Deep Learning (Homework #1)

Due Data: March 17, 2017

1. **Neural Network**

In this exercise, you will implement a Deep Neural Network (DNN) model for classification of Street View House Number (SVHN) dataset. SVHN is a real-world image dataset for developing machine learning and object recognition algorithms with minimal requirement on data preprocessing and formatting. This dataset is obtained from house numbers in Google Street View images.

The file SVHN.mat contains four matrices (train_x, train_label, test_x, test_label). And all images are grayscale. The input images (train_x, test_x) are reshaped from 28*28 into 784*1. And all the labels (train_label, test_label) are one-hot vectors of size 10.

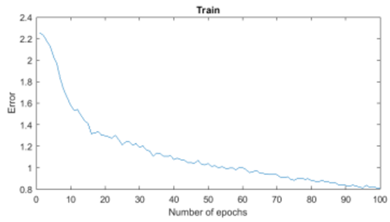
(1) Please construct a DNN for SVHN classification. Your objective function is:

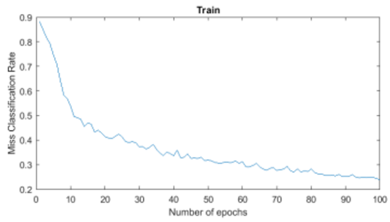
$$E(\mathbf{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_k(\mathbf{x}_n, \mathbf{w})$$

Minimize the objective function $E(\mathbf{w})$ by error backpropagation algorithm and Stochastic Gradient Descent (SGD):

$$\mathbf{w}^{\tau+1} = \mathbf{w}^{\tau} - \eta \nabla E(\mathbf{w}^{\tau})$$

You should decide the following variable: number of hidden layer, number of hidden unit, learning rate, number of iteration and mini-batch size. You have to show your learning curve, train error rate and test error rate in the report.





(2) Bonus: Regularization is one of techniques which can reduce overfitting. The idea of regularization is to add an extra term to the cost function:

$$E(\mathbf{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_k(\mathbf{x}_n, \mathbf{w}) + \frac{\lambda}{2}||\mathbf{w}||^2$$

The effect of regularization term is to make DNN prefer to learn small weights. The coefficient λ governs the relative importance of the regularization term compared with the cross-entropy term. Please minimize the objective function with regularization term by SGD optimizer, and show your learning curve, train error rate and test error rate in the report.

(3) Bonus: Adaptive Moment Estimation (ADAM) is another method that computes adaptive learning rates for each parameter. Please optimize your DNN by ADAM optimizer and compare with SGD, and show your learning curve, train error rate and test error rate in the report.

Reference

ADAM: A Method for Stochastic Optimization