Deep Learning Assignment 2
Arsalan Syed 961117-6331
22/04/2017
CDATE 3
DD2424 Deep Learning in Data Science

The analytic gradients was checked by computing the numerical values using the methods `ComputeGradsNum` and `ComputeGradsSlow` on one training example at a time. The relative error between the analytic and numerical gradients was found to be around 2e-8. Also, because of the fact that the classifier is performing better than the one in Assignment 1, the gradient computations should be reliable, otherwise the performance would be near random.

To test the effect of momentum on the model, no momentum was used and the training ran for 10 epochs. Then rho was set 0.95 and the training was to stop as soon as the accuracy on the validation set (computed each epoch) exceeded the accuracy provided when there was no momentum. On one of the experiments, the model achieved an accuracy 23.89% after 10 epochs without momentum . Using momentum made the model only require 2 epochs in order to get an accuracy of 26.95% making training around 80% faster.

While performing the coarse search, 50 different random combinations were chosen and each of them ran for 3 epochs. The range of values for lambda was [0, 0.001] and for eta it was [0.0001, 0.1]. Below are the best three models that were found from this sampling:

Eta	**Lambda**	**Accuracy on validation set**
0.032040	0.000676	0.418500
0.030010	0.000808	0.415900
0.031692	0.000576	0.414900

In the fine search, lambda ranged from [0.0001, 0.001] and eta ranged from [0.02, 0.06]. Thirty different combinations were formed from theses ranges and the number of epochs was increased to 5.

Eta	**Lambda**	**Accuracy on validation set**
0.024780	0.000531	0.429200
0.024805	**0.000929**	**0.432000**

0.034622	0.000619	0.429100

Using the parameter settings in bold above, the model was trained on all available data except for 1000 examples which were used in the validation set. This produced an accuracy of 0.5164 on the test set. In the graph showing the accuracies, the blue line is the accuracy on the training set and the red line is for the accuracy on the validation set.