Answers For Assignment 2

Part 1:

QUESTION 1: What are the problems with sigmoid?

ANSWER: sigmoid has variety of problems.

1. The function isn't centered around 0.
2. Vanishing gradient. On multi-layerd deep networks the gradient becomes very close to 0.
So, weight doesn't get updated much leading to very slow convergence
also, the gradient might reach 0, so learning wont occur.

****QUESTION 2.1****: What are the accuracy and loss values? Explain the loss and accuracy graphs

****ANSWER****: The accuracy value is 10.28 and the loss value is around 2.570895195007324

****QUESTION 2.2****: What are the accuracy and loss values? Explain the loss and accuracy graphs.

****ANSWER****: The accuracy is 11.42 and the loss value is 2.334763431549072

In the graph value we can see that throughout epochs
0-8 the accuracy isn't improving and after epoch 9 to 14 there is an improvment but not much (11.30 to 11.34)
that because the dim of the network is small (only 5), and the learning rate is also too small (0.0001)
In the loss graph we can see the graph barerly improving the loss value from 2.36 to 2.33
that because the dim of the network is small (only 5), and the learning rate is also too small (0.0001)

****QUESTION 2.3****: What are the accuracy and loss values? Explain the loss and accuracy graphs.

****ANSWER****: The accuracy is 12.32 and the loss value is 2.3116324186325072

We can see That the accuracy is **slightly** improving over the course of 14 epochs, that because the learning rate is very low.
The accuracy is improving faster from the previous test because the hidden dim is now 100 instead of 5.
The loss is also slight improving over the course of 14 epochs, for the same reasons as the accuracy.

****QUESTION 2.4****: What are the accuracy and loss values? Explain the loss and accuracy graphs.

****ANSWER****: The accuracy is 23.64 and the loss value is 2.278033423423767

Over the duration of 14 epocs the accuracy is slightly improving. the accuracy is now better than the previous
test and thats because the number of hidden dim is 200 instead of 100 in the previous test.
the accuracy is slightly improving because the learning rate is small.
The loss is slightly improving and is better than the previous test for the same reasons as the accuracy.

****QUESTION 3.1****: What are the accuracy and loss values? Explain the loss and accuracy graphs.

****ANSWER****: The accuracy is 10.620000000000001 and the loss value 2.3410150527954103

We can see that the best accuracy is between epoch 2 to 5, but the difference is minimal.
That because the learning rate is so small (really very close to 0), so the gradient doesn't improve and no learning occurs.
The hidden dim in this case doesnt help to improve much the accuracy.
We can see that the loss improvment is minimal, thats for the same reasons as the accuracy.

****QUESTION 3.2****: What are the accuracy and loss values? Explain the loss and accuracy graphs.

****ANSWER****: The accuracy is 92.09 and the loss is 0.2875345215201378

We can see that the accuracy is signifanctly improving for the first 5 epochs, afterwards its still improves but at a slower pace (because there isnt much to improve, the accuracy is high).
We got to that accuracy because the learning rate is bigger than the previous tests, and is more fitting to the model (not too big, not too small).
The loss at the graph decreases fast for the first 6 epochs (that because the learning rate is big), afterwards its decreases slowly.
In comparison to the previous tests, the loss is getting significantly smaller, that is because the learning rate is bigger in comparison and learning occurs.

****QUESTION 3.3****: What are the accuracy and loss values? Explain the loss and accuracy graphs.

****ANSWER****: The accuracy is 51.01 and the loss value is 2.1381322622299193

In the graph we can see the accuracy constantly improving over the course of 14 ephocs. That’s because the learning rate is not too small,
and learning occurs. In comparison to the previous test, the accuracy is lower because the learning is small.
The loss value slightly decreases over the epochs, that because the learning rate is low.

****QUESTION 4:**** : Suggest a way to improve the results by changing the networks's architecture

****ANSWER****: From the previous examples, we can see that the best test was when the lr was 0.1,
But, we can also see, from the previous tests that when the lr was constant and we "played" with the hidden dim, that the best result was when the hidden dim was the highest (200)
So, a way to improve the result would be to add more hidden dimensions while keeping the learning rat e at 0.1.
A way to really improve the results would be to use the ADAM optimizer instead of STG, as was shown in class, the ADAM optimzer may outclass the STG (depending on the situation)
and will probably improve The accuracy and the loss value.
Another thing that may improve the result may be normalization of the data, as was taught, greatly increases the learning.

****QUESTION 5:**** Where are the bright regions? why?

****ANSWER:**** the bright regions are in the center of the pictures, because the numbers are in the center of the picture, and the network is looking for numbers.

Part 2:

Question 1

Analyze your results from experiment 1.1. In particular,

1. Explain the effect of depth on the accuracy. What depth produces the best results and why do you think that's the case?
2. Were there values of `L` for which the network wasn't trainable? what causes this? Suggest two things which may be done to resolve it at least partially.

****ANSWER****:

1. At the training accuracy graph we can see the getting better after each epoch (for l_2 and l_4, for l_8 and l_16 the accuracy doesnt improve). As for the test accuracy we can see that for the first 10 epochs the accuracy for L_2 and L_4 is improving, after wards its stays more of the same, where as the accuracy for L_8 and L_8 doesnt not improve and stays the same.

As shown in the graph, The best accuracy results coming from the depths 2 and 4.
for the training result we can see that between depth 2 and depth 4 there isnt much any different, but at the test result we can see that depth 4 is more accurate. In contrast, we can see that depth 8 and depth 16 shows poor results (are untrainable).
The cause for the poor results in depth 8 and 16 may be because the depth is too much and causing exploding gradient/vanishing gradient.
We think depth 4 produces better results than depth 2 because there are more layers for the CNN to work with, which as taught in class, results in better performance.

2. The values of L which were not trainable are L_8 and L_16.
What may cause them to not be trainable is fact that they are too deep, which may cause exploding gradient/vanishing gradient.

We can resolve the problem partially with the following 2 steps:
1) We can normilaize the batches, which will improve performance
2) For exploding gradient we can use the clipping gradient method, which will partially help in this case.

Question 2

Analyze your results from experiment 1.2. In particular, compare to the results of experiment 1.1. For a spesific value of L, how the preformance change with respect to K? Does we saw the same phenomena in 1.1 for a specific value of K?

****ANSWER****:

First, as seen in experiment 1.1, the L_8 isnt learning for the presented values of k, so the same phenomena happens for any presented value of K.

For L_4 we can see that the best training results comes from K128 and K256.
The best test results comes from K256, but the second best results comes from L32. In comparison to 1.1 the training accuracy is higher for the values K128 and K256, but the test accuracy is similiar for K128 and K256, and isnt much different with K32 and K64 in comparison to experiment 1.1
The best Accuracy comes from K256.
The perfomance doesn't change much for a specific value of K, but in this case, for better performance we will get the higher K.

For L_2 we can see that best training results comes from K64 and K256.
The best test results comes K256 and K64 (the differences are minimal).
In this case, in comparison to experiment 1.1 we dont see much changes in the training accuracy results and the test accuracy results for the presented values of K.
The perfomance doesnt change much in this case for a specific value of K.
We can see that the same phenomena happens in compirason to experiment 1.1 for values of K=32,64, 256 in the test results and training result, as for K=128 we got lower accuracy result in compirason to experiment 1.1 (for training and the test).