Homework 7

Kaizhao Liang(kl2)

Part 1. **Conceptual Questions**

1. The sample space for this problem is $\{' ',' \#',' +'\}^{784}$.
2. A multi-layer perceptron is more expressive. The VC dimension of the perceptron is only d+1, while the neural network with one hidden layer can represent any linear or non-linear functions.
3. A multi-layer perceptron is more likely to over-fit the data.Because the perceptron can only learn a linear separator and there is no way it can overfit the non-linear data, while the multilayer perceptron can learn more complex structures, thus more likely to overfit.
4. No, because the samples with each label do not seem linearly separable to the samples with rest of the labels. Thus, perceptron will always make mistakes on some of the training samples.

Part 2. **Implementing a Perceptron**

1. The best set of learning rate with decay rate $O(\frac{1}{n})$ tuned by cross-validation for 10 separators is $[0.01, 0.001, 0.01, 0.001, 0.01, 0.1, 0.001, 0.001, 0.01, 0.001]$ in order for each label. The best initialization of the weights is to initialize them all to zeros and the randomization of the training set after each epoch helps the convergence to converge.
2. 25 epoches are necessary to train my perceptron.
3. 94.8% accuracy on the training data set.
4. 84.5% accuracy on the testing data set.

Part 3. **Using a Multilayer Perceptron**

1. I tried: [785] with tanh, [785] with sigmoid, [785] with relu, [1570] with tanh, [1570] with sigmoid, [1570] with relu, [785,785] with tanh, [785,785] with sigmoid, [785,785] with relu.**[1570] with relu** has the best performance.
2. 1000 epochs are necessary to train your multilayer perceptron.
3. 99.2% accuracy on the training data set.
4. 90.2% accuracy on the testing data set.
5. Both the training accuracy and testing accuracy are higher than the ones in the second question. It's understandable because neural network is more expressive than the perceptrons, thus fits the both the training data and the testing data better.