Oscar Azrak
000912-1690
DD2424

Oscar Azrak
Oazrak@kth.se
DD2424
Assignment 1

Oscar Azrak
000912-1690
DD2424

## Abstract

In this report, a linear classifier for the CIFAR-10 image dataset has been implemented using mini-batch gradient descent. Different parameter settings for regularization (lambda), learning rate (eta), and batch size (n_batch) have been tested. After training, the test accuracy and visualize the learned weight matrix as images have been reported. The effects of regularization and learning rate on the classifier's performance is also discussed.

## Introduction

The CIFAR-10 dataset contains 60,000 32x32 color images in 10 classes, with 6,000 images per class. In this experiment, we have implemented a linear classifier to categorize the images into their respective classes. Mini-batch gradient descent was used to train the classifier and the effects of different parameter settings on the classifier's performance was also analyzed.

## Methodology

Before training the linear classifier, it is essential to ensure that the gradients are correctly computed. In this assignment, gradient checking was performed using the central difference approximation method. The central difference approximation method calculates the numerical gradient by comparing the loss function values at two points near the parameter value.
For a small subset of the data, the following results were obtained:
Relative error W: 9.029321809444876e-07
Relative error b: 9.843060966979009e-10

The linear classifier using mini-batch gradient descent was trained with the following parameter settings:

a. lambda=0, n_epochs=40, n_batch=100, eta=.1
b. lambda=0, n_epochs=40, n_batch=100, eta=.001
c. lambda=.1, n_epochs=40, n_batch=100, eta=.001
d. lambda=1, n_epochs=40, n_batch=100, eta=.001

For each setting, the loss and cost function was plotted on the training data and validation data after each epoch of the mini-batch gradient descent algorithm. The learned weight matrix as images after the completion of training is also visualized for each parameter setting.

Oscar Azrak
000912-1690
DD2424
Results and Discussion

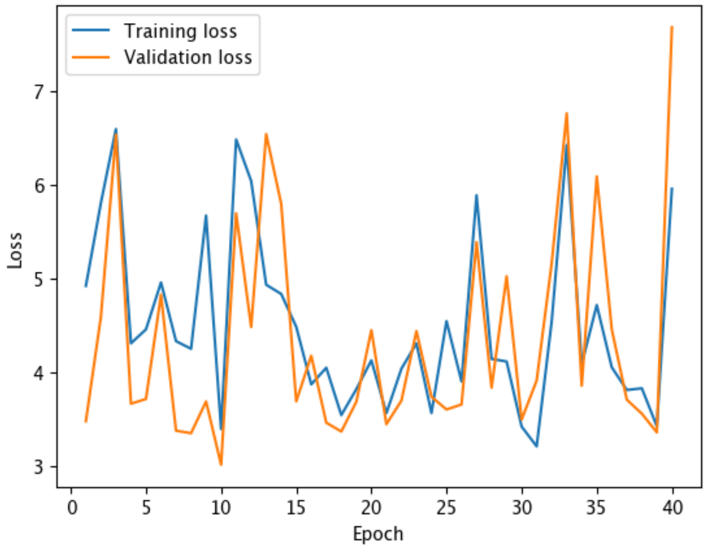


Figure 1. Graph of the loss and cost function for parameters: lambda=0, n epochs=40, n batch=100, eta=.1

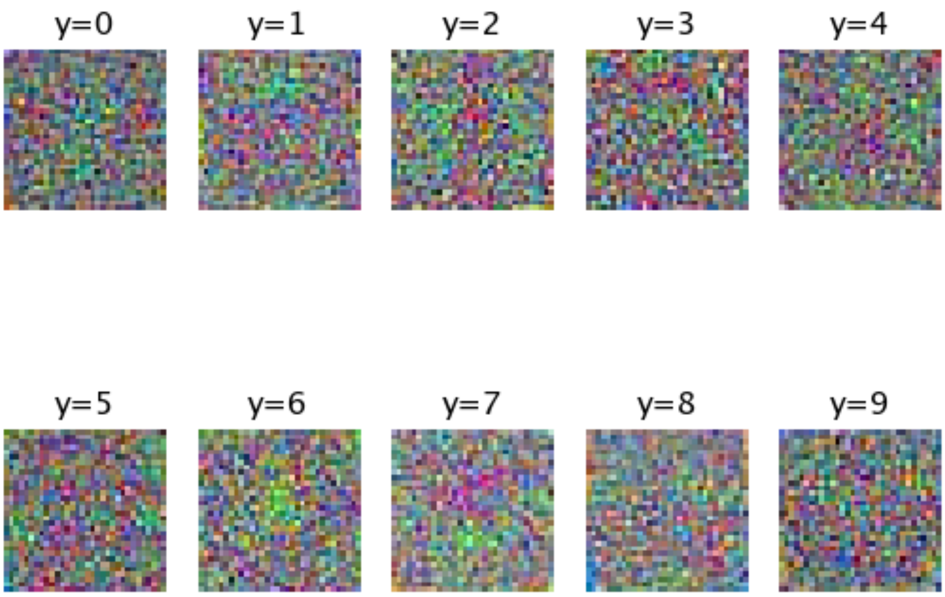


Image 1. Images of weight matrix using parameters: lambda=0, n epochs=40, n batch=100, eta=.1

Oscar Azrak
000912-1690
DD2424

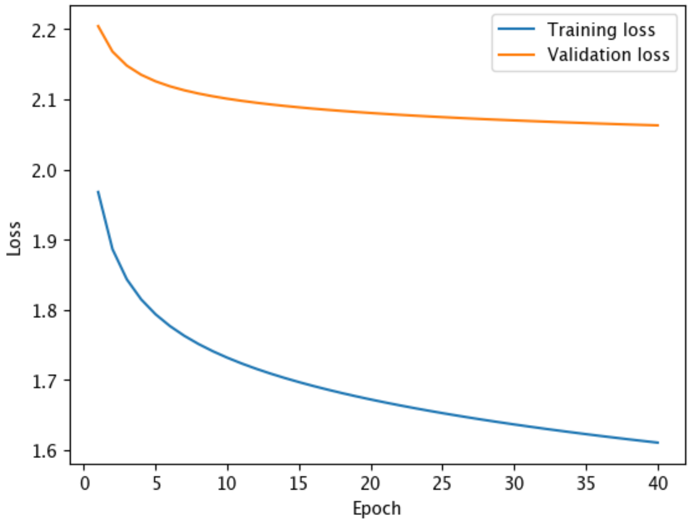


Figure 2. Graph of the loss and cost function for parameters: lambda=0, n epochs=40, n batch=100, eta=.001

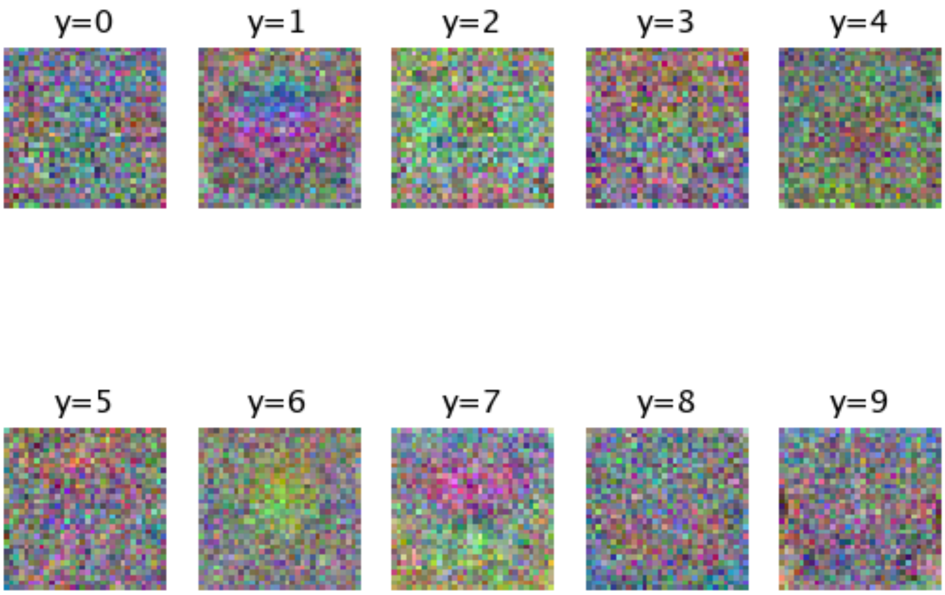


Image 2. Images of weight matrix using parameters: lambda=0, n epochs=40, n batch=100, eta=.001

Oscar Azrak
000912-1690
DD2424

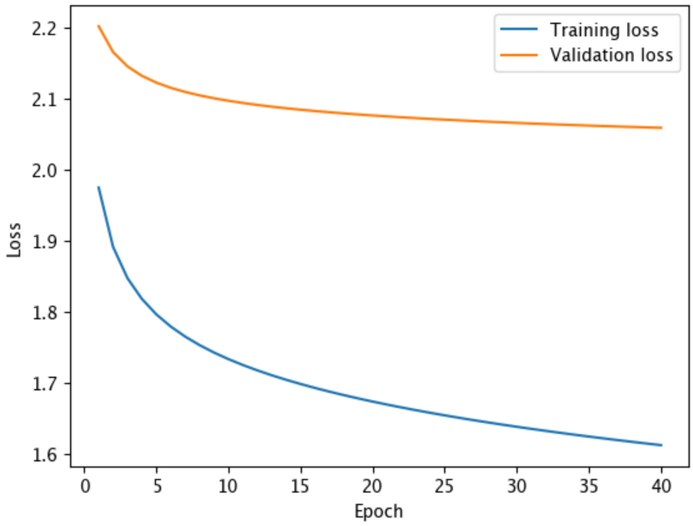


Figure 3. Graph of the loss and cost function for parameters: lambda=.1, n epochs=40, n batch=100, eta=.001

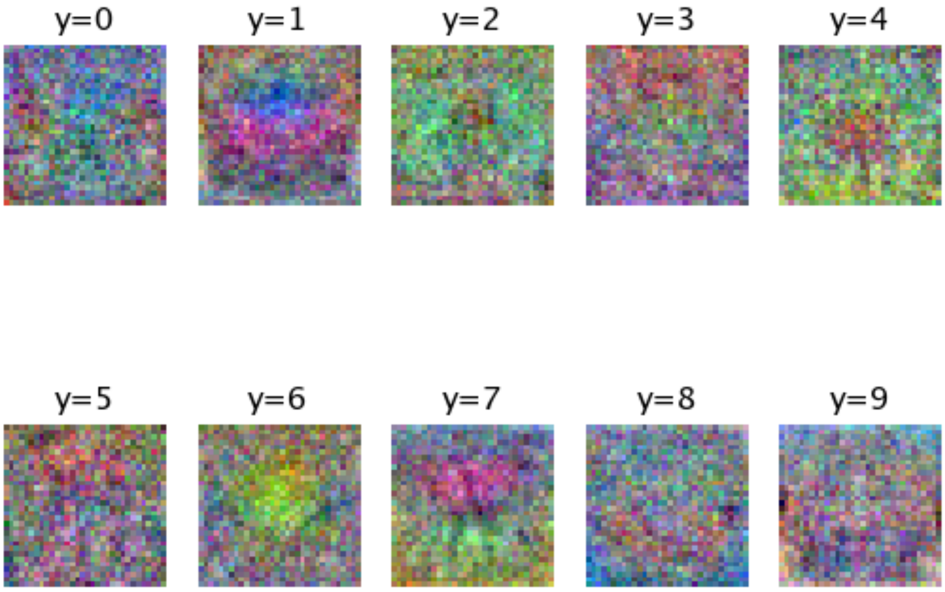


Image 3. Images of weight matrix using parameters: lambda=.1, n epochs=40, n batch=100, eta=.001

Oscar Azrak
000912-1690
DD2424

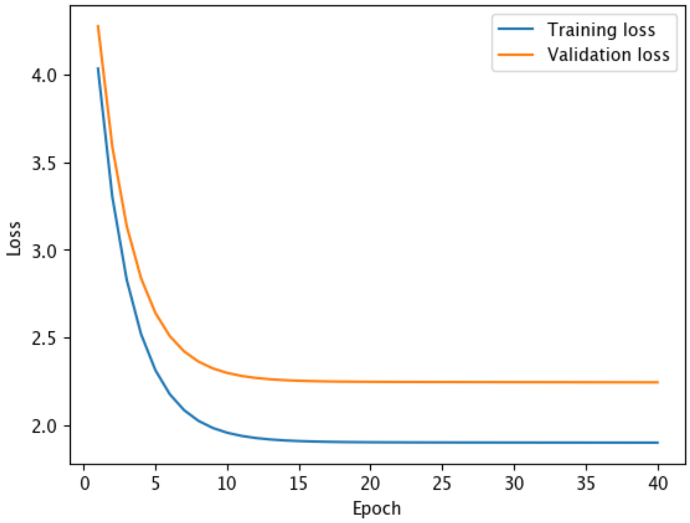


Figure 4. Graph of loss and cost function for parameters: lambda=1, n epochs=40, n batch=100, eta=.001

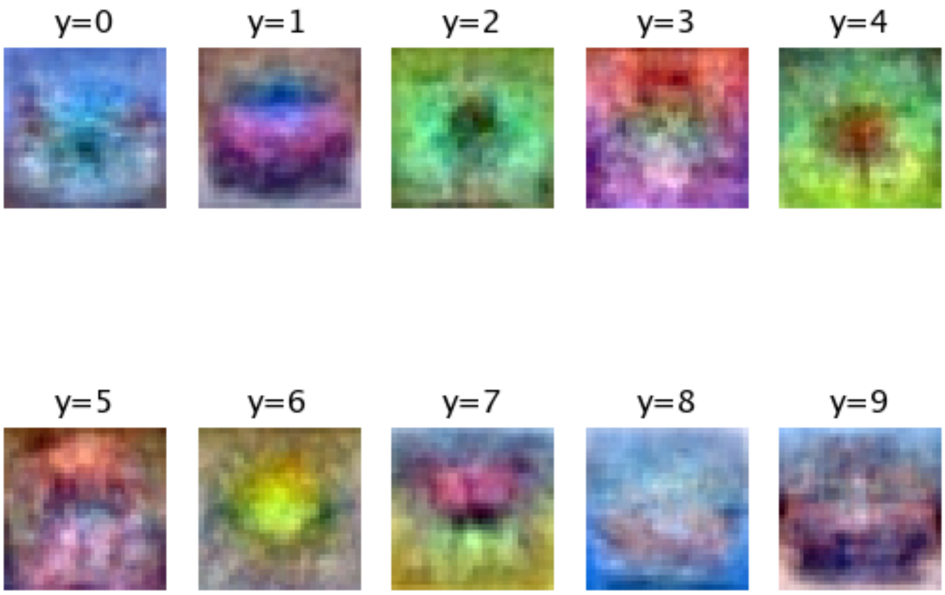


Image 4. Images of weight matrix using parameters: lambda=1, n epochs=40, n batch=100, eta=.001

The test accuracies for the different parameter settings are:
a. Test accuracy (lambda=0, eta=.1): 25.36%

Oscar Azrak
000912-1690
DD2424
b. Test accuracy (lambda=0, eta=.001): 38.80%
c. Test accuracy (lambda=.1, eta=.001): 39.04%
d. Test accuracy (lambda=1, eta=.001): 37.45%

By analyzing the results, we can observe the following:

1. The importance of the correct learning rate: A higher learning rate (eta=.1) leads to faster convergence, while a lower learning rate (eta=.001) results in slower convergence. However, a learning rate that is too high may cause the model to overshoot the optimal weights and result in a lower test accuracy.

2. The effect of regularization: Increasing the amount of regularization (lambda) helps prevent overfitting and can lead to better generalization. However, too much regularization can result in underfitting, as the model becomes too constrained.

3. Conclusion: T implementation of a linear classifier on the CIFAR-10 dataset have been demonstrated using mini-batch gradient descent. The effects of different parameter settings on the classifier's performance have been analyzed and the importance of selecting the correct learning rate and regularization is highlighted. The best test accuracy was achieved with the setting (lambda=.1, eta=.001).