DD2424 Assignment 1 Report

Student: Alberto Xamin xamin@kth.se

Exercise 1

Analitical gradients

	Numerical	Analytical
time [s]	5.5991	0.0004

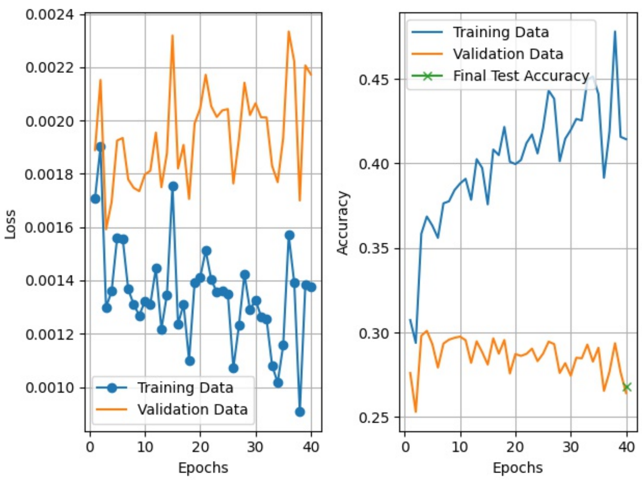
Absolute difference between mean weights `5.464830700232145e-18`

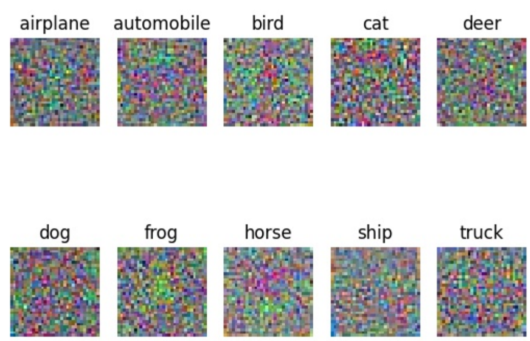
Absolute difference between mean biases `1.0842021724855044e-14`

We can observe that the difference is so small, that we can consider it to be zero.

Training runs

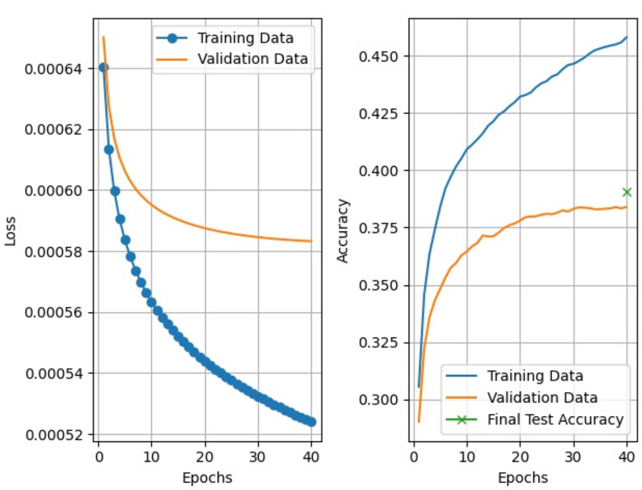
lambda=0, n epochs=40, n batch=100, eta=.1

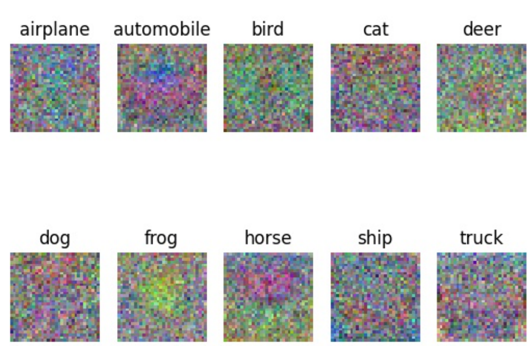




---

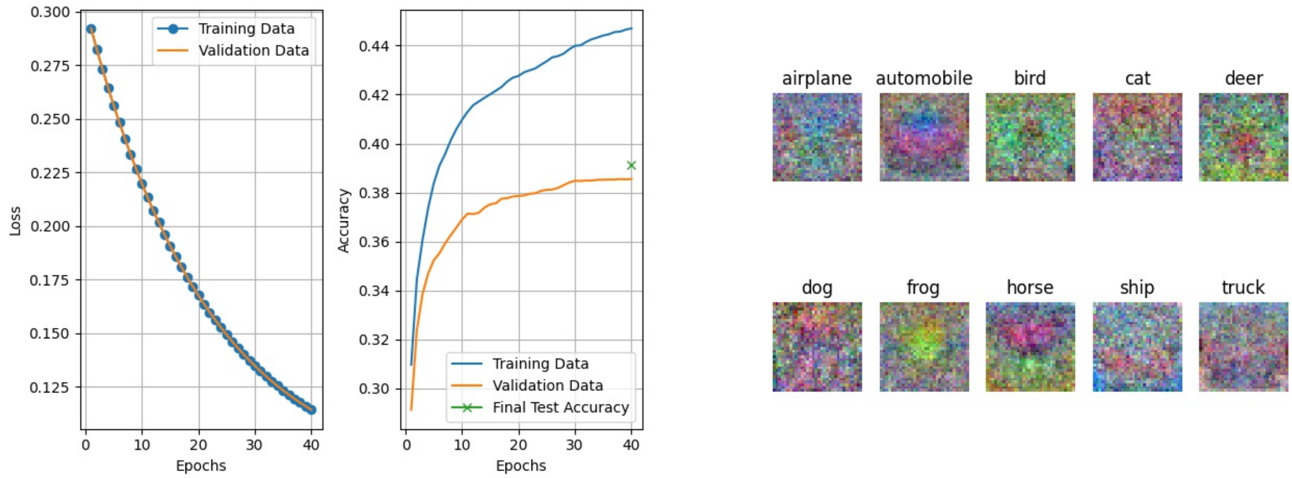
lambda=0, n epochs=40, n batch=100, eta=.001





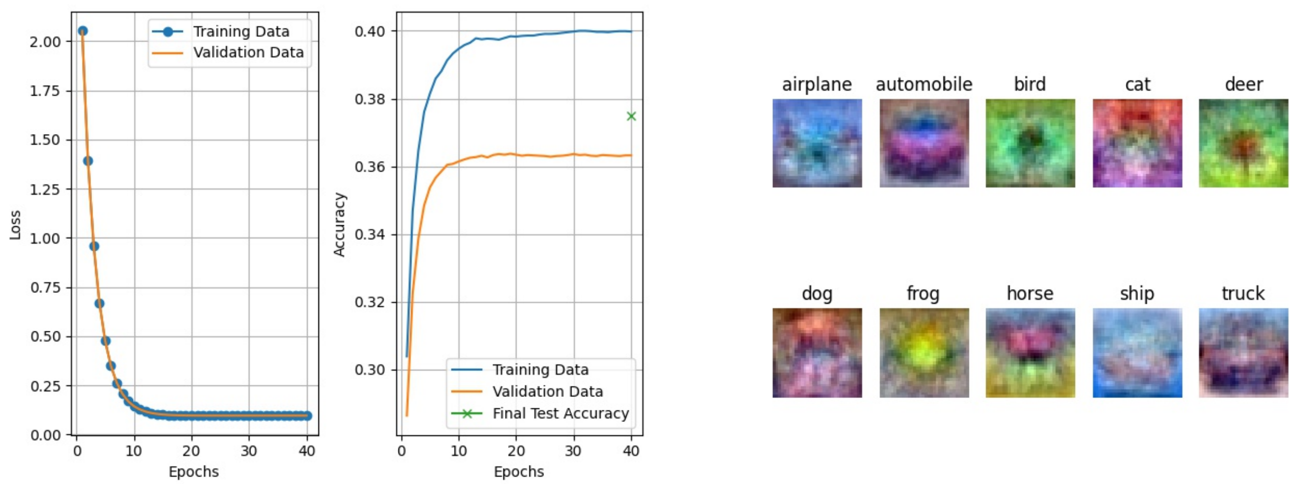
---

lambda=.1, n epochs=40, n batch=100, eta=.001



---

lambda=1, n epochs=40, n batch=100, eta=.001



Comments on regularization and learning rates

Learning rate (eta) To investigate the effect of the learning rate (eta) we have to look at the first 2 runs, as they only differ with the learning rate. It appears that a learning rate of 0.1 is too high and that may cause the algorithm to overshoot and undershoot the minimum as the epochs increase. We can observe that with `eta=0.001` the learning is smoother, so we can assume that this value works better for this data.

Regularization (lambda) Three different values for lambda were tested $(0, 0.1, 1)$. This regularization should help reduce overfitting of the network by reducing the variance and increasing the bias. We can observe that with `lambda=`0.1 the accuracy in the training data decreases but it increases on the validation and test set. We can also see that when we set the lambda to the maximum value of 1 it has a negative impact on the accuracy as expected with a high bias.