DD2424 Deep Learning Assignment 2

Ershad Taherifard

May 2020

1 Introduction

This paper trains and tests a two layer network with multiple outputs to classify images from the CIFAR-10 dataset. The dataset consists of 60000 32x32 color images in 10 classes, with 6000 images per class. They are divided into five training batches and one test batch, each with 10000 images.

2 Methods

As in assignment 1, mini-batch gradient descent will be used to a cost function that computes the cross-entropy loss of the classifier applied to the labelled training data and an L_2 regularization term on the weight matrix.

In this assignment, a cyclical learning rate is used which means the parameter η cycles between η_{max} and η_{min} as the model is learning. Furthermore, a coarse and a fine search is performed in order to find the most optimized value for λ.

3 Results

3.1 Gradients

The analytically calculated gradients were in fact correct. This is seen by calculating the relative error between the a numerically computed gradient value and an analytically computed gradient. The numerically computed gradient is based on the centered difference formula. As seen in the table below the relative difference is very low and thus we can conclude the gradients are correctly calculated.

$$\frac{|g_a - g_n|}{\max(\text{eps}, |g_a| + |g_n|)}$$

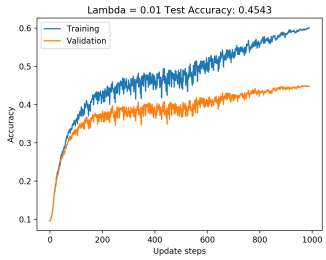
$i =$	1	2
$\sum gn_{W\,i} =$	$-2.8976820942716586\text{e}-14$	0.0
$\sum gn_{b\,i} =$	0.2530874338478766	$6.938893903907228\text{e}-18$
$\sum ga_{W\,i} =$	$1.13464793116691\text{e}-08$	$3.3306690738754696\text{e}-10$
$\sum ga_{b\,i} =$	0.25308743394347744	$-4.440892098500626\text{e}-11$
$\varepsilon_{b_i} =$	$7.071116152165262\text{e}-10$	$7.071116152165262\text{e}-10$
$\varepsilon_{W_i} =$	$6.922096175862951\text{e}-11$	$6.922096175862951\text{e}-10$

3.2 Plots

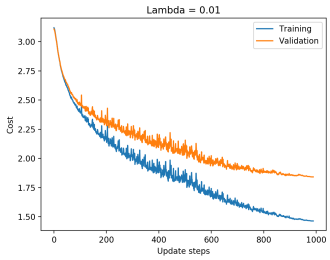
3.2.1 1 Cycle

We see that during the beginning of the cycle when the learning rate is small, the loss decreases rapidly but then it decreases at a slowing rate. The accuracy increases a lot in at start and then stabilises in the halfway through the cycle, when the learning rate is high. Then during the second half of the cycle the accuracy increases yet again, further suggesting that a lower learning rate improves learning.

(a) Accuracy



(b) Costs



(c) Loss

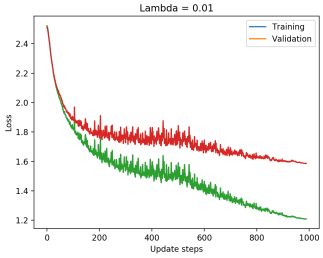
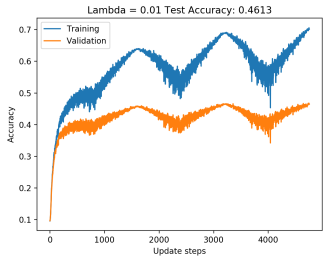


Figure 1: $\text{n}_{batch} = 100, n_s = 500, epochs = 10$

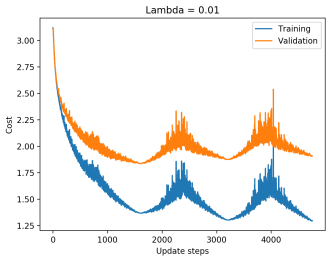
3.2.2 3 Cycles

In this case we see the same pattern as before but now it is repeating itself for each cycle. We see that the plots are shaped as waves due to the nature of cyclical learning rates. As the learning rate increase for within each cycle, the loss increases and the accuracy decreases. Once the learning rate decrease the loss decrease and the accuracy increases.

(a) Accuracy



(b) Costs



(c) Loss

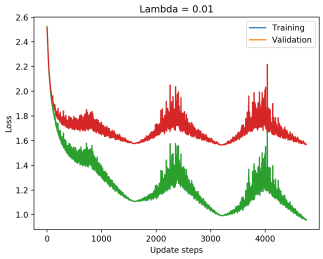


Figure 2: $\mathrm{n}_{batch} = 100, n_s = 800, epochs = 48$

3.3 Random Search

To find a suitable value for λ a coarse-to-fine approach is performed. First a search within a wide range of λ on a log scale is done. The values are extracted from a uniform distribution of exponents $\lambda_{min} = -5$ and $\lambda_{max} = -1$. Then the number ten is raised by these values.

Once then coarse search is done, a narrower search is done. This time $\lambda_{min} = -3$ and $\lambda_{max} = -2$

The final accuracy is a bit lower then expected. One possible reason is implementation issues. The theoretical value for the final accuracy should be higher than then once show so far in this assignment.

3.3.1 Coarse search (2 cycles)

Validation Accuracy: 0.415 Lambda:0.00013810168311317438
Validation Accuracy: 0.414 Lambda: 7.73677066392006e-05
Validation Accuracy: 0.406 Lambda: 0.07267109420026695
Validation Accuracy: 0.458 Lambda: 0.0003362254632232712
Validation Accuracy: 0.477 Lambda: 0.006551518774129204
Validation Accuracy: 0.440 Lambda: 5.171915544622242e-05
Validation Accuracy: 0.433 Lambda: 0.0005389838990043178
Validation Accuracy: 0.384 Lambda: 0.08597426886535235

3.3.2 Fine search (2 cycles)

Validation Accuracy: 0.426 Lambda: 0.0019277456992253224
Validation Accuracy: 0.421 Lambda: 0.0016677844907711113
Validation Accuracy: 0.466 Lambda: 0.009232948748803633
Validation Accuracy: 0.437 Lambda: 0.0024080064737038507
Validation Accuracy: 0.439 Lambda: 0.005059242510786357
Validation Accuracy: 0.415 Lambda: 0.001508039982543899
Validation Accuracy: 0.418 Lambda: 0.0027095299004427586
Validation Accuracy: 0.471 Lambda: 0.009629242473790354

3.3.3 Final

(a) Accuracy

(b) Loss

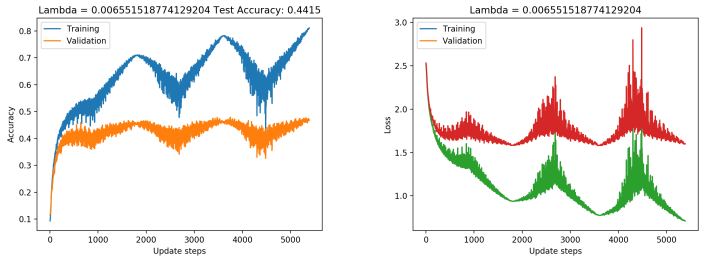


Figure 3: 3 cycles