1. Go through the following two tutorials.  Answer these questions:



Why is the cross entropy optimization function used in the text classification exercise?

Binary cross entropy is used for the text classification in the text classification exercise because it is a binary classification problem, meaning there are only 2 possible outputs so this is the best optimization function for this test.



Why is the text classification model ‘overfit’ to the training data?  How might you prevent this and have generalized learning? Why is general learning better than high accuracy with the training data?

It is overfit because it is learning over the test data for too long. In order to prevent this, we could lessen the number of epochs or the enlarge the size of the test data to give it more to look at. We want general learning because that means when presented with raw data that the neural net hasn’t encountered before, it is more likely to predict correctly.



What does the first layer in the image classification neural net do when it flattens?  How are images represented when they enter the neural net?

The first layer basically takes the individual pixels of each image and extracts them from a 2 dimensional array and places it in a one dimensional array. All it is doing is destructuring the image so that we can do some learning on it.



Why is a softmax activation function used as an output layer in the first image classification test and a sigmoid function used in the second text classification test?  Why are there ten neurons in the softmax layer and only 1 in the sigmoid layer?

The sigmoid activation function is used with only 1 neuron because we are only going to have a single output between 0 and 1 which is what the sigmoid function will do for us. However, a softmax activation function with 10 neurons will give us an output with our confidence between each of the 10 possible categories.



Which is better: getting really high accuracy on test data, or getting  really high accuracy on training data? Why?

Getting really high accuracy on test data is more important because that means that our neural net has been trained generally enough that it predicts accurately on data it has never seen before.



Vary each of the following hyperparameters and report what the difference in accuracy and loss is for this particular dataset.  Try and understand why varying that hyperparameter makes that difference and record your answer.



Optimizer



Loss function



Number of epochs



Number of batches or minibatches



Number of hidden layer units (neurons)



Type of neuron (vary activation function)



Class challenge: Get the best accuracy for test data on:



Text classification: <https://www.tensorflow.org/tutorials/keras/text_classification>



Image recognition: <https://www.tensorflow.org/tutorials/keras/classification>