HW3 Report

Q1.

1. Hidden layer dimensions

	2	4	8	16	64	512
Test Acc.	0.5780	0.7010	0.7900	0.7980	0.8100	0.8050
Epochs	59	84	25	20	10	3

   a. **Do the test accuracies increase with the hidden dimension? Why or why not?**
   -> The test accuracies are increasing with the hidden dimension because with more nodes in the hidden layer, the model is able to capture more information as opposed to fewer hidden nodes capturing fewer features. This way the model is able to fit a more complex function. We do observe that beyond 8 dimensions the test accuracy is not improving much. We can say that more hidden dimensions means better test accuracies until the improvement plateaus.

   It's important to note that there is a possibility that it's likely to start to overfit beyond a point if the number of layers are fixed and if there's no regularization. So the model will start memorizing the training data instead of generalizing it. Due to this, finding the right number of hidden layer dimensions for our problem becomes imperative.

   b. **Does the model converge faster with increasing hidden dimension? Why or why not?**
   -> It does converge faster as more hidden nodes allow better flexibility for the model in the solution space. But there is a good chance that if there is no regularization or if the dataset is small, it's starting to overfit. And with the cora dataset that indeed could potentially be the case. From 8 dimensions onwards, we observe that there is not much change. So in short, it does converge faster, but beyond a point, it's not beneficial.

2. Number of layers

	1	2	4	8	16
Test Acc.	0.7470	0.8050	0.7700	0.6110	0.1440
Epochs	24	17	24	87	74

   a. **Do the test accuracies increase with the number of layers? Why or why not?**
   -> There is any clear continuous increment happening. So we cannot say that an increase in the number of layers will lead to an increase in test accuracy. This

could be due to the model overfitting. The dataset is too small for such a complex model.

b. Does the model converge faster with increasing the number of layers? Why or why not?
-> Not at all. Actually, it's getting progressively worse. This could work better with a bigger dataset with more features but not for this case in particular. In general, this parameter could be useful to play around with. But in this case, it works well with 1 or 2 layers.

3. Learning rate

	1	0.5	0.1	0.01	0.0001
Test Acc.	0.6420	0.7750	0.8030	0.8040	0.8040
Epochs	88	61	3	0	0

a. How do the test accuracies change with different learning rates?
-> The test accuracies are improving and after a point plateauing.

b. How do the convergence rates change with different learning rates? Explain your answer.
-> The time it takes to converge is getting progressively much better. This could be due to a simple model and simple dataset needing to take only small gradient steps to reach its nearest local minimum in the solution space. The larger learning rate could be off-shooting the gradient steps by a large margin causing it to oscillate back and forth a lot before it converges finally.