IEOR 290 002 - Spring 2017
Homework 7
Due 05/01/2017 at 11:59 pm

Instruction

- We suggest you type your answer using LaTeX or Microsoft Word.
- This homework contains both theoretical and coding parts. For questions with `[coding]` mark, you need to write a small program to answer the question. We recommend you use Python, R or MATLAB to do the problem. Append your code after the answer.
- Submit a single PDF of your write-up on bCourses. Make sure your answer is legible.

1. **(Neural network on MNIST)**

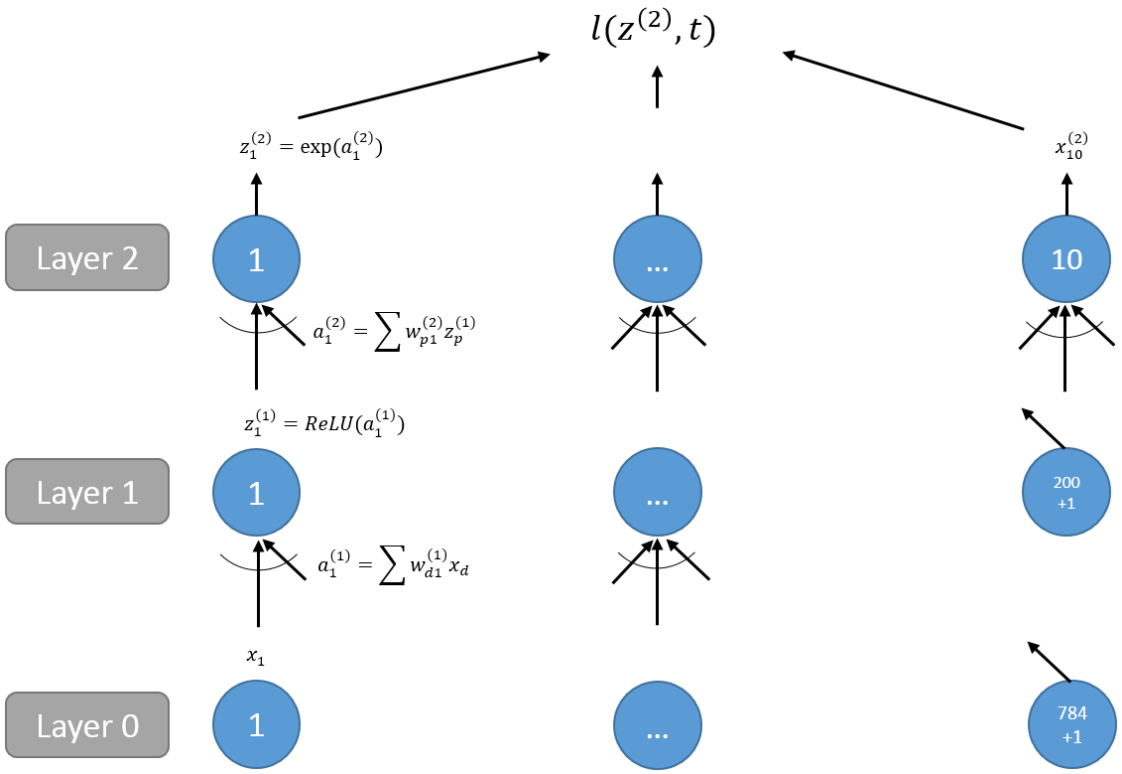
   `[Coding]` In this final homework we'll train a simple neural network on the handwritten digit recognition MNIST dataset. Consider a neural network with one hidden layer of size p. For the hidden layer use the rectified linear unit (ReLU) function as the activation function given as

   $$z_q^{(1)} = \max(0, a_q^{(1)}) \quad q = 1, \ldots, p+1$$

   and for the output layer use the softmax function as the activation function with cross entropy loss given as

   $$\begin{aligned} z_k^{(2)} &= e^{a_k^{(2)}} \quad k = 1, \ldots, 10 \\ l(\boldsymbol{z}^{(2)}, \boldsymbol{t}) &= -\sum_{k=1}^{10} t_k \ln \frac{z_k^{(2)}}{\sum_{j=1}^{10} z_j^{(2)}} \end{aligned}$$

The neural network with $p = 200$ is roughly depicted below:



In this problem you need to

- Train the model using the above neural network.
- Report the training and testing accuracy. Compare the accuracy to the ridge regression with random features in Homework 4.
- Plot the first mis-classifed data in the testing data and visualize the data as an image. What is your prediction before seeing the true label? Are you doing better than the neural network?

Some programming tips if you use Python

- Use `sklearn.neural_network.MLPClassifier`
- You only need to specify the activation function for the hidden layer as ReLU. The activation function and loss function for the output layer is automatically softmax with cross entropy loss.
- Experiment with different `MLPClassifier` options: `hidden_layer_sizes`, `solver`, `random_state`, `early_stopping` etc. to get a better testing accuracy.