KTH Royal Institute of Technology

DD2424 Deep Learning in Data Science

Assignment 1

Name: Polyxeni-Parthena Ioannidou
E-mail: ppio@kth.se

September 29, 2019

---

1 Introduction

In this assignment we built an one-layer network to classify images from the CIFAR-10 dataset. We trained the model using the Mini-Batch Gradient Descent, where we made use of a subset of training samples per epoch. We used the CIFAR-10 dataset, which consists of 60000 32x32 colour images in 10 classes. The dataset is divided into five training batches and one test batch, each with 10000 images. For each batch we ran the Mini-Batch Gradient Descent for 40 epochs, where each batch had step=100 and size=100 (so as 100x100=10000 images per batch). Moreover, a cost function that computes the cross-entropy loss of the classifier is implemented and the L2 regularization term is also applied. For computing and testing my gradients we used the function assert_almost_equal from numpy library, where it was found that the arrays were equal up to 6 decimal.

2 Results of Classification

The model was tested with 4 different parameter settings, which are presented below. For each case there are the cost plots of training and validation sets.

1. Model results for lambda = 0 and learning rate = 0.1
   -The accuracy on the training set is: 0.2745
   -The accuracy on the validation set is: 0.2316
   -The accuracy on the testing set is: 0.233

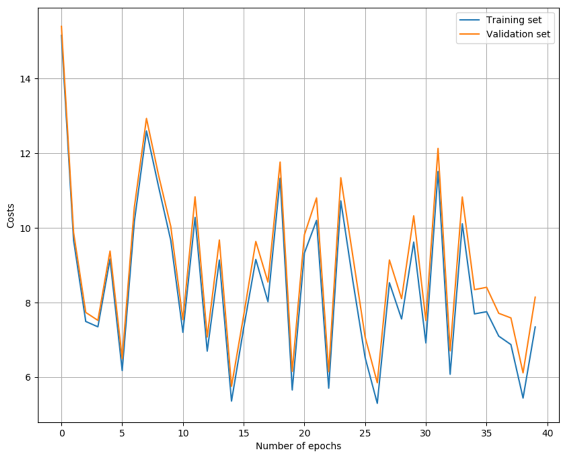

Figure 1: Costs per epoch for eta 0.1 and l = 0

2. Model results for lambda = 0 and learning rate = 0.01
   -The accuracy on the training set is: 0.4845
   -The accuracy on the validation set is: 0.3546
   -The accuracy on the testing set is: 0.3612

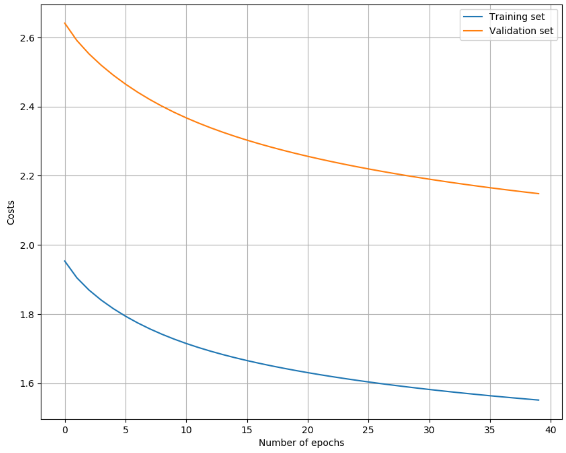

Figure 2: Costs per epoch for eta 0.01 and l = 0

3. Model results for lambda = 0.1 and learning rate = 0.01
   -The accuracy on the training set is: 0.3442
   -The accuracy on the validation set is: 0.3225
   -The accuracy on the testing set is: 0.334

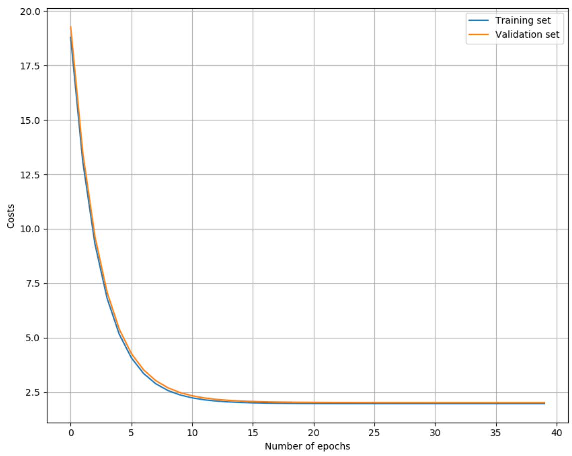

Figure 3: Costs per epoch for eta 0.01 and l = 0.1

4. Model results for lambda = 1 and learning rate = 0.01
   -The accuracy on the training set is: 0.2275
   -The accuracy on the validation set is: 0.2166
   -The accuracy on the testing set is: 0.2218

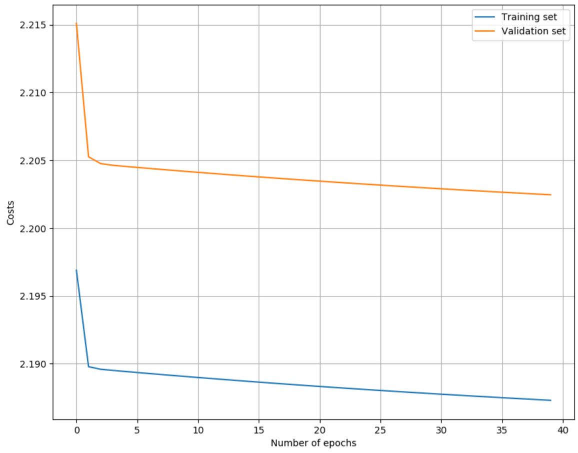

Figure 4: Costs per epoch for eta 0.01 and l = 1

From the results, we concluded that when we had higher learning rate, which means 0.1 (Figure 1) instead of 0.01 (Figures 2,3,4), the training had significant oscillations, which means that it was unstable. The plot in Figure 1 is not smooth comparing to Figures 2 and 3, as well as the test accuracy for Case 1 is much lower than test accuracy in Case 2 and 3.

Furthermore, we concluded that when we increased the regularization term from 0 (Figures 1, 2) to 0.1 (Figure 3), the convergence is better because the network generalizes more. However, if the regularization is too high , like in Figure 4, the network learns nothing, because increasing regularization term too much means that we actually penalize more the data that have not been classified correctly, and this makes the learned hyperplane almost linear which is also a bad result.

See also Figures 5-8 to see how the learned weight matrix changes for different model parameters in the 4 cases above. As we can observe, the higher the regularization the smoother the results.

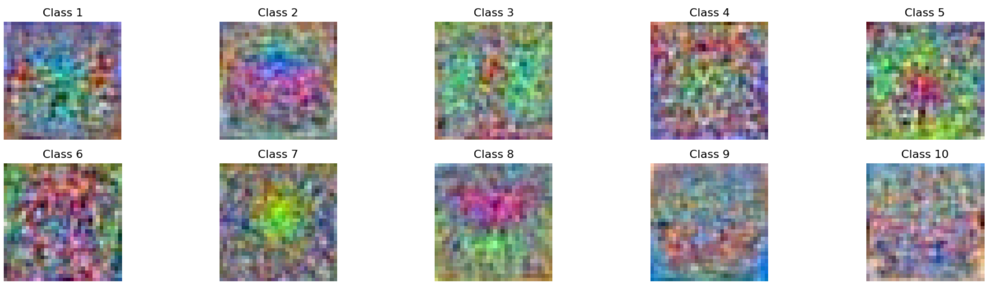

Figure 5: Learned weight matrix for eta 0.1 and l = 0

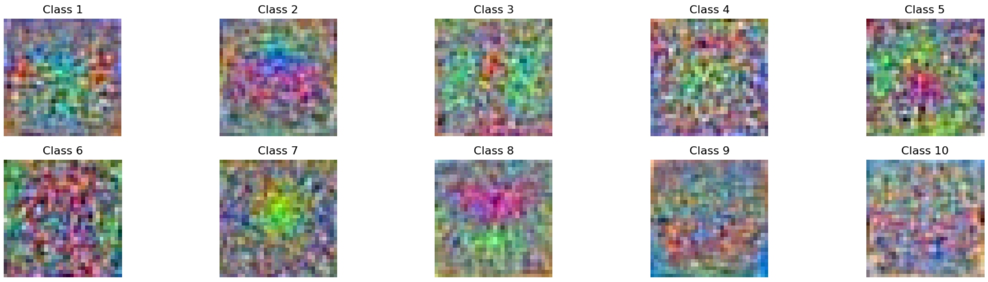

Figure 6: Learned weight matrix for eta 0.01 and l = 0

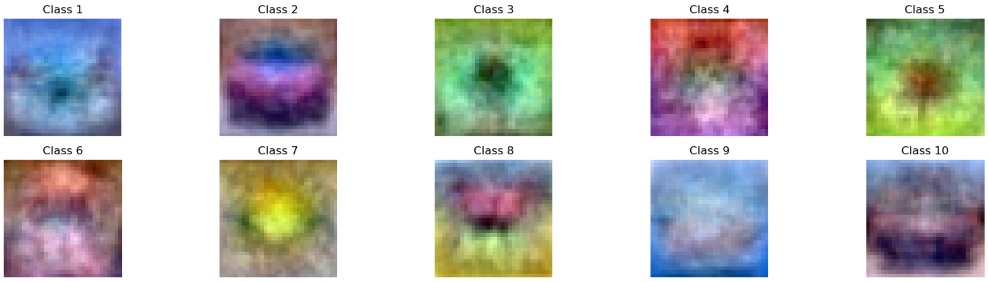

Figure 7: Learned weight matrix for eta 0.01 and l = 0.1

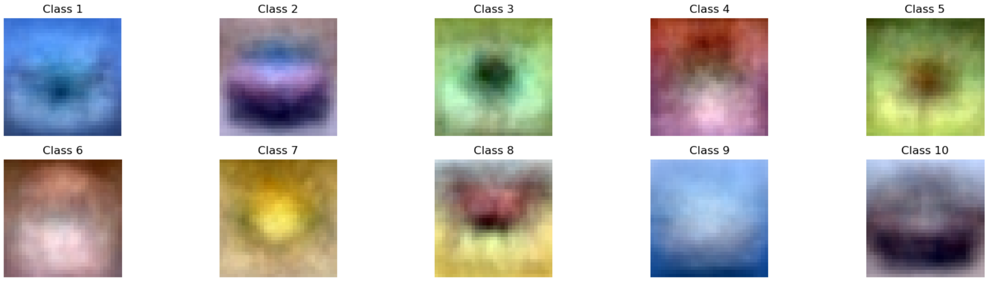

Figure 8: Learned weight matrix for eta 0.01 and l = 1