

1. [2 marks]

`python mnist_deep.py`

Copy the last lines of the output (test performance and confusion matrix) into your answers.

How does the accuracy compare with that of the MLP model from Assignment 1?

Which three digits are most likely to be misclassified?

test accuracy 0.9902

```
[[ 969  0  0  0  0  0  3  1  6  1]
 [  0 1129  2  2  0  0  1  0  1  0]
 [  0  3 1025  2  1  0  0  0  1  0]
 [  0  0  1 1007  0  0  0  0  2  0]
 [  0  0  3  0 967  0  1  1  0 10]
 [  2  0  0 13  0 872  2  1  0  2]
 [  7  3  0  1  2  3 939  0  3  0]
 [  0  4 11  6  0  0  0 1003  0  4]
 [  1  0  4  2  0  2  1  2 960  2]
 [  2  1  0  5  3  2  0  5  4 987]]
```

Comparing with that of the MLP model from assignment 1, the test accuracy greatly increased to 0.9902. The following three cases are most likely to be misclassified 4->9, 5->3, 7->2.

2. [4 marks]

The network used in this code is similar to the LeNet model shown in Slide 5 of the supplementary slides, but with some differences. The original image is 28x28 instead of 32x32, but zero-padding with $p=2$ ensures that the first convolutional layer is also 28x28. The first pooling layer reduces the size to 14x14, and zero-padding again ensures that the second convolutional layer is also 14x14. You can find out other details about the network by looking through the code (including comments).

For each layer in the network (first and second convolutional layer, fully connected layer and output layer) calculate the number of:

- weights per neuron?
- neurons?
- connections? (including connections from the "zero-padding" regions)
- independent parameters?

According to the code for the following calculation->

conv layer 1:

5*5 patch size of 1 channel, 32 features.

weights per neuron? $1 + 5*5*3 = 76$

neurons? $28*28*32 = 25088$

connections? $28*28*32*76 = 1906688$

independent parameters? $32*76 = 2432$

conv layer 2:

5*5 patch size of 32 channel, 64 features.

weights per neuron? $1 + 5*5*3 = 76$

neurons? $14*14*64 = 12544$

connections? $14*14*64*76 = 953344$

independent parameters? $64*76 = 4864$

full connected layer:

7*7*64 tensor map to 1024 neurons

$7*7*64*1024$

Readout layer:

1024 neurons to 10 digits. $1024*10$

3. [4 marks]

a. In the context of CNN with ReLU, briefly explain (in one or two sentences) why Max Pooling may be a good strategy.

b. List two different methods of Data Augmentation for image classification.

c. Briefly explain how dropout simulates a bagged ensemble of network architectures.

a. Max pooling is to down-sample an input. It helps over-fitting. It significantly reduces computational complexity and helps the network achieve spatial invariance.

b. Crop: we just randomly sample a section from the original image. Scale: the image can be scaled outward or inward

c. Neurons are randomly chosen to not be used. So dropout forces the network to achieve redundancy that some features are missing. It simulates ensemble of different network architectures.