DAMA61 - 4th assignment

For both exercises

1 - 3
All the directions of the exercises were followed

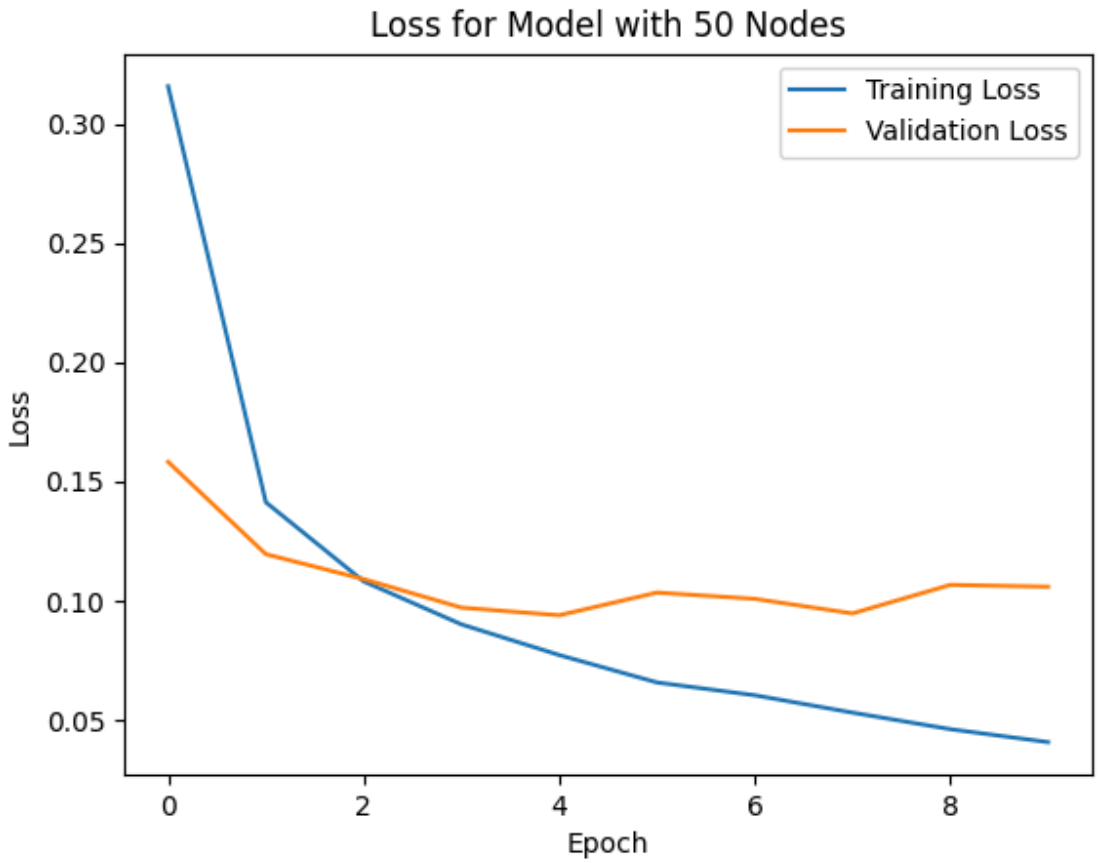
4
As printed by the python console when running the code, the total parameters of the 50 node run were 44860 which were all trainable, and the same applies for the 200 node code run where there were 239410 parameters.

5
When running the code, it was noticed that the training stopped before reaching the maximum specified number of epochs (which was 100) which was expected due to the early stopping mechanism implemented with a patience of 5 epochs.
The same was noticed when running the 200 node model, but the training stopped at a much larger number of epochs because we used the Dropout regularization with the value of 0.5, thus giving each neuron a 50% chance of being dropped out during each training iteration. Even though it takes longer to run the seems to be a very productive way to train the model because it introduces noise and uncertainty, forcing the model to learn more redundant representations and reducing the risk of overfitting to the training data and makes the network learn more robust and generalized features by reducing the reliance on specific neurons or connections

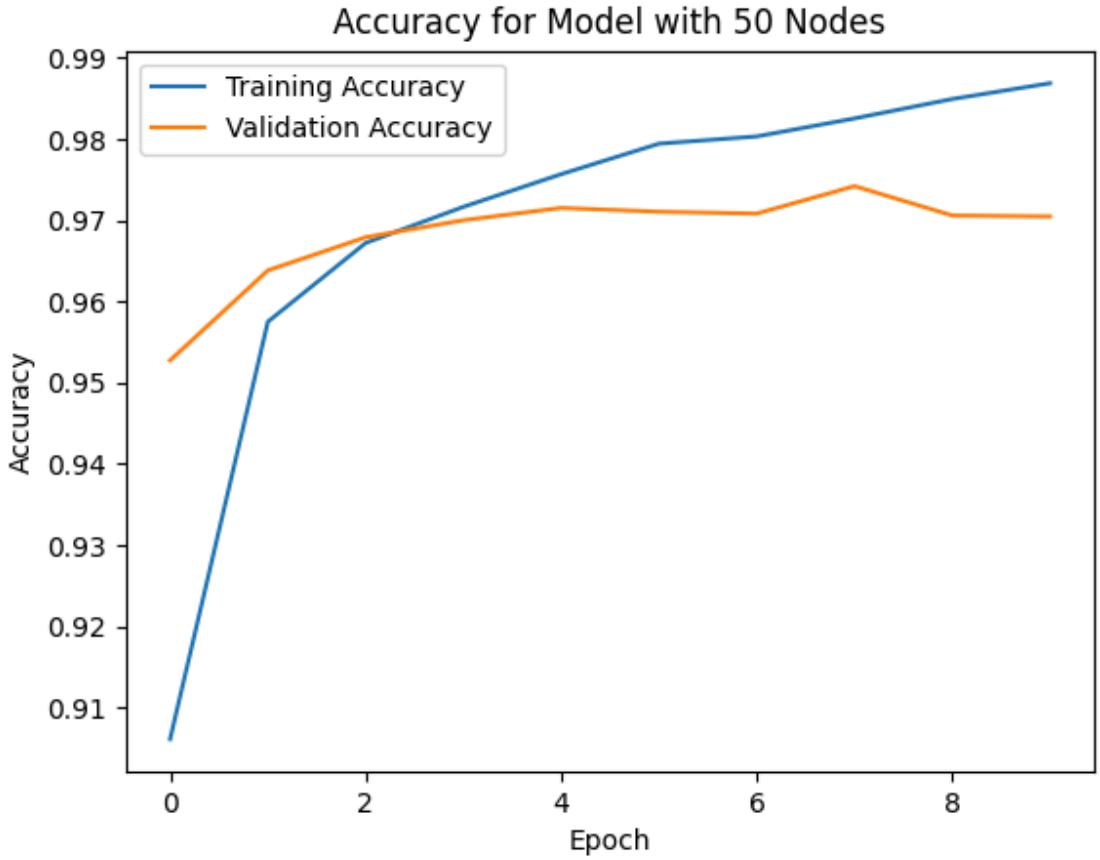
DAMA61 - 4th assignment

6

As we can see below in the two plots for Loss and Accuracy for the 50 node run, as the training progresses over epochs, the loss drops significantly for the training set but not in the same way for the validation set. The same applies for accuracy, as the training progresses over epochs, it reaches a high level for the training set but not for the validation set. As the training progresses they are expected to follow a similar trajectory if the model is optimal. As we can see in the plots they even interject so we can see that the model with 50 nodes is not optimal.

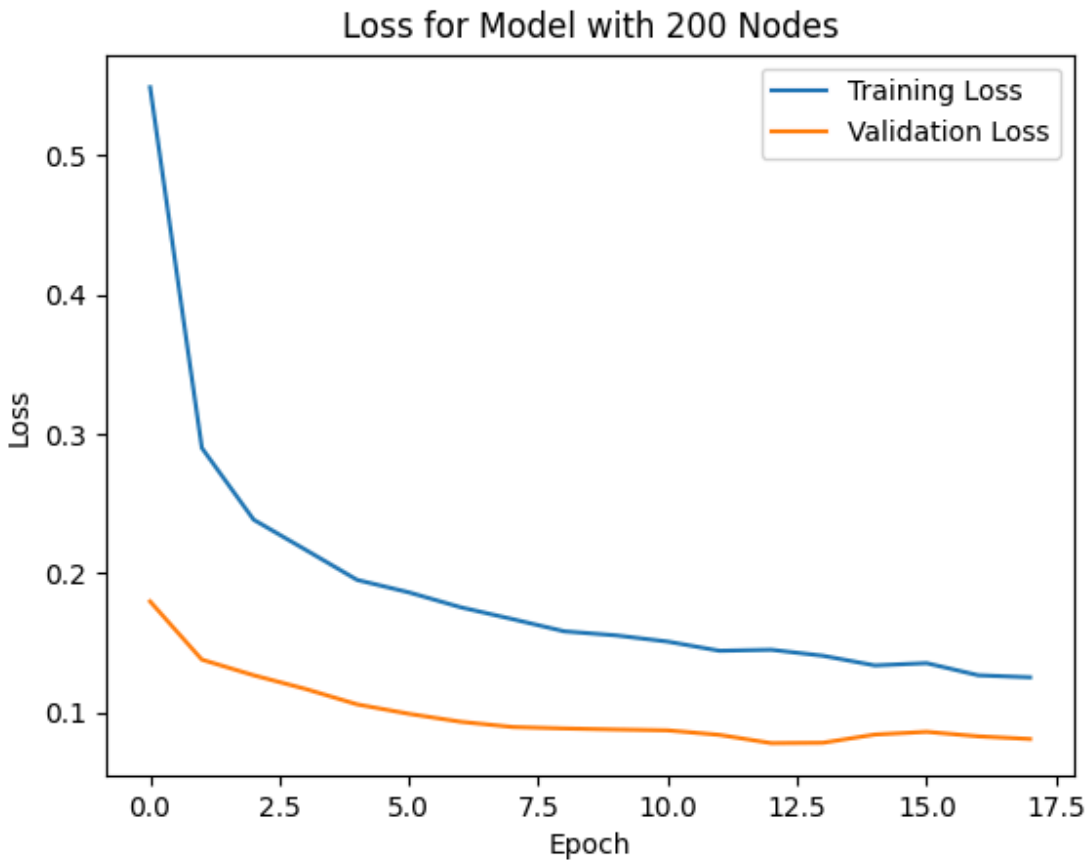


DAMA61 - 4th assignment

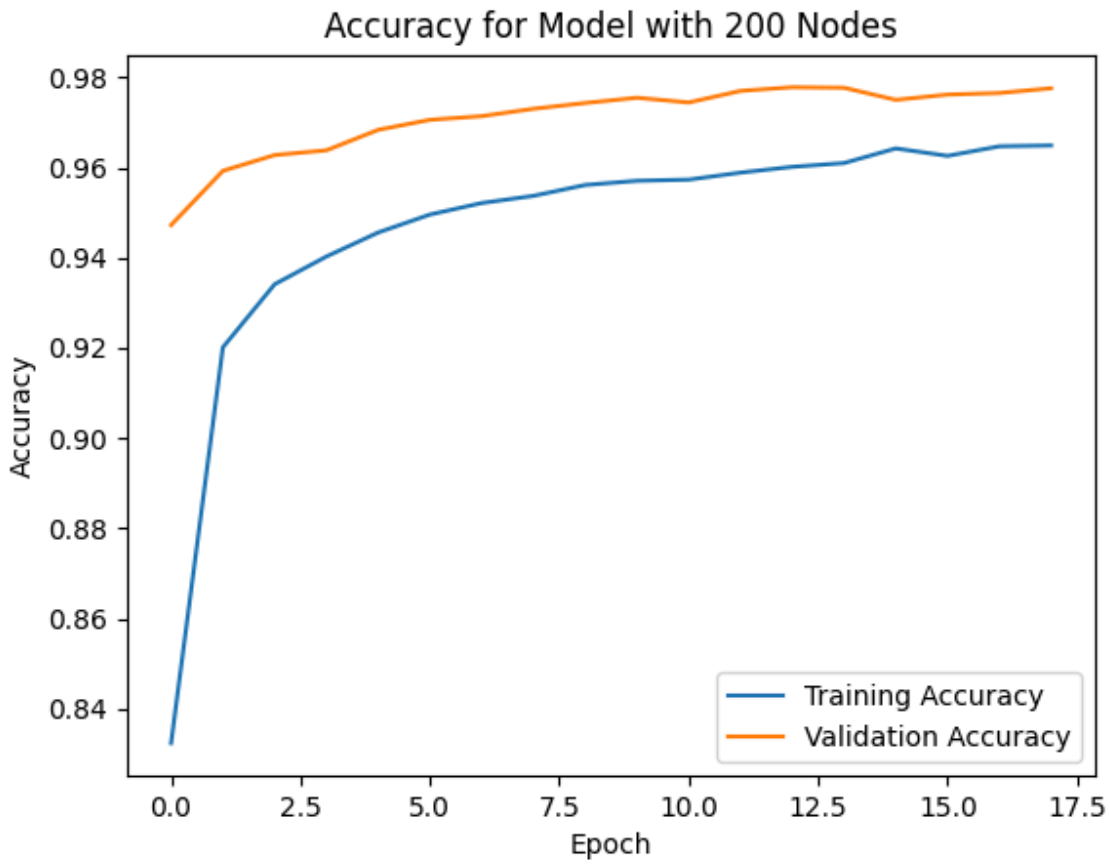


And below there are the two plots for Loss and Accuracy for the 200 node run. In this training model's plots we see a more expected trajectory as the training progresses over epochs. The lines in the plot follow a similar flow and they seem to get closer together which means that the model fits the validation data in a good manner without overfitting them.

DAMA61 - 4th assignment



DAMA61 - 4th assignment



7

Despite the observations we made above, we see that when we run the two models with the test data the loss and the accuracy are very similar with the 200 node model being slightly better than the 50 node one.

```
Test Loss for 50 nodes: 0.10910116136074066
Test Accuracy for 50 nodes: 0.9696999788284302

Test Loss for 200 nodes: 0.09260601550340652
Test Accuracy for 200 nodes: 0.9761000275611877
```