	v_3	$= \sin(v_0)$	$= \sin(5)$
	v_4	$= v_1 + v_2$	$= 0.6931 + 10$
	v_5	$= v_4 - v_3$	$= 10.6931 + 0.9589$
	y	$= v_5$	$= 11.6521$

	Reverse adjoint trace		
	$\bar{x}_1$	$= \bar{v}_{-1}$	$= 5.5$
	$\bar{x}_2$	$= \bar{v}_0$	$= 1.7163$
	$\bar{v}_{-1}$	$= \bar{v}_{-1} + \bar{v}_1 \frac{\partial v_1}{\partial v_{-1}}$	$= 5.5$
	$\bar{v}_0$	$= \bar{v}_0 + \bar{v}_2 \frac{\partial v_2}{\partial v_0}$	$= 1.7163$
$\Uparrow$	$\bar{v}_{-1}$	$= \bar{v}_2 \frac{\partial v_2}{\partial v_{-1}}$	$= 5$
	$\bar{v}_0$	$= \bar{v}_3 \frac{\partial v_3}{\partial v_0}$	$= -0.2837$
	$\bar{v}_2$	$= \bar{v}_4 \frac{\partial v_4}{\partial v_2}$	$= 1$
	$\bar{v}_1$	$= \bar{v}_4 \frac{\partial v_4}{\partial v_1}$	$= 1$
	$\bar{v}_3$	$= \bar{v}_5 \frac{\partial v_5}{\partial v_3}$	$= -1$
	$\bar{v}_4$	$= \bar{v}_5 \frac{\partial v_5}{\partial v_4}$	$= 1$
	$\bar{v}_5$	$= \bar{y}$	$= 1$

$$\tilde{\omega}_{ji}^{(1)} = \frac{1}{2}\omega_{ji}^{(1)}$$
$$\tilde{\omega}_{j0}^{(1)} = \frac{1}{2}\omega_{j0}^{(1)} + 1$$
$$\tilde{\omega}_{k0}^{(2)} = \omega_{k0}^{(2)}$$

Subsequently, we can express $\theta'(\tilde{\omega}^{(1)}, \tilde{\omega}^{(2)})$ in terms of $\theta(\omega^{(1)}, \omega^{(2)})$,

$$\theta' = (\tilde{\omega}^{(1)}, \tilde{\omega}^{(2)}) = (\left[\tilde{\omega}_{j0}^{(1)}, \tilde{\omega}_{ji}^{(1)}\right], \left[\tilde{\omega}_{k0}^{(2)}, \tilde{\omega}_{kj}^{(2)}\right]) = (\left[\tfrac{1}{2}\omega_{j0}^{(1)} + 1, \tfrac{1}{2}\omega_{ji}^{(1)}\right], \left[\tfrac{1}{2}\omega_{k0}^{(2)}, \omega_{kj}^{(2)}\right])$$

Question 4 (5-5)**.** Fundamentally, back-propagation is just a special case of reverse-mode Automatic Differentiation (AD), applied to a neural network. Based on the "three-part" notation shown in Table 1 and 2, represent the evaluation trace and derivative (adjoint) trace of the following examples. In the last columns of your solution, numerically evaluate the value up to 4 decimal places.

1. Forward AD, with $y = f(x_1, x_2) = 1/(x_1 + x_2) + {x_2}^2 + \cos(x_1)$ at $(x_1, x_2) = (3, 6)$ and setting $\dot{x}_1 = 1$ to compute $\partial y/\partial x_1$.

2. Reverse AD, with $y = f(x_1, x_2) = 1/(x_1 + x_2) + {x_2}^2 + \cos(x_1)$ at $(x_1, x_2) = (3, 6)$. Setting $\bar{y} = 1$, $\partial y/\partial x_1$ and $\partial y/\partial x_2$ can be computed together.

Answer 4. Reuse the tables to prepare your answer.

	Forward evaluation trace		
	v_{-1}	$= x_1$	$= 3$
	v_0	$= x_2$	$= 6$
	v_1	$= \cos(v_1)$	$= \cos(3) = 0.9900$
	v_2	$= \frac{1}{v_{-1}+v_0}$	$= \frac{1}{9} = 0.1111$
$\Downarrow$	v_3	$= (v_0)^2$	$6^2 = 36$
	v_4	$= v_1 + v_2$	$= -0.8789$
	v_5	$= v_4 + v_3$	$= 35.1211$
	y	$= v_5$	$= 35.1211$

	Forward derivative trace		
	$= \dot{v}_{-1}$	$\dot{x}_1$	$= 1$
	$= \dot{v}_0$	$\dot{x}_2$	$= 0$
	$\dot{v}_1$	$= \dot{v}_{-1} \times -\sin(v_{-1})$	$= -0.1411$
	$\dot{v}_2$	$= \frac{-1}{(v_{-1}+v_0)^2} \times (v_0 \times \dot{v_{-1}} + \dot{v}_0 \times v_{-1})$	$= -0.6667$
$\Downarrow$	$\dot{v}_3$	$= 2\dot{v}_2 v_2$	$= 0$
	$\dot{v}_4$	$= \dot{v}_1 + \dot{v}_2$	$= -0.8078$
	$\dot{v}_5$	$= \dot{v}_4 - \dot{v}_3$	$= -0.8078$
	$= \dot{y}$	$\dot{v}_5$	$= -0.8078$

	Forward evaluation trace		
	v_{-1}	$= x_1$	$= 3$
	v_0	$= x_2$	$= 6$
	v_1	$= \cos(v_1)$	$= \cos(3) = 0.9900$
	v_2	$= \frac{1}{v_{-1}+v_0}$	$= \frac{1}{9} = 0.1111$
$\Downarrow$	v_3	$= (v_0)^2$	$6^2 = 36$
	v_4	$= v_1 + v_2$	$= -0.8789$
	v_5	$= v_4 + v_3$	$= 35.1211$
	y	$= v_5$	$= 35.1211$

	Reverse adjoint trace		
	$\bar{x}_1$	$= \bar{v}_{-1}$	$= -0.1534$
	$\bar{x}_2$	$= \bar{v}_0$	$= 11.9877$
	$\bar{v}_{-1}$	$= \bar{v}_{-1} + \bar{v}_1 \frac{\partial v_1}{\partial v_{-1}}$	$= \bar{v}_{-1} - \sin(v_{-1}) = -0.0123 - \sin(3) = -0.1534$
	$\bar{v}_0$	$= \bar{v}_0 + \bar{v}_2 \frac{\partial v_2}{\partial v_0}$	$= \bar{v}_0 + \frac{1}{(v_{-1}+v_0)^2} = 11.9877$
$\Uparrow$	$\bar{v}_{-1}$	$= \bar{v}_2 \frac{\partial v_2}{\partial v_{-1}}$	$= \frac{-\bar{v}_2}{(v_{-1}+v_0)^2} = -0.0123$
	$\bar{v}_0$	$= \bar{v}_3 \frac{\partial v_3}{\partial v_0}$	$= 2\bar{v}_3 v_0 = 12$
	$\bar{v}_2$	$= \bar{v}_4 \frac{\partial v_4}{\partial v_2}$	$= 1$
	$\bar{v}_1$	$= \bar{v}_4 \frac{\partial v_4}{\partial v_1}$	$= 1$
	$\bar{v}_3$	$= \bar{v}_5 \frac{\partial v_5}{\partial v_3}$	$= 1$
	$\bar{v}_4$	$= \bar{v}_5 \frac{\partial v_5}{\partial v_4}$	$= 1$
	$\bar{v}_5$	$= \bar{y}$	$= 1$

Question 5 (6)**.** Compute the *full*, *valid*, and *same* convolution (with kernel flipping) for the following 1D matrices : $[1, 2, 3, 4] * [1, 0, 2]$

Answer 5. Full : $[1, 2, 5, 8, 6, 8]$; Valid : $[5, 8]$; Same : $[2, 5, 8, 6]$.

Question 6 (5-5)**.** Consider a convolutional neural network. Assume the input is a colorful image of size 256×256 in the RGB representation. The first layer convolves 64 8×8 kernels with the input, using a stride of 2 and no padding. The second layer downsamples the output of the first layer with a 5×5 non-overlapping max pooling. The third layer convolves 128 4×4 kernels with a stride of 1 and a zero-padding of size 1 on each border.

1. What is the dimensionality (scalar) of the output of the last layer ?
2. Not including the biases, how many parameters are needed for the last layer ?

Answer 6.

1. Input : (256, 256, 3)
   First layer : $\frac{256-8+2\times 0}{2} + 1 = 125$, (125, 125, 64)
   Second layer : $\frac{125-5}{5} + 1 = 25$, (25, 25, 64)
   Third layer : $\frac{25-4+2\times 1}{1} + 1 = 24$, (24, 24, 128)
   Output shape : (128, 24, 24)
2. $4 \times 4 \times 128 \times 64 = 131072$

Question 7 (4-4-6)**.** Assume we are given data of size $3 \times 64 \times 64$. In what follows, provide a correct configuration of a convolutional neural network layer that satisfies the specified assumption. Answer with the window size of kernel (k), stride (s), padding (p), and dilation (d, with convention $d = 1$ for no dilation). Use square windows only (e.g. same k for both width and height).

1. The output shape (o) of the first layer is $(64, 32, 32)$.
   (a) Assume $k = 8$ without dilation.
   (b) Assume $d = 7$, and $s = 2$.
2. The output shape of the second layer is $(64, 8, 8)$. Assume $p = 0$ and $d = 1$.
   (a) Specify k and s for pooling with non-overlapping window.
   (b) What is output shape if $k = 8$ and $s = 4$ instead ?
3. The output shape of the last layer is $(128, 4, 4)$.

(a) Assume we are not using padding or dilation.

(b) Assume $d = 2$, $p = 2$.

(c) Assume $p = 1$, $d = 1$.

Answer 7. Fill up the following table,

		i	p	d	k	s	o
1.	(a)	(3,64,64)	3	1	8	2	(64,32,32)
	(b)	(3,64,64)	7	7	3	2	(64,32,32)
2.	(a)	(64,32,32)	0	1	4	4	(64,8,8)
	(b)	(64,32,32)	0	1	8	4	(64,7,7)
3.	(a)	(64,8,8)	0	1	5	1	(128,4,4)
	(b)	(64,8,8)	2	2	3	2	(128,4,4)
	(c)	(64,8,8)	1	1	3	2	(128,4,4)