STAT665 Lab 5

Bo Song

For the MNIST set, I chose 1 hidden layer with 100 nodes, used crossEntropy cost function, 20 epochs, mini batch size is 10, learning rate eta is 0.25 and lambda is 5. The accuracy rate I achieved is 97.82%. For the CIFAR set, I chose 1 hidden layer with 200 nodes, used crossEntropy cost function, 20 epochs, mini batch size is 10, learning rate eta is 0.03 and lambda is 1. The accuracy rate I achieved is 51.72%.

The selection of cost function affects the accuracy very much. The quadratic cost function performs 1% worse than that using crossEntropy cost function. When the number of nodes or depth of the model increases, which means our network becomes larger and more complex, the accuracy at the first few loops of epochs drops a lot. However, when we loop further, the accuracy increases rapidly. Halving the learning rate makes the accuracy rate increases slowly but steady. While doubling makes the rate increases quickly but vibrates. In the first few rounds of epochs, the large regularization rate penalizes the accuracy rate very much. However when the epoch goes to 20 or more, an appropriate regularization rate will lead to a high accuracy rate, avoid overfitting. There are some categories easy or difficult to distinguish. For MNIST set, 5 is the most difficult one to recognize. For CIFAR-10 set, label 3 cat is the most difficult one to recognize. I think this result make sense since cat and ‘5’ are also difficult for human to recognize, comparing to other digits or objects.