HW # 6 & 7 Due: 6/18/2021

1. If you use a 3-layer (one hidden layer) neural network as a classifier for a problem. How do you determine whether your classifier has high bias? If so, will you increase or decrease the number of hidden units? How do you determine whether your classifier has high variance? If so, how do you cope with it (increase or decrease the number of hidden units)?
2. If you want to use a multilayer neural network for regression, can you use cross entropy as the loss function? If no, why not?
3. Assuming that the following is a part of convolution neural networks. Compute the resultant values if it has two input channels, stride of **two**, no zero-padding, and using the ReLU activation function. Remember that the output from this part of network has only ONE value.

Input channel 1

6	0	-4	0	1
4	4	0	2	1
3	-7	1	4	2
-2	2	1	-4	2
5	1	2	4	-1

Kernel

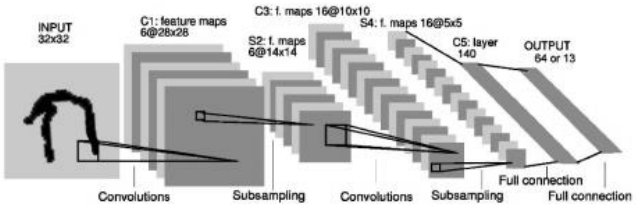
0	1	0
-2	0	-2
0	1	0

Input channel 2

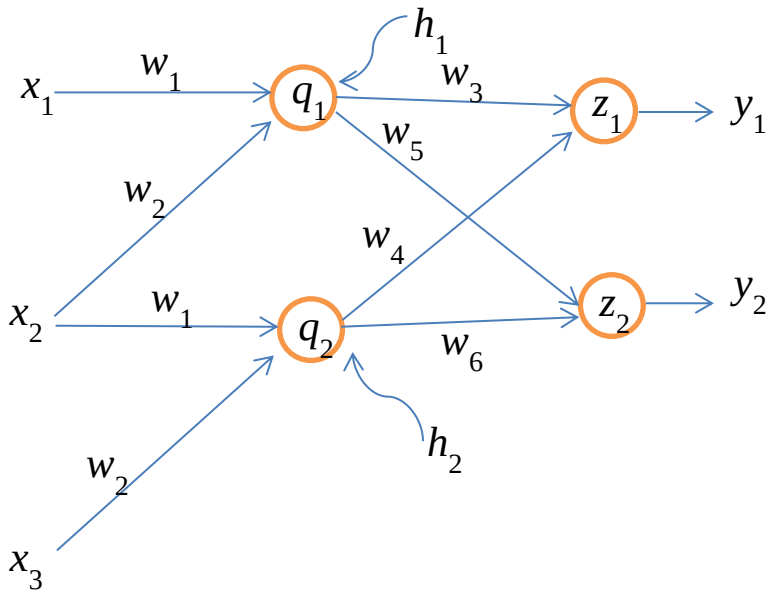
3	1	-4	0	-1
3	0	3	4	-1
3	7	-2	-2	2
-5	-2	2	0	-2
2	1	-1	1	2

0	1	0
1	-4	1
0	1	0

4. The following shows the LeNet-5 architecture. If we use the weight-sharing approach for the convolutional layers, compute the number of connections and trainable weights of the network. To compute the results, you need the following parameters for the convolutional layers: kernel size = 5×5, stride = 1, and no zero-padding. To simplify the computations, ignore the bias weights and let the output units be 10.



5. Find $\Delta w_1 = \eta \frac{\partial J}{\partial w_1}$ of the following CNN using backprop. The activation function from q_1 to h_1 and q_2 to h_2 is ReLU, the outputs y_1 and y_2 are softmax ouput, and the cost function is $J = -\log y_1$. Let w_1 to w_6 be 1.0, $x_1 = 0.5$, $x_2 = 1.0$, $x_3 = -0.5$, and $\eta = 0.1$.



6. Show the following is true. If $y_1 = \frac{1}{1+exp(-z_1)}$, then $\frac{\partial y_1}{\partial z_1} = y_1(1 - y_1)$.
7. The back propagation algorithm does NOT work if all weights are initially set to zero. Why? To make life easier, use the Relu activation function in the discussion.
8. Use the logistic regression to classify the wind data set. As usual, use 70/30 split and average 10 times to report the accuracy. In the given example, we can see a warning message saying that the program does not converge. Check the sklearn document to figure out what options we have in the field of "solver." A solver is an optimization algorithm to compute the optimal solution numerically. Which optimizer does not have the convergence problem?
9. Build a 3-layer neural network using Keras for wine dataset. Vary the hidden units from 10 to 100 in the increment of 10 to observe the change of accuracy

along with the number of hidden units. As usual, repeat the experiments 10 times to obtain the average accuracy. Use 10 epochs to train the network.

10. Build LeNet-5 by using keras to classify the MNIST dataset.