Part 1: MNIST Logistic Regression

Hyperparameters:
I modified the epoch and learning rate hyperparameters. For learning rate, I used a staircase schedule with exponentially decreasing learning rate. This is because I found that early convergence was very slow if I used a consistently low learning rate, while allowing the learning rate to decrease in the later epochs can help reduce overfitting. I started with a ratio of 0.2 as higher ratios (such as 0.7) sometimes led to divergent behaviour, and then it decays by 0.6 per epoch. I picked epoch as 10 because it allowed to model to converge to 99.9% on training data within a reasonable amount of time (90 seconds on my computer). i set the batch size to 32 which is also the default, so I didn’t actually change anything there.

Model:

```
[tf.keras.layers.Conv2D(8, (3, 3), padding = ’same’, input_shape =
(28, 28, 1)),
tf.keras.layers.MaxPooling2D((2, 2)),
tf.keras.layers.Conv2D(8, (3, 3), padding = ’same’),
tf.keras.layers.MaxPooling2D((2, 2)),
tf.keras.layers.Flatten(),

tf.keras.layers.Dense(128, activation=’relu’),
tf.keras.layers.Dense(64, activation=’relu’),
tf.keras.layers.Dense(10, activation=’softmax’)]
```

The basic structure of this model was taken from the convolution slides in lecture. It starts with a 3x3 convolution with 8 filters, apply a 2x2 max pooling to reduce the size, another 3x3 convolution with 8 filters, and another 2x2 max pooling. The convolution layers can detect local patterns. The pooling layers reduce the size of the model to reduce complexity without sacrificing much accuracy, because downsizing an image still results in a similar image. After this, two more densely connected layers with recitfied linear activiation before the output layer was enough to handle the complexity of the MNIST dataset.

Results:
The resulting model has 99.9% accuracy on the traning set and 99.0% on the testing set. The 0.9% difference means that there was some overfitting present, but it wasn’t too serious. The model took 90 seconds to train on my computer running on a 12 thread CPU which is a reasonable timeframe.