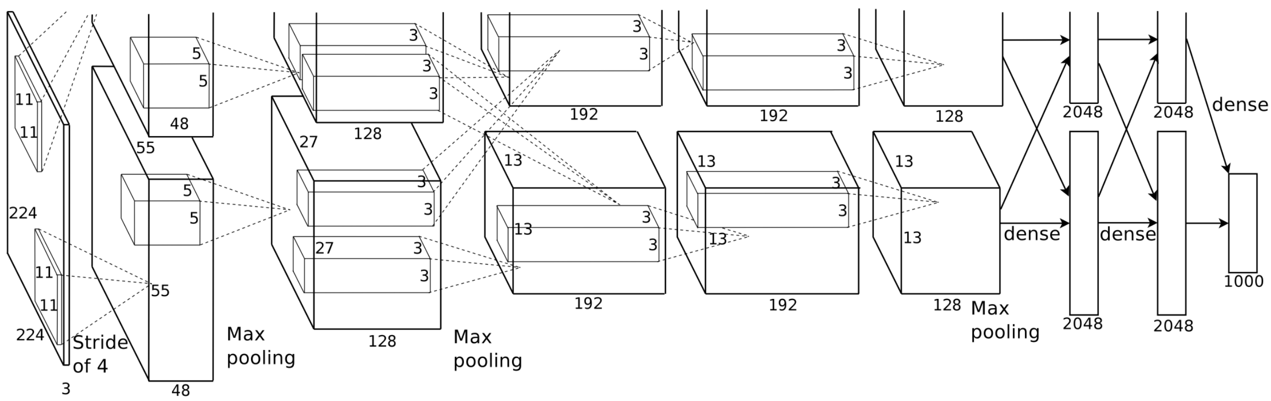
CS461 Quiz Three

CS461 Section #:	
Name:	
NetID:	

0. True / False Questions.

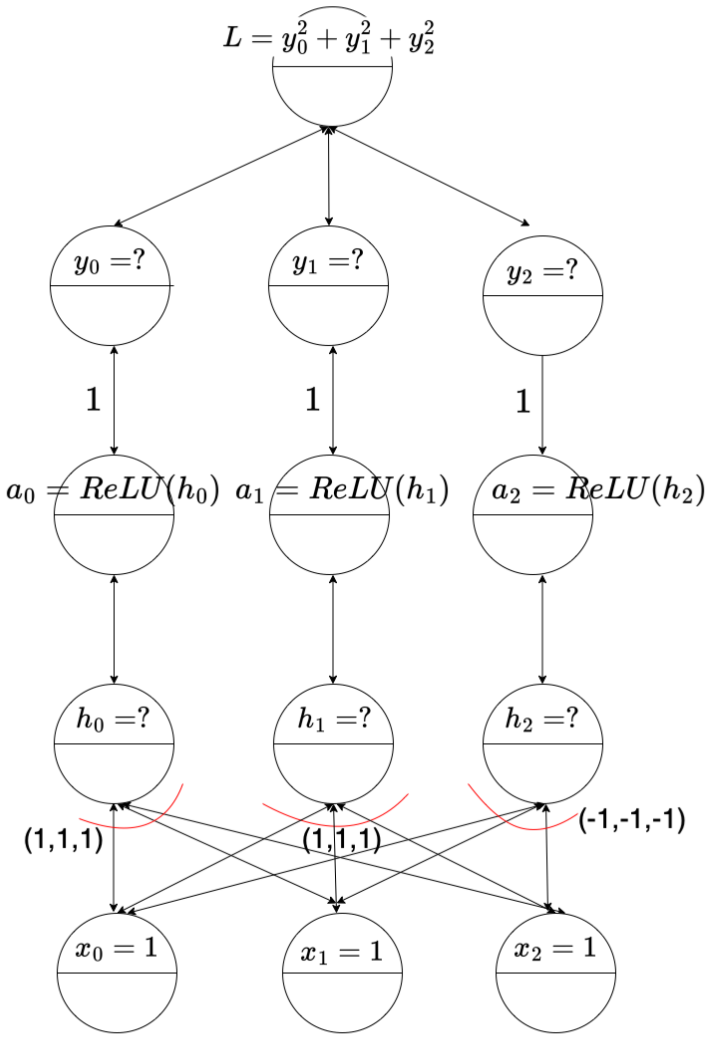
- The stochastic descent algorithm aims to find the global minimum of $J(w)$ of a deep-CNN. (True / False)
- In momentum optimization, a larger momentum will reduce the oscillations in the iterative steps. (True / False)
- Both of Ridge and Lasso regularization are applicable to deep-CNN training. (True / False)
- Max pooling layer promotes invariance. (True / False)
- Dropout layer (0.5) discards 50% of the hidden unit values in training stage. The dropout operation is still effective during inference for generalization. (True / False)
- Tanh or Sigmoid function are prone to the gradient vanishing problem. (True / False)
- In deep-CNNs, we can input various sized images for inference. (True / False)
- The polynomial $y = x^5$ 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)
- Newton's gradient method is often used in training a deep neural net because it promotes fast convergence. (True / False)
- When an objective function is quadratic, Newton's gradient method ($w_{k+1} = w_k + \alpha \nabla^2 J(w_k)^{-1} \nabla J(w_k)$) can find an optimal solution in a single iteration. (True / False)

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



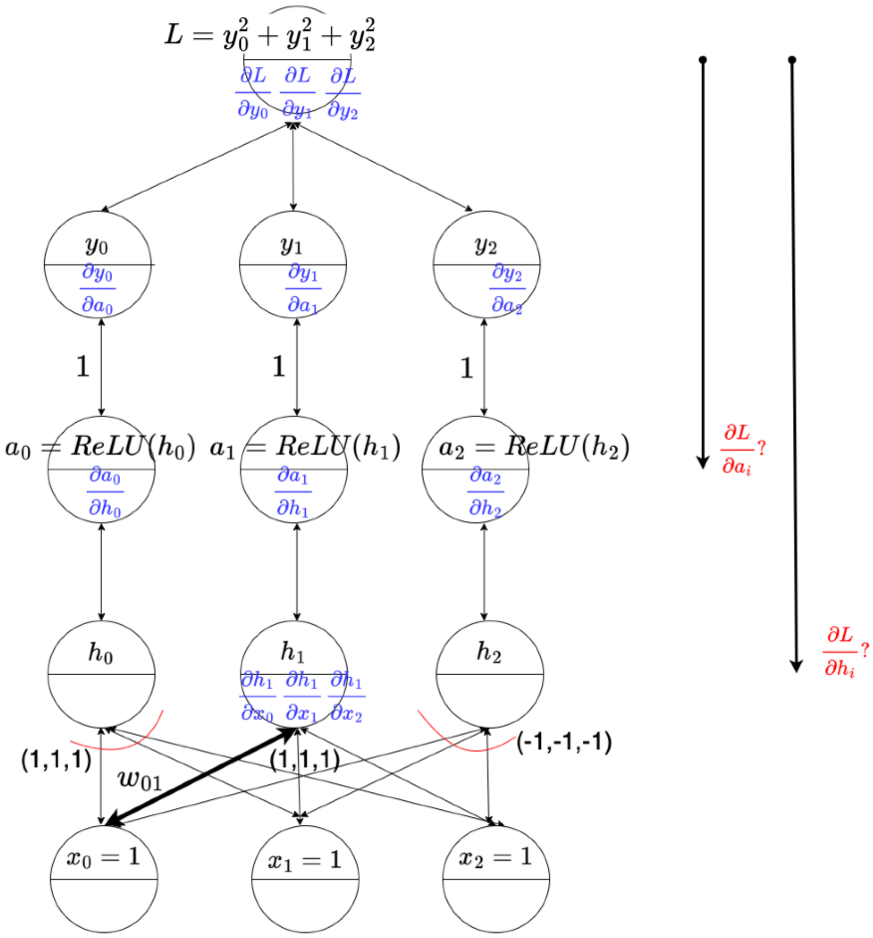
1. In the figure, the size of input image is 3 × __________ × 224.
2. Suppose the size of the second convolution filter is 64 × 5 × 5. And, the size of the third convolution filter is 192 × 3 × 3. Based on the information, we can make an inference: (1) the number of the second convolution filters will be __________ and (2) the number of bias of the second convolution layer will be __________.
3. The size of input and output in the last fully connected (FC) layer is 4096 and 1000. When there is a bias term, the number of parameters of the last FC layer is __________.
4. AlexNet learns the discriminant functions for 1,000 objects classification. How many discriminant functions does Alexnet learn? __________

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	
h_1	
h_2	
y_0	
y_1	
y_2	
$Loss = y_0^2 + y_1^2 + y_2^2$	

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
$\partial L/\partial y_1$	
$\partial y_1/\partial a_1$	
$\partial a_1/\partial h_1$	
$\partial h_1/\partial x_0$	

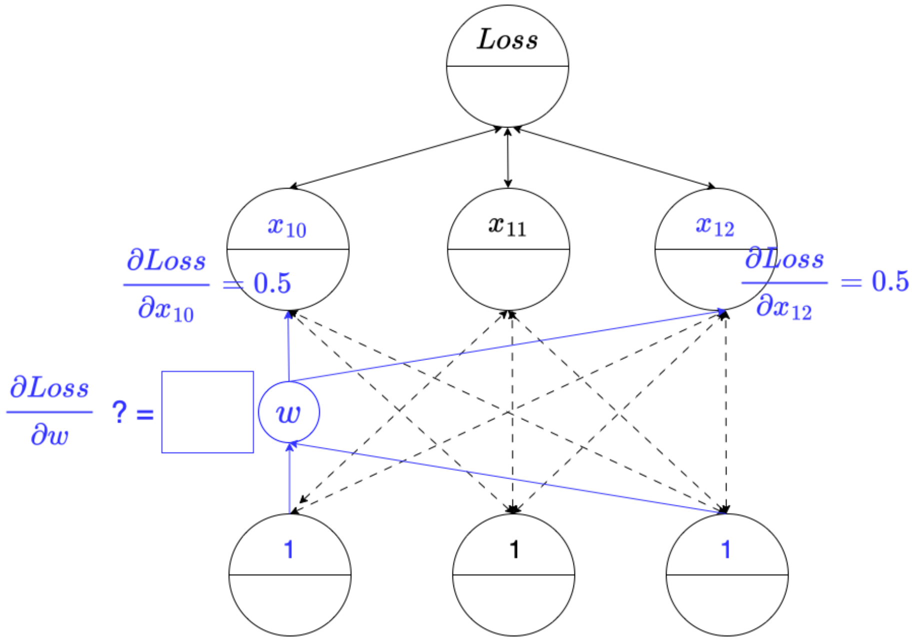
Units	Unit Derivative
$\partial L/\partial a_1$	
$\partial L/\partial h_1$	
$\partial L/\partial w_{01}$	

3.2 Compute a new value of w_{01} by using $w_{01}^{'} = w_{01} - \eta \frac{\partial L}{\partial w_{01}}$ where $\eta = 1$ and w_{01} denotes the parameter between h_1 and x_0.

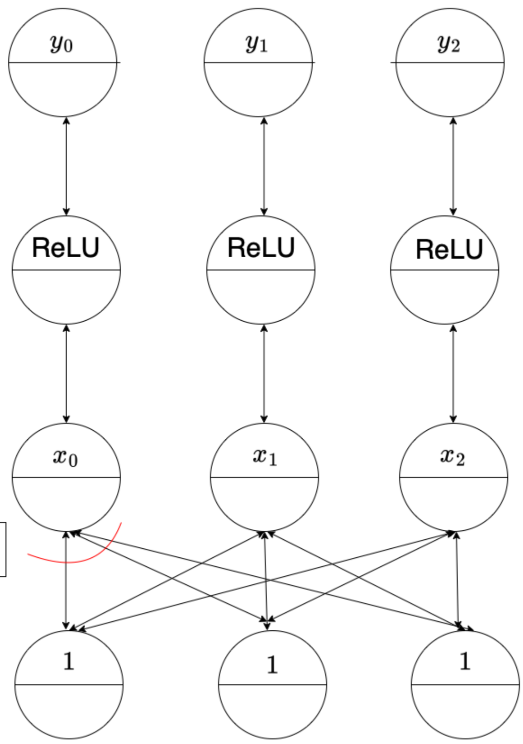
- $w_{01}^{'} = ?$

4. [ReLU]

4.1 (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}.



4.2 (10 points) Write the possible weights (w_1, w_2, w_3) that would result in inactivation for the current step but possibly revive later steps.



$(w_1 :\_\_\_\_, w_2 :\_\_\_\_, w_3 :\_\_\_\_)$

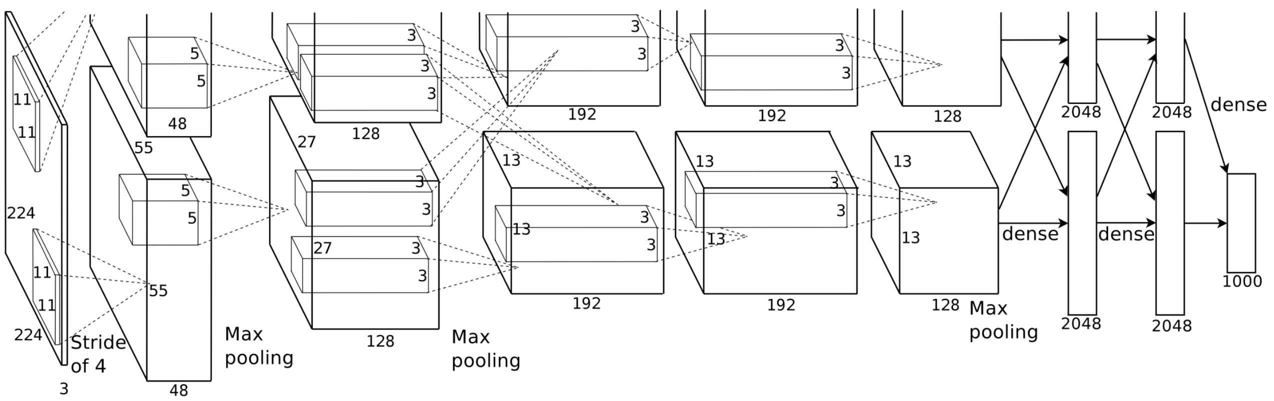
CS461 Section #:	
Name:	
NetID:	

CS461 Quiz April 10

0. True / False Questions.

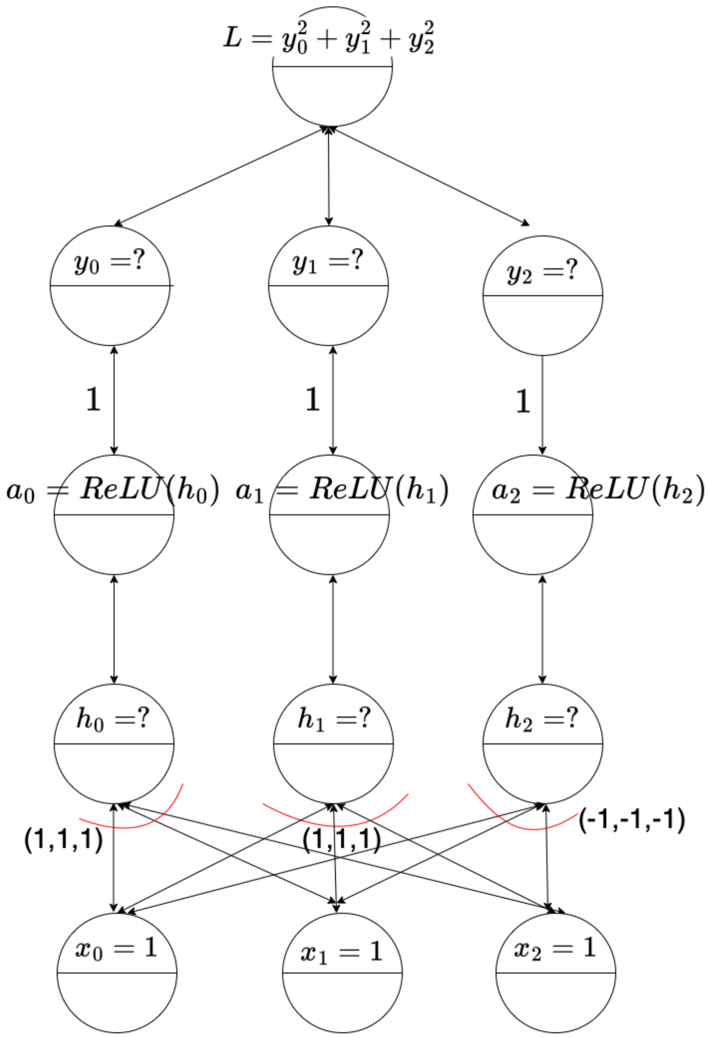
- The stochastic descent algorithm aims to find the global minimum of $J(w)$ of a deep-CNN. (True / **False**)
- In momentum optimization, a larger momentum will reduce the oscillations in the iterative steps. (**True** / False)
- Both of Ridge and Lasso regularization are applicable to deep-CNN training. (**True** / False)
- Max pooling layer promotes invariance. (**True** / False)
- Dropout layer (0.5) discards 50% of the hidden unit values in training stage. The dropout operation is still effective during inference for generalization. (True / **False**)
- Tanh or Sigmoid function are prone to the gradient vanishing problem. (**True** / False)
- In deep-CNNs, we can input various sized images for inference. (**True** / False)
- The polynomial $y = x^5$ 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**)
- Newton's gradient method is often used in training a deep neural net because it promotes fast convergence. (**True** / False)
- When an objective function is quadratic, Newton's gradient method ($w_{k+1} = w_k + \alpha \nabla^2 J(w_k)^{-1} \nabla J(w_k)$) can find an optimal solution in a single iteration. (**True** / False)

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



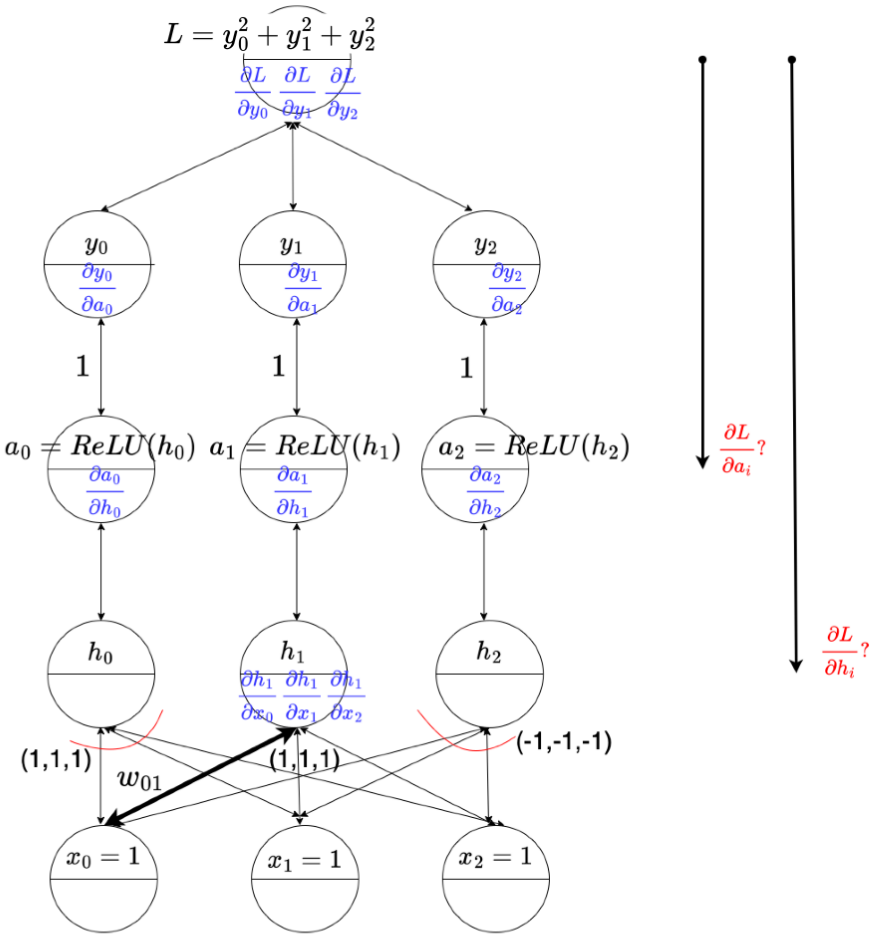
1. In the figure, the size of input image is 3 × (224) × 224.
2. Suppose the size of the second convolution filter is 64 × 5 × 5. And, the size of the third convolution filter is 192 × 3 × 3. Based on the information, we can make an inference: (1) the number of the second convolution filters will be (192) and (2) the number of bias of the second convolution layer will be (192).
3. The size of input and output in the last fully connected (FC) layer is 4096 and 1000. When there is a bias term, the number of parameters of the last FC layer is 4,097,000.
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
$Loss = y_0^2 + y_1^2 + y_2^2$	18

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
$\partial L/\partial y_1$	$2 \cdot y_1 = 6$
$\partial y_1/\partial a_1$	1
$\partial a_1/\partial h_1$	1
$\partial h_1/\partial x_0$	1

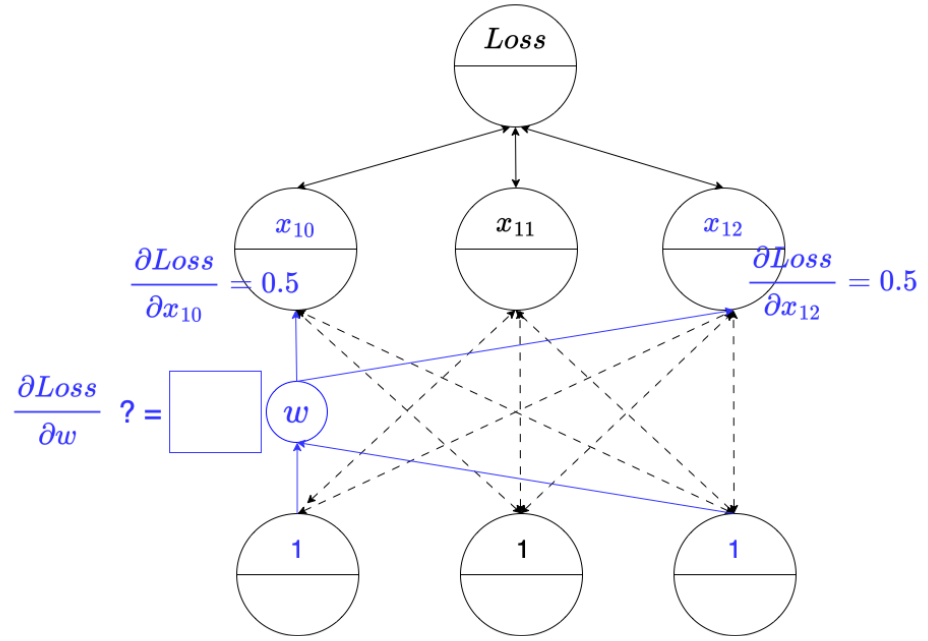
Units	Unit Derivative
$\partial L/\partial a_1$	6
$\partial L/\partial h_1$	6
$\partial L/\partial w_{01}$	$6 \cdot 1 = 6$

3.2 Compute a new value of w_{01} by using $w_{01}^{'} = w_{01} - \eta\frac{\partial L}{\partial w_{01}}$ where $\eta = 1$ and w_{01} denotes the parameter between h_1 and x_0.

- $w_{01}^{'} = 1 - 1 \cdot 6 = 1 - 6 = -5$

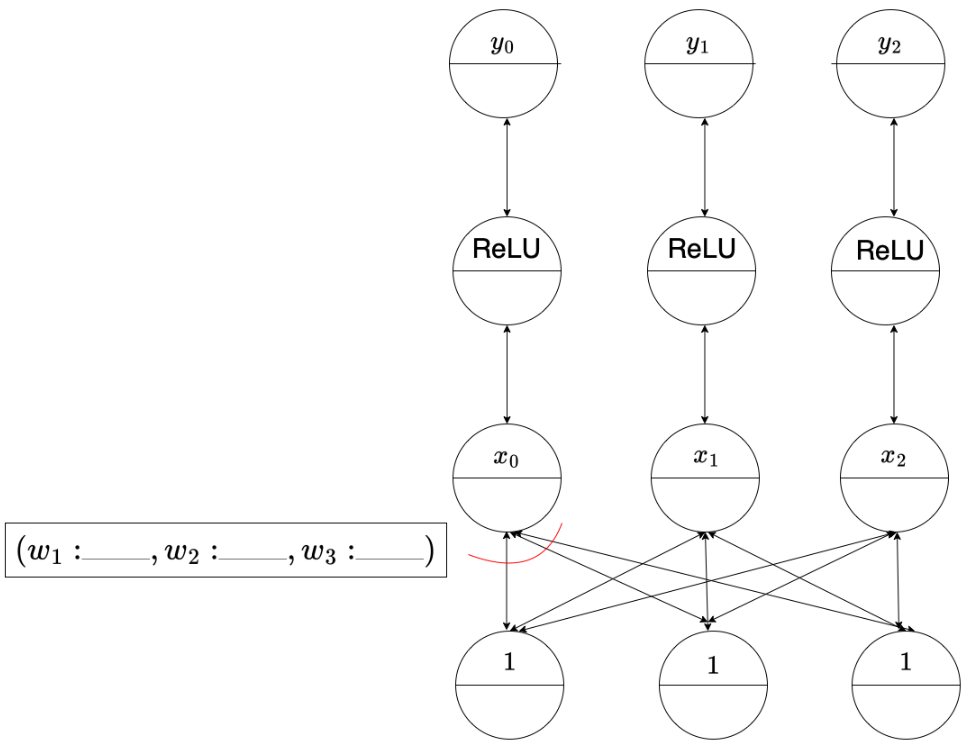
4. [ReLU]

4.1 (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) $0.5 + 0.5 = 1$

4.2 (10 points) Write the possible weights (w_1, w_2, w_3) that would result in inactivation for the current step but possibly revive later steps.



sol) any combination $w_1 + w_2 + w_3 < 0$ but at least one of the weights must be positive. for example, $(-5, -5, 1)$.