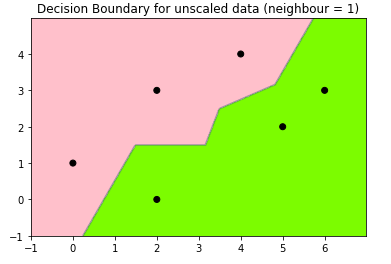
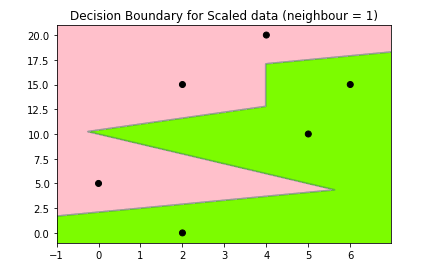
Question 1:

1. See plot below



1. See Plot below



1. Scaling has a bad effect on our data. Because we scaled by 5 and not moving . We put a heavier emphasize to . Making the scaled data has much less smooth decision boundary compared to the unscaled ones, making it harder to generalize.

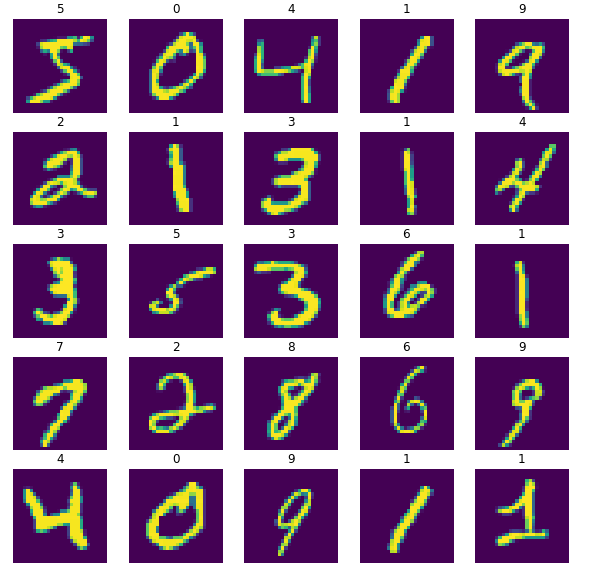
Question 2:

Question 4:

1. Number of training images: 60000

Number of testing images: 10000

Image size: (28, 28)



1. Model 1 for dense=10 no activation function, all are Dense layers, 10% of the training data for validationTest loss: 0.12677660584449768

Test accuracy: 0.972000002861023

Model 2 for dense=10 no activation function, add two Dropout layers, add the epochs to 40, 10% of the training data for validation

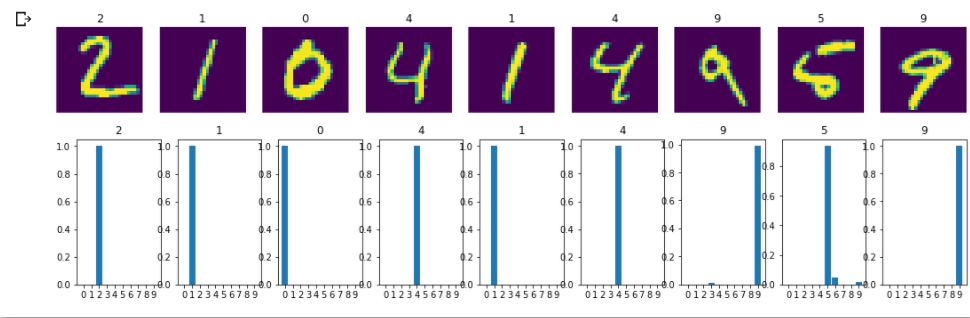
Test loss: 0.11696485430002213

Test accuracy: 0.9735999703407288

Model 3 for dense=10 no activation function, a deeper network, and add the epochs to 40, 15% of the training data for validation

Test loss: 0.1272159367799759

Test accuracy: 0.9713000059127808



As we can tell from the graph, our model did a pretty good job (excelled) in predicting 1 and 2 and 4. But for terribly written 9 and 5, our model struggled a little bit about the right answer. Overall, our model did a pretty good job, as the first 10 handwritten digits suggest.

Question