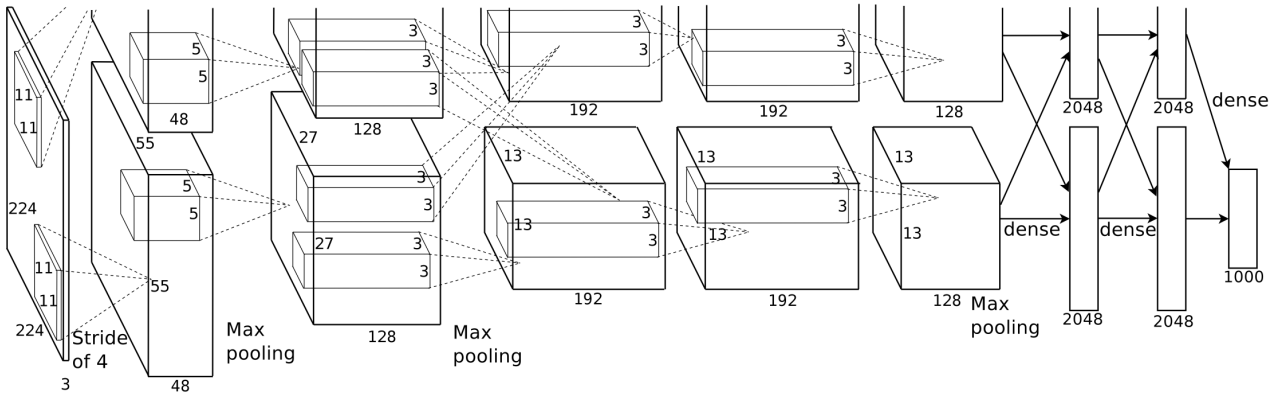
CS461 Quiz Three

CS461 Section #:	
Name:	
NetID:	

0. True / False Questions.

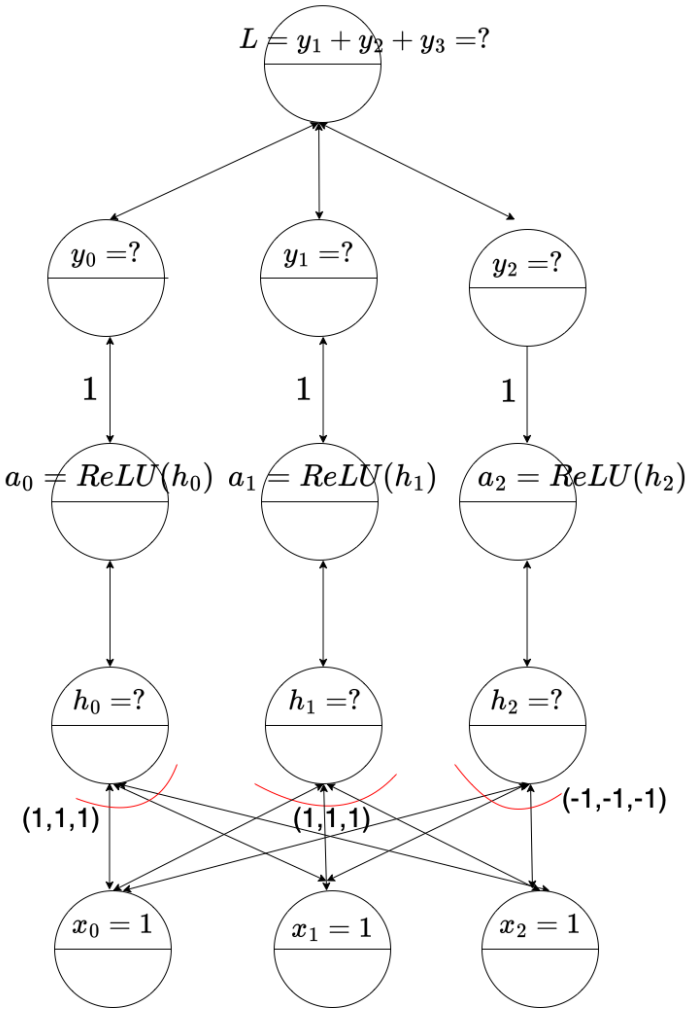
- By computing the gradient of $J(w)$ of a deep-CNN, we can find its global minimum. (True / **False**)
- In momentum optimization, a larger momentum will reduce oscillations in the iterative steps. (**True** / False)
- Both of Ridge and Lasso regularization are applicable to deep CNN training. (**True** / False)
- Max pooling helps to achieve equivariance. (True / **False**)
- Dropout layer (0.5) discards 50% of the hidden units in training stage, but the layer is disregarded during inference. (**True** / False)
- A max pooling layer does not reduce the size of input feature map. (True / **False**)
- In deep-CNNs, we can input various sized images for inference. (**True** / False)
- The polynomial $y = x^3$ is a desirable activation function for its non-linearity. (True / **False**)
- Once a unit value becomes zero in a deep-CNN with ReLU, the unit never revive again. (True / **False**)

1. [The architecture of Alexnet and the input and output]



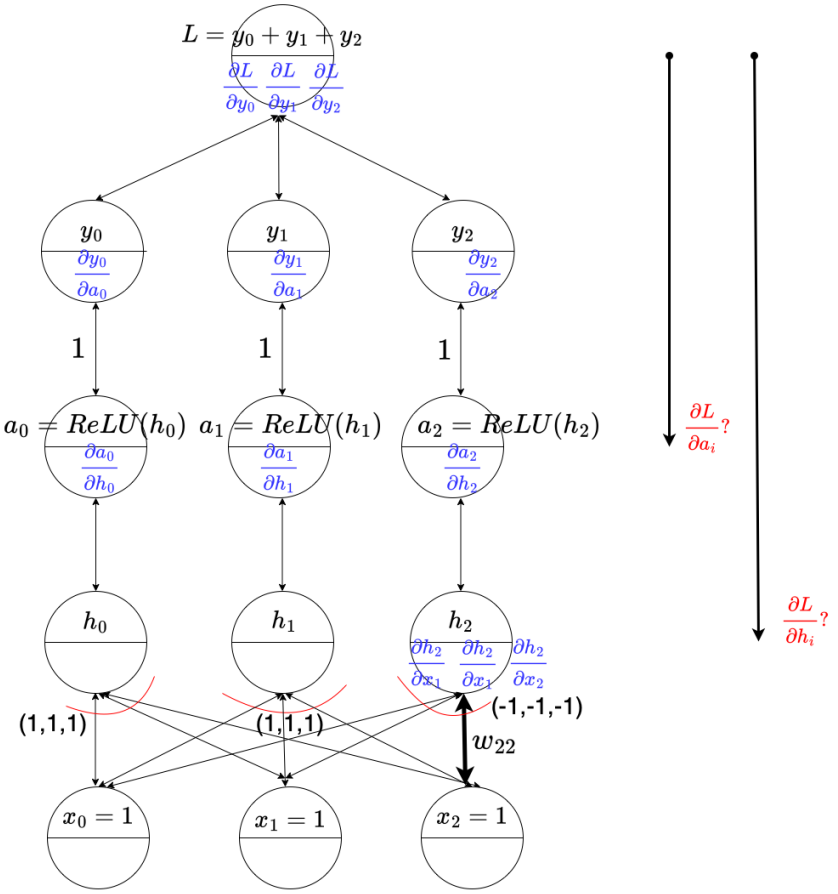
1. In the figure, the size of input image is $3 \times$ **224** $\times$ 224.
2. Suppose the size of the first convolution filter is $3 \times 11 \times 11$. And, the size of the second convolution filter is $96 \times 5 \times 5$. Based on the information, we can make an inference: (1) the number of the first convolution filters will be **96** and (2) the number of bias of the first convolution layer will be **96**.
3. The size of input and output in the last fully connected (FC) layer is 4096 and 1000. The number of parameters of the last FC layer is **4,096,000, or 4,097,000 (bias)**.
4. AlexNet learns the discriminant functions for 1,000 objects classification. How many discriminant functions does Alexnet learn **1,000**?

2. [Feed-forward] evaluate unit values in feed-forward step in the network below. The current parameter values are $w_{i,0} = (1, 1, 1), w_{i,1} = (1, 1, 1), w_{i,2} = (-1, -1, -1)$, where $w_{i,j}$ is the edge connecting the units: x_i and h_j. Fill out the table below and the table will be graded.



units	unit value in feed-forward step
h_0	3
h_1	3
h_2	-3
y_0	3
y_1	3
y_2	0
$Loss = y_0 + y_1 + y_2$	6

3 [Backpropagation] evaluate the derivatives and fill out the table below; the table will be graded.



3.1 Fill out the table.

Units	Unit Derivative
$\partial L/\partial y_2$	1
$\partial y_2/\partial a_2$	1
$\partial a_2/\partial h_2$	0 for $h_2 < 0$
$\partial h_2/\partial x_2$	-1

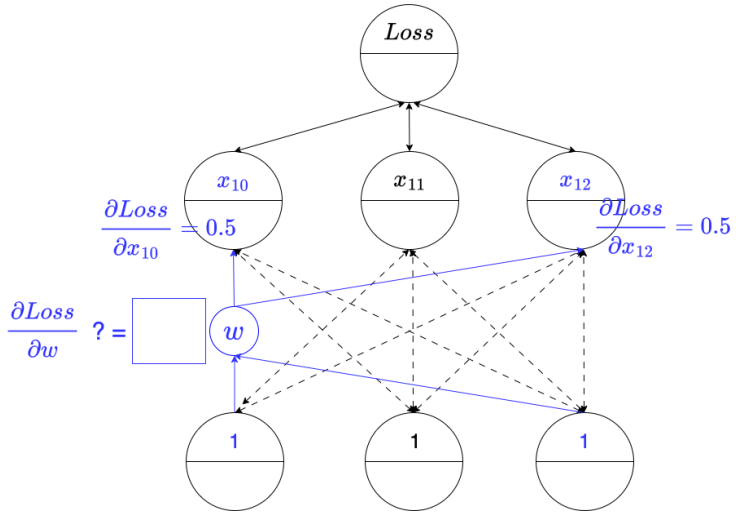
Units	Unit Derivative
$\partial L/\partial a_2$	$\partial L/\partial y_2 \cdot \partial y_2/\partial a_2 = 1$
$\partial L/\partial h_2$	$\partial L/\partial a_2 \cdot \partial a_2/\partial h_2 = 0$
$\partial L/\partial w_{22}$	$\partial L/\partial h_2 \cdot x_2 = 0$

3.2 Compute a new value of w_{22} by using $w_{22}^{'} = w_{22} - \eta \frac{\partial L}{\partial w_{22}}$ where $\eta = 1$ and w_{22} denotes the parameter between h_2 and x_2.

- $w_{22}^{'} = -1 + 0 = -1$

[Extra Credits]

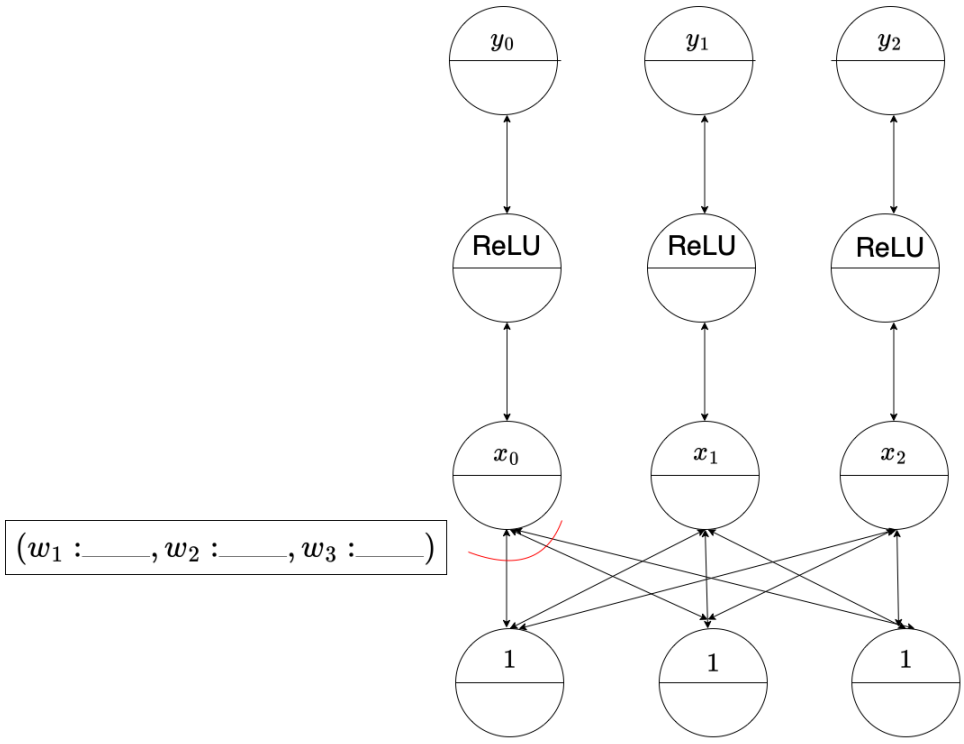
E.1 (10 points) Compute the derivative of loss respect to w ($\frac{\partial L}{\partial w}$) when the parameter w is shared to compute output x_{10} and x_{12}.



sol)

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial x_{10}}\frac{\partial x_{10}}{\partial w} + \frac{\partial L}{\partial x_{12}}\frac{\partial x_{12}}{\partial w} = 0.5 + 0.5 = 1$$

E.2 (10 points) Write the possible weights (w_1, w_2, w_3) that would result in permanent inactivation for the unit x_0.



sol) all negative or zero parameters will make unit x_0 inactive permanently.