Deep Learning CS583 Fall 2022
Homework Assignment 1

October 10, 2022

Instructor: Jia Xu

Student name: _____

Student ID: _____

Student email address: _____

- **Read these instructions carefully**
- Fill-in your personal info, as indicated above.
- There is one question, in total 10 points.
- You should work on the assignment only by yourself.
- Submit your zip file **by 1AM Oct 24th** on Canvas under Homework.

good luck!

1 Question

Choose a software framework. A few candidates:

- Marvin "http://marvin.is/"
- Tensorflow "https://www.tensorflow.org/"
- Caffe "http://caffe.berkeleyvision.org/"
- Pylearn2 "http://deeplearning.net/software/pylearn2/"

Download the benchmark dataset MNIST from "http://yann.lecun.com/exdb/mnist/". Implement multiclass logistic regression and try it on MNIST. Comments: MNIST is a standard dataset for machine learning and also deep learning. It is good to try it on one-layer neural networks (i.e., logistic regression) before multilayer neural networks. Downloading the dataset from other places in preprocessed format is allowed, but practicing how to read the dataset prepares you for other new datasets you may be interested in. Also, it is recommended to try different initializations and learning rates to get a sense of how to tune the hyperparameters (remember to create and use validation dataset!).

Check the tutorials for some of the parameters (e.g., "https://machinelearningmastery.com/how-to-develop-a-convolutional-neural-network-from-scratch-for-mnist-handwritten-digit-classification/").

- Build a three-layer feedforward network: [6 points]

$$x \to h^1 \to h^2 \to p(y|h^2). \tag{1}$$

  The hidden layers h^1 and h^2 have dimension 500. Train the network for 250 epochs[1] and test the classification error. Do not use regularizations. Plot the cross-entropy loss on the batches and also plot the classification error on the validation data.

- Repeat the above experiment, but train the network with the following regularizations and compare with the results in the previous experiment:

  L2 regularization [4 points]

  Comments: no need to implement them on your own; the software framework typically provides implementations for L2 regularization and dropout.

[1] Each epoch is a pass over the training data. Suppose you use batches of size b, and the training data set has n points, then an epoch consists of n/b batches. Note that you can divide the data set into batches and then round-robin over the batches. You can also randomly sample, say 64 points for each batch. Either way is OK, and typically there is no performance difference between them. When these batches are randomly sampled, it is possible that some points are not in any of them, but we still call these batches a pass over the data. Acknowledgement: Thanks Princeton COS 495 for this homework.