DD2424 Deep Learning in Data Science

Assignment 1 - 24th June, 2019

Musab Tayyib Gelisgen - gelisgen@kth.se

In this assignment, a single layer neural network is trained with CIFAR-10 dataset by using mini batch gradient descent and the obtained results are provided in this report. I successfully managed to write the functions to correctly compute the gradient analytically. The tests done are provided below.

1. Tests

- **lambda=0, n_epochs=40, n_batch=100, eta=.1**

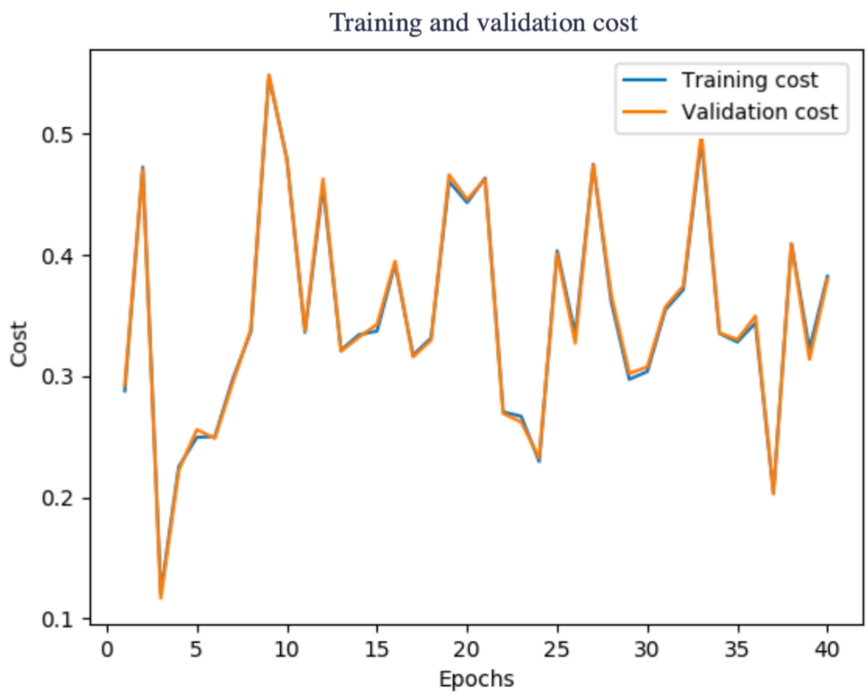

Figure I: Value of cost over epochs

Train Accuracy	Test Accuracy
%25	%24

Table I: Train and test accuracy when lambda=0 & learning rate=0.1

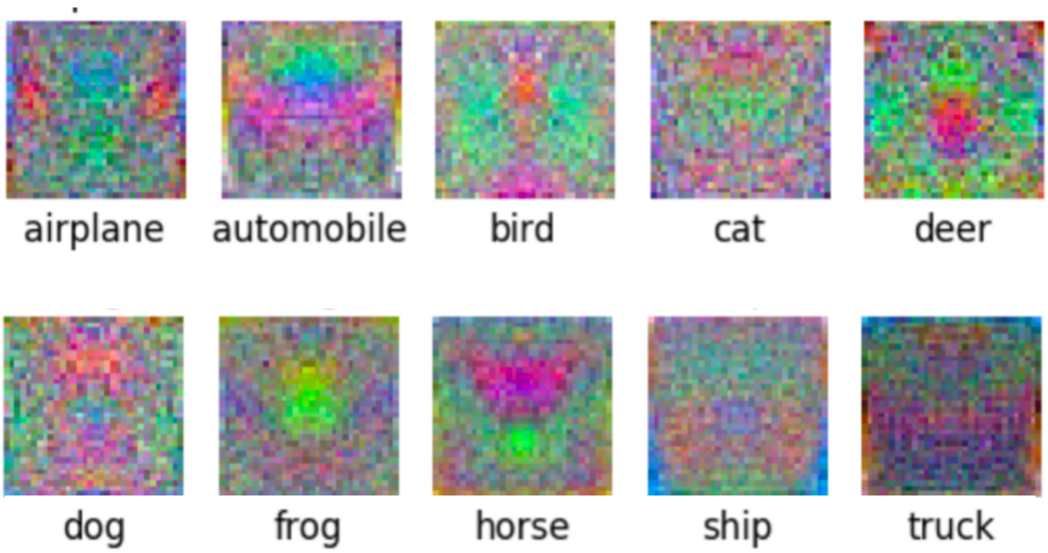

Figure II: Weight matrices of different classes

- **lambda=0, n_epochs=40, n_batch=100, eta=.01**

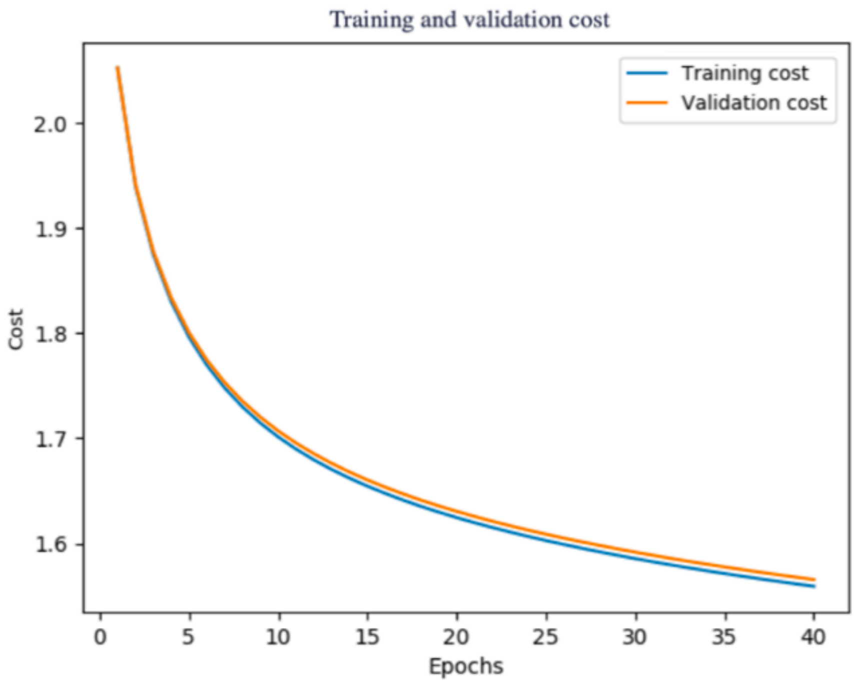

Figure III: Value of cost over epochs

Train Accuracy	Test Accuracy
%42	%38

Table II: Train and test accuracy when lambda=0 & learning rate=0.01

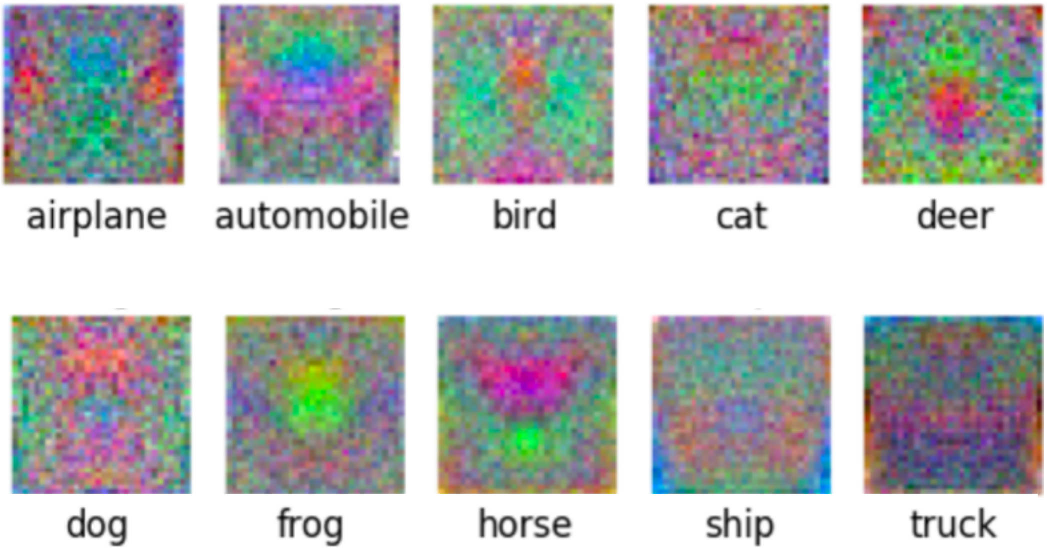

Figure IV: Weight matrices of different classes

- **lambda=.1, n_epochs=40, n_batch=100, eta=.01**

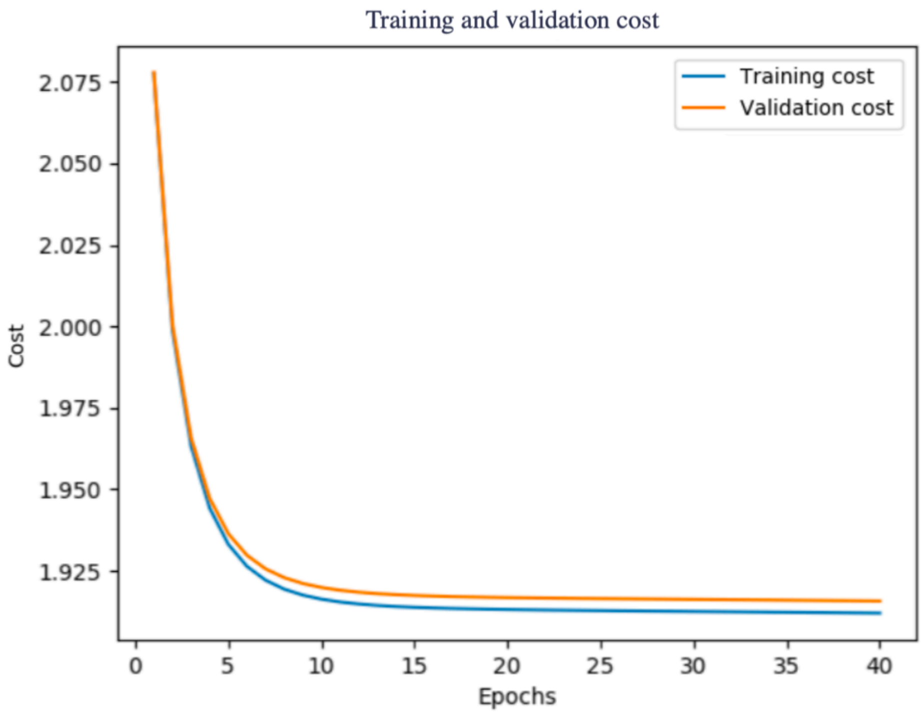

Figure V: Value of cost over epochs

Train Accuracy	Test Accuracy
%36	%35

Table III: Train and test accuracy when lambda=0.1 & learning rate=0.01

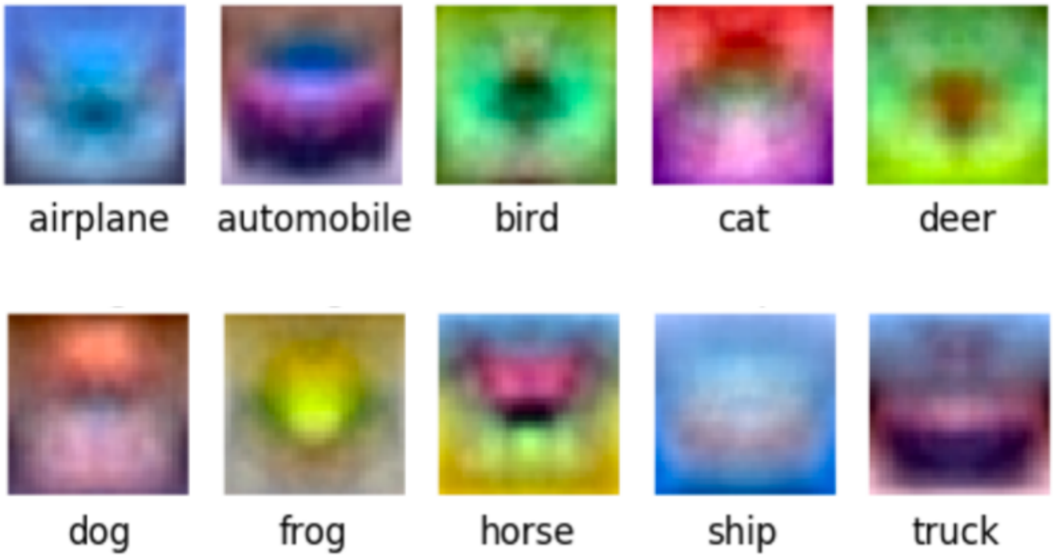

Figure VI: Weight matrices of different classes

- **lambda=1, n_epochs=40, n_batch=100, eta=.01**

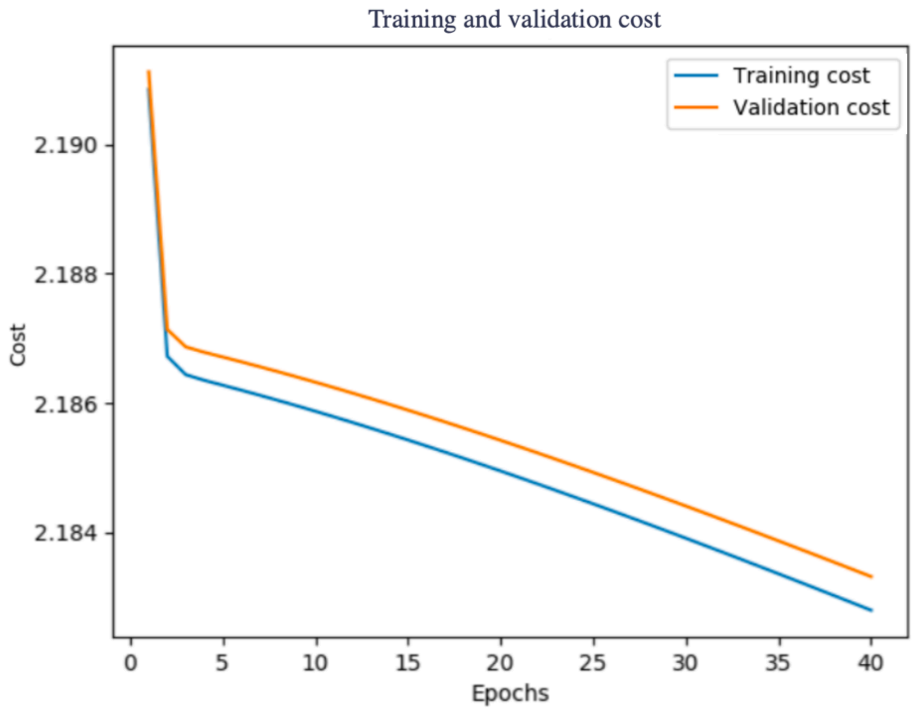

Figure VII: Value of cost over epochs

Train Accuracy	Test Accuracy
%24	%24

Table IV: Train and test accuracy when lambda=1 & learning rate=0.01

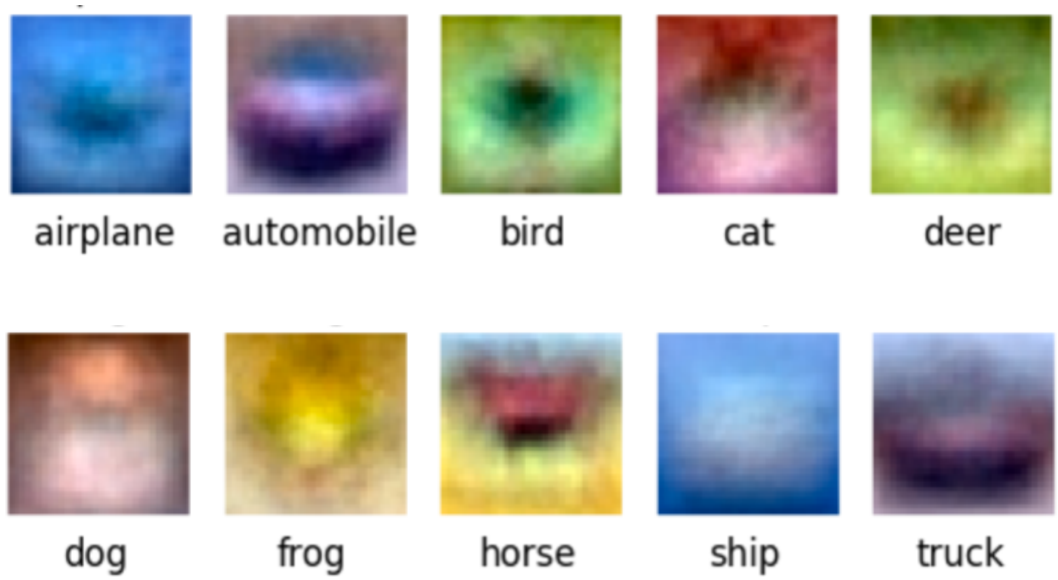

Figure VIII: Weight matrices of different classes

2. Conclusion

In general, we can say the regularization reduces the influence of outlier data points. Thus, it helps when overcoming overfitting issues (and it reduces the variance of the model). Even though the better accuracies are obtained with lower lambda (such as experiment no.2), but the data is probably overfit. As the lambda increases, the model tends to get better, but the reduction in test data accuracy points the model is not the best (such as experiments no.3 and no.4).

When a higher learning rate is used, the steps taken are bigger at each step. This causes the program the reach local minima (or possibly global minima) faster. However, if this rate is too high, it may cause to fluctuate a lot when finding the minimal point (such as experiment no.1). Then it results in a less stable convergence and this can be seen in the first experiment. With this, the model can have higher loss in an epoch than the previous epoch and thus, it can take a lot more time to learn. With smaller learning rates, the learning will be slower with a more stable convergence. However, if this value is kept too small, it may lead the model to stuck at a local minima. The selection of learning rate is critical when training the model since we ideally want to move fast to our goal while also showing progress at each step.