Computer Science COMP-8590-1 --- Assignment #3

Question #1 Download the benchmark dataset MNIST from http://yann.lecun.com/ exdb/mnist/. Implement multiclass logistic regression and try it on MNIST.

Use your favorite deep learning platform. Marvin, Caffe (http://caffe.berkeleyvision.org), TensorFlow (https://www.tensorflow.org), Theano, Torch, Lasagne. For more DL platforms, see the link: http://deeplearning.net/software_links/.

Comments: MNIST is a standard dataset for machine learning and also deep learning. It's good to try it on one shallow neural networks before deep 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.

1. Try the basic minibatch SGD. It is recommended to try different initializations, different batch sizes, and different learning rates, in order to get a sense about how to tune the hyperparameters (batch size, and, learning rate). Remember to create and use validation dataset!. it will be very useful for you to read Chapter-11 of the textbook.

2. It is recommended to try, at least, another optimization method (SGD momentum, RMSProp, RMSProp momentum, AdaDelta, or Adam) and compare its performances to those of the basic minibatch SGD on the MNIST dataset. Which methods you want to try and how many you want to try and compare is up to you and up to the amount of time you have left to complete the assignment. Remember, this is a research course. You may want to read Chapter-8, which I will cover this week.

Question #2 Consider the L^2-regularized multiclass logistic regression. That is, add to the logistic regression loss a regularization term that represents L^2 norm of the parameters. More precisely, the regularization term is

$$\lambda \Sigma_i (\|w_i\|2 + \|b_i\|2)$$

where $\{w_i, b_i\}$ are all the parameters in the logistic regression, and $\lambda \in R$ is the regularization hyper-parameter. Typically, λ is about C/n where n is the number of data points and C is some constant in [0.01,100] (need to tune C). Run the regularized multiclass logistic regression on MNIST, using the basic minibatch SGD, and compare its results to those of the basic minibatch SGD with non-regularized loss, in Question #1.

Question #3 Going above and beyond Question-1 and Question-2, investigate the basic minibatch SGD with, at least, another regularization method discussed in class (L^1, data augmentation, noise robustness, early stopping, sparse representation, bagging, or dropout). Currently, L^2 norm, early stopping, and dropout are the most frequently used regularization methods. You may need to read Chapter-7, which I have started to cover in class. You may even try CNN if time allows you.