Assignment 2
DD2424 Oscar Azrak

1. Introduction

This assignment aimed to develop, train and evaluate a k-layer neural network for image classification using the CIFAR-10 dataset. The network was trained using mini-batch gradient descent which was applied to a cost function that consisted of the cross-entropy loss and an L2 regularization term on the weight matrices. This assignment also implemented batch normalization. The network's performance was optimized by fine tuning the regularization parameter and utilizing a cyclical learning rate. This allowed me to achieve best test accuracy for this assignment.

2. Method

1. Download and extract the contents of CIFAR-10 dataset
2. Load the dataset and split the data to train, validation and test sets.
3. Normalize the relevant datasets.
4. Initialize model parameters
5. Implement relevant functions
6. Compute and compare the numerical and analytical gradients
7. Train the model using default settings
8. Find the best parameters
9. Evaluate the model on the test set

3. Results

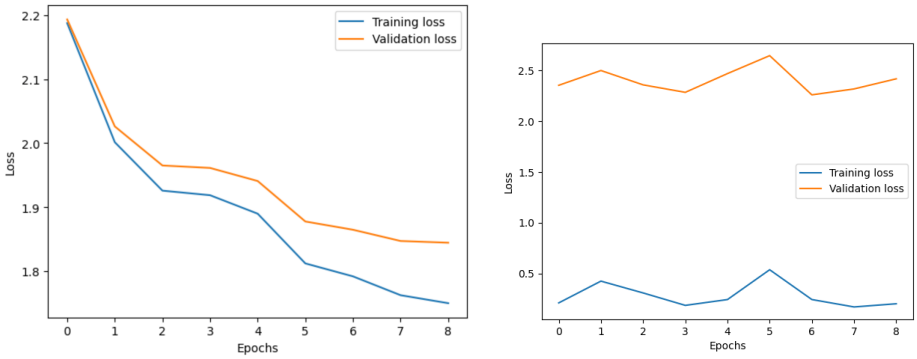
The central difference, h of 10^{-5}, was used where these errors were observed while using Batch Normalization:

Layer	Relative error Weights	Relative error biases
1	0.00013	$9.37*10^{-19}$
2	$4.81*10^{-11}$	$3.05*10^{-11}$

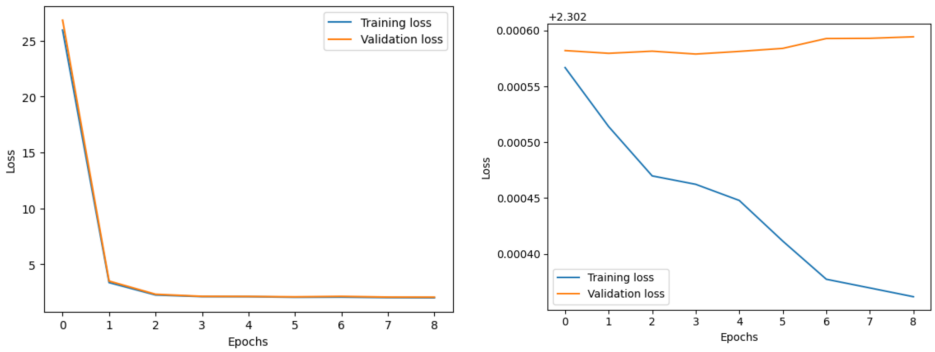
Layer	Relative error Weights	Relative error biases
1	$3.56*10^{-05}$	$8.76*10^{-19}$
2	0.00341	$1.28*10^{-17}$
3	$5.44*10^{-11}$	$5.45*10^{-11}$

The relative errors were obtained by comparing the difference of the analytical and numerical gradients for each variable. The fact that the errors were small indicates high accuracy of the gradient calculations.

Graph of loss evolution with and without batch normalization for 3-layer training:



Loss evolution for 9 layers with and without batch normalization:



The range for retrieving best lambdas for the coarse search with batch normalization used values generated on a log space in the range of 10^{-5} to 10^{-1}.

Lambda	Accuracy
0.00697	52.77%

This accuracy is from fine searching after the coarse search. The coarse search managed to achieve a higher accuracy with lambda: 0.00774, with an accuracy of: 53.51%

Sensitivity to initialization: