First, I defined the kernel size and number of inputs, outputs, and channels. In my case, the kernel size was 3*3 and number of inputs were 28*28 and outputs were 10 and channels were 8. I also defined the softmax and relu function. I used relu in the hidden units because relu worked better than sigmoid and tanh. Next I initialized the filter, weight and bias and used the stochastic gradient descent. In the stochastic gradient descent, I calculated the forward step and backpropagation and updated the parameters with 50000 iterations. For the first 10000 iterations, the learning rate was 0.05 and for the next 20000 iterations, it was 0.005 and for the final 20000 iterations, it was 0.0005. Finally, I calculated the test accuracy and it was 94.4% but test accuracy will change because of the randomness of the initialization.