Foundations of Deep Learning - HW1

Matan Ben-Tov 316048321, Ofir Gaash - 315395210

April 18, 2023

1 Setup and Baseline

We sampled both from the training-set and test-set of CIFAR10 10% of the samples, resulting in 5K, 1K samples respectively. We trained and evaluated SVM classifier on this data and the results are as followed -

SVM's Kernel Type	Train Accuracy	Test Accuracy
Linear	98.94%	33.10%
RBF	72.30%	44.90%

2 Feed Forward Neural Network

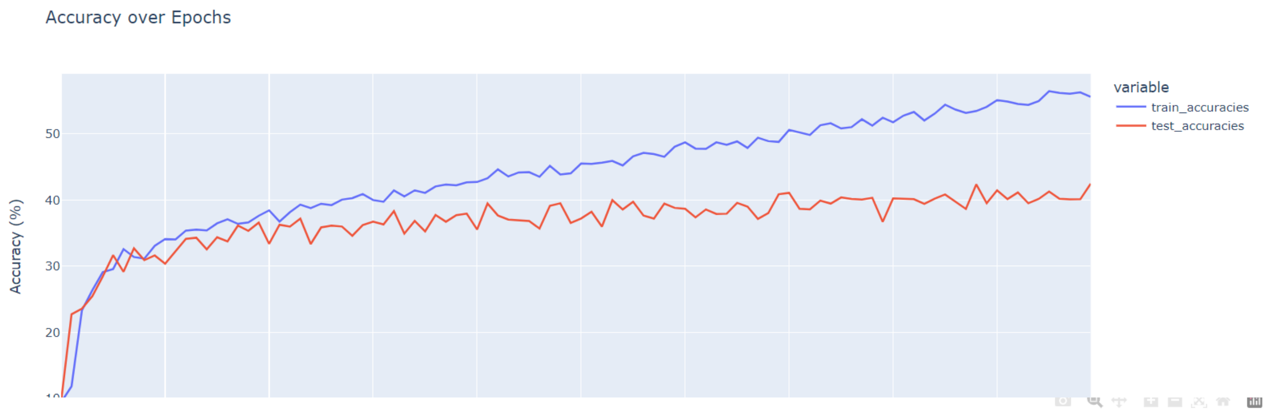
We trained a FFNN on the aforementioned data, found its best hyperparameters, and then modified/added certain components to better understand their impact.

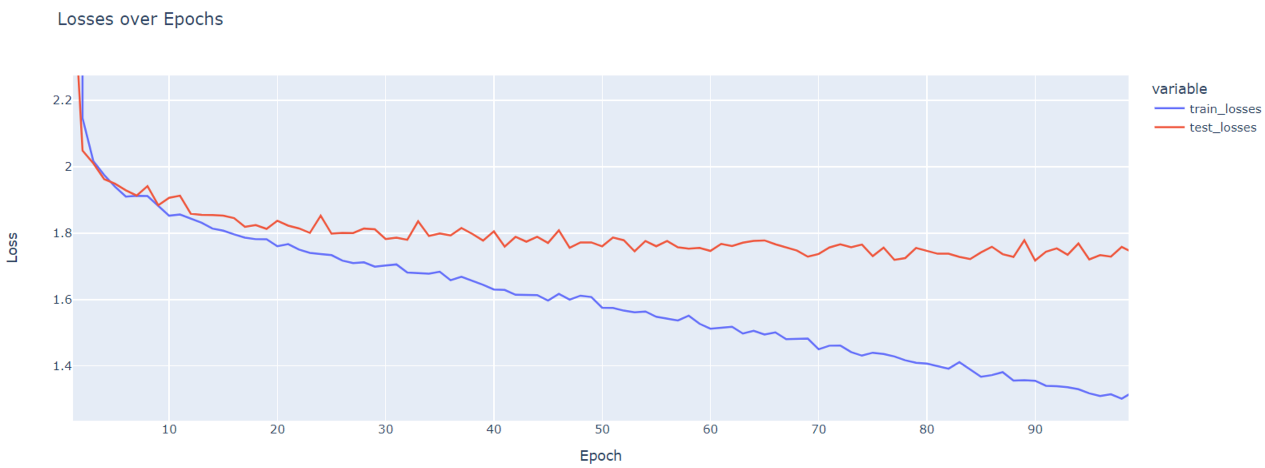
2.1 Baseline

Firstly we implemented a baseline FFNN, and tuned the following hyperparatmeters - momentum, learning rate and initialization STD. We performed a grid search (of which results are documented in q2-baseline-grid-search.csv) and found the following values to perform the best on the test set -

$$momentum = 0.9, learningRate = 0.001, initSTD = 0.1$$

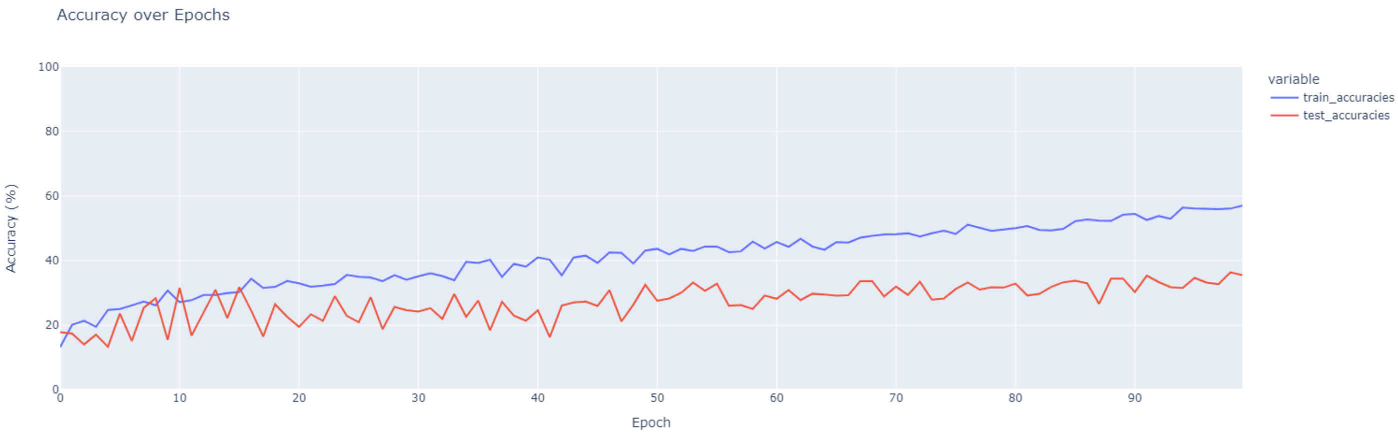
We also observed that this baseline converges to **test-set accuracy of 42.46% and loss of 1.75** within **<70 epochs**. The accuracy and loss curve of this optimal baseline are -

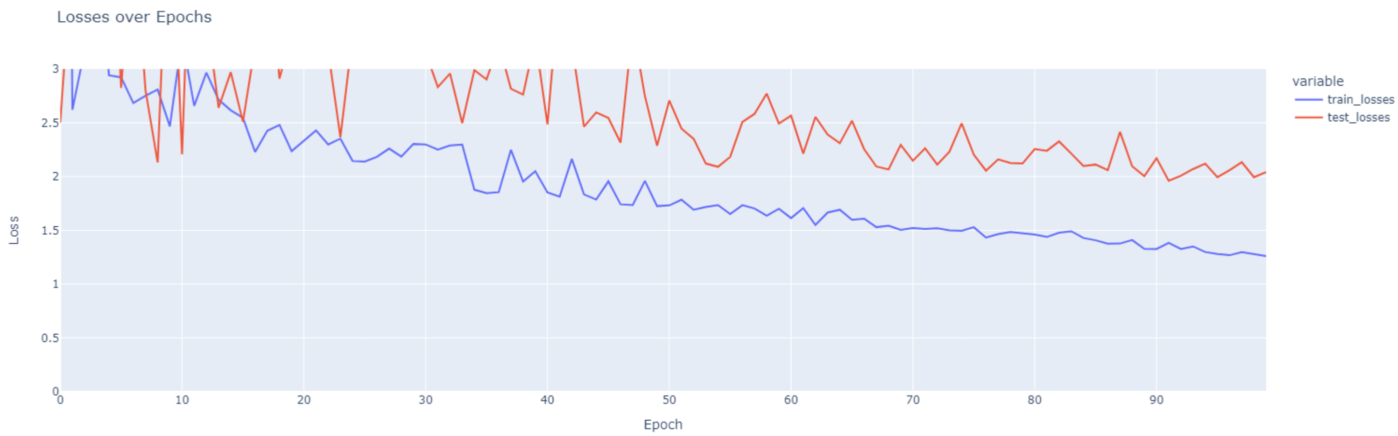




2.2 Optimization

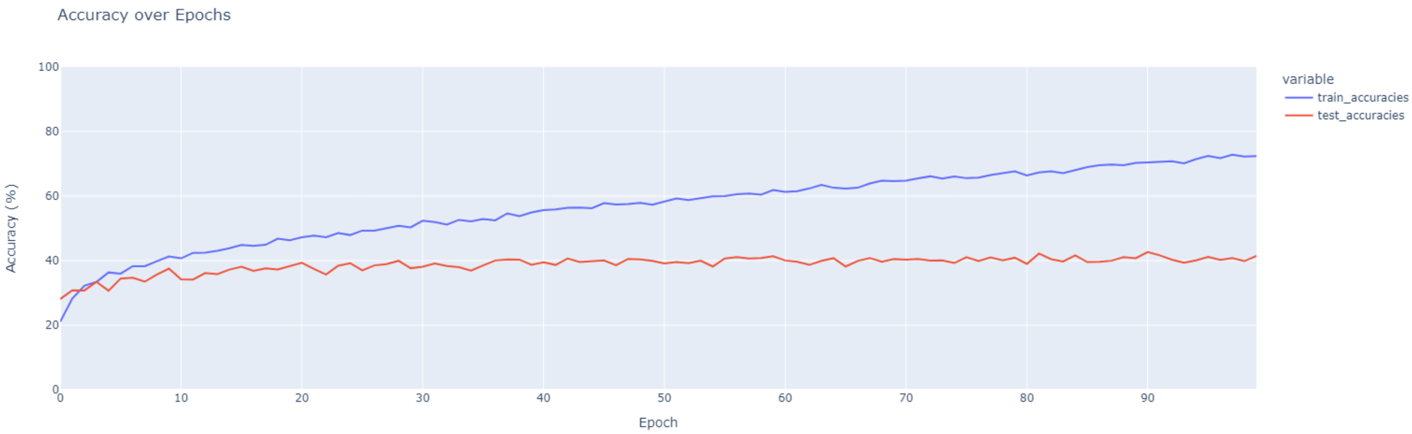
Next we test the Adam optimizer on our data. It achieves worse accuracy and loss, both on the training data and the test data. The process converges to **test-set accuracy of 36% and loss of 2.14.**

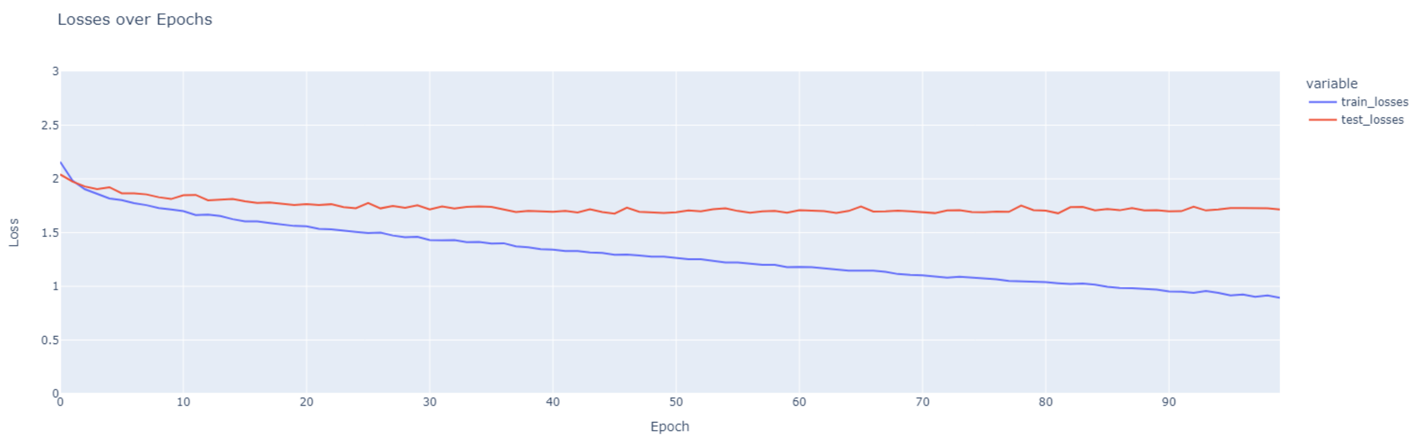




2.3 Initialization

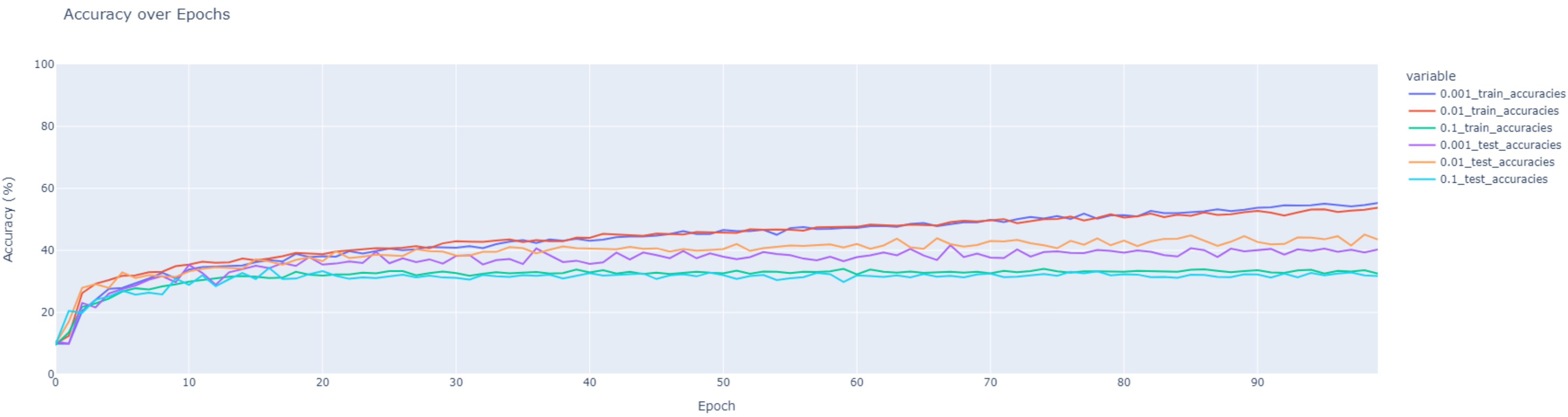
We move on to use Xavier initialization. While the test-loss and test-loss remain similar, the convergence rate seems faster. The process converges to **test-set accuracy of 41.5% and loss of 1.72.**

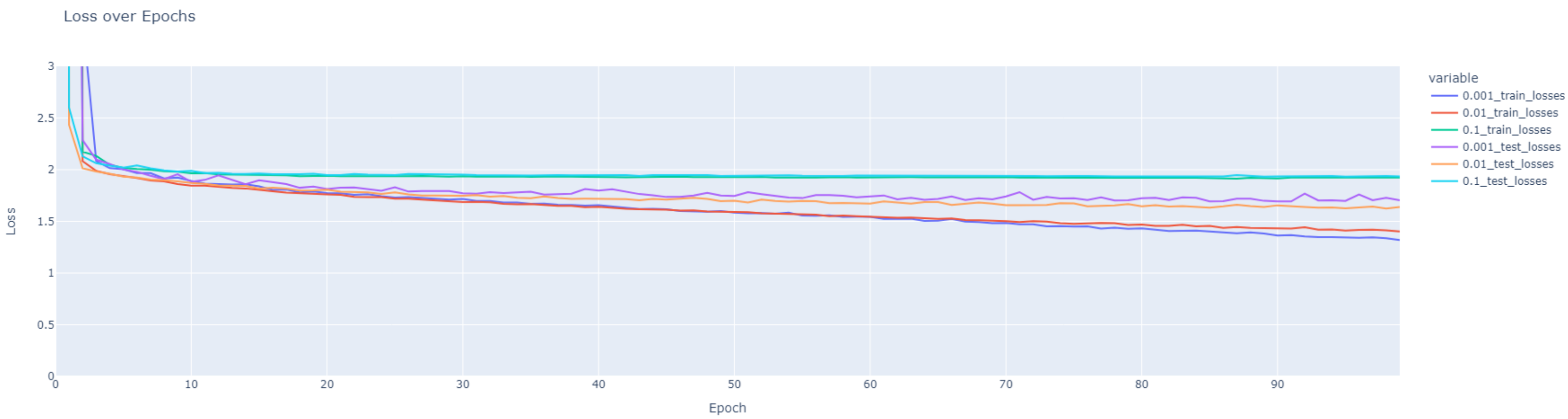




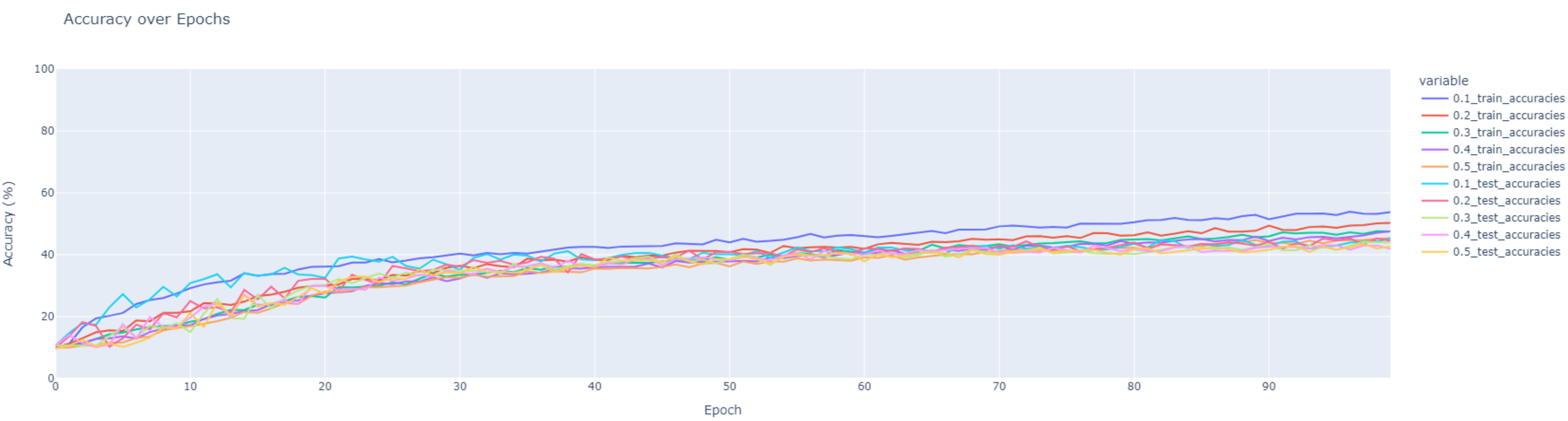
2.4 Regularization

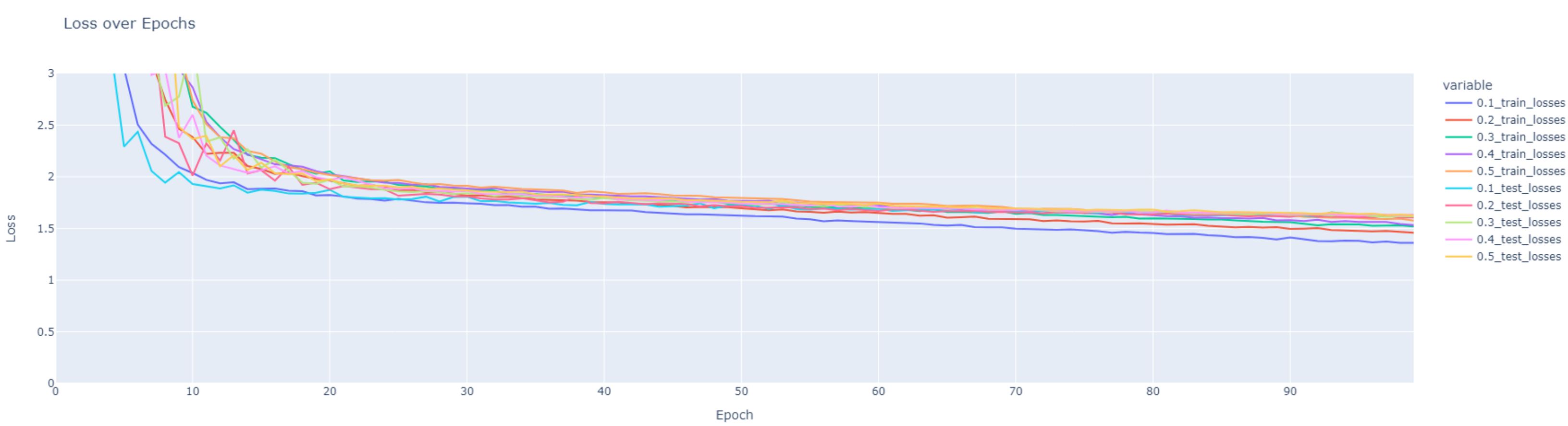
Experiments with weight decay showed lower performance for decay 0.1 and similar performance for decay 0.01, 0.001. It is noteworthy that for decay 0.1, the train accuracy and test accuracy were very similar. The optimal decay value achieved **test-set accuracy of 43% and loss of 1.63**.





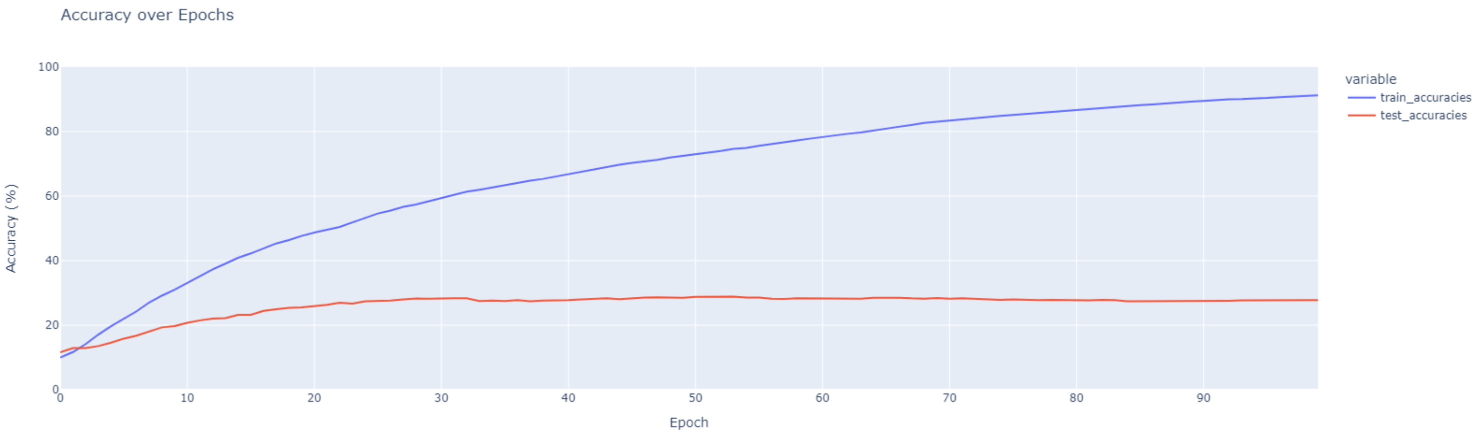
Experiments with dropout showed mildly better losses. The difference between dropout ratios of 0.1-0.5 was in the loss convergence rate. The optimal dropout value achieved **test-set accuracy of 45% and loss of 1.6**.

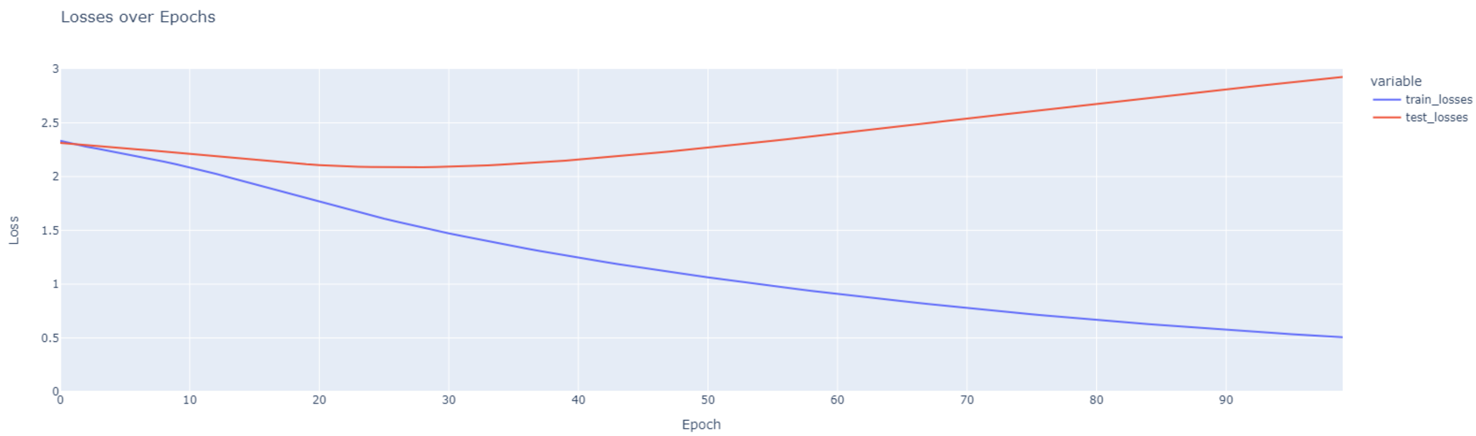




2.5 Preprocessing

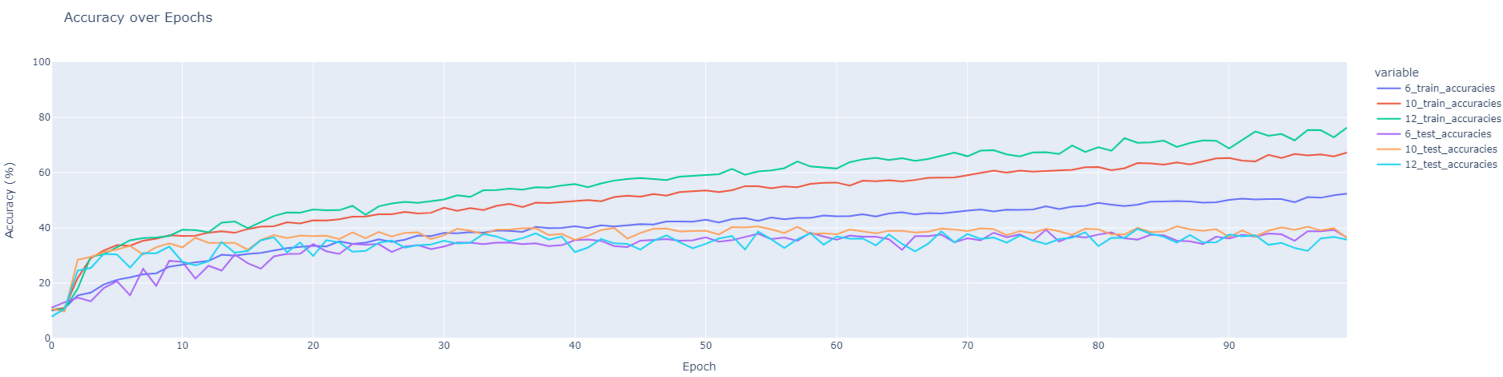
Performing PCA on the data (500 components) seems to increase the training accuracy & loss significantly, yet degrade the test accuracy & loss. We also note that the process seems more stable. The process converges to **test-set accuracy 28.9% and loss of 2.92**

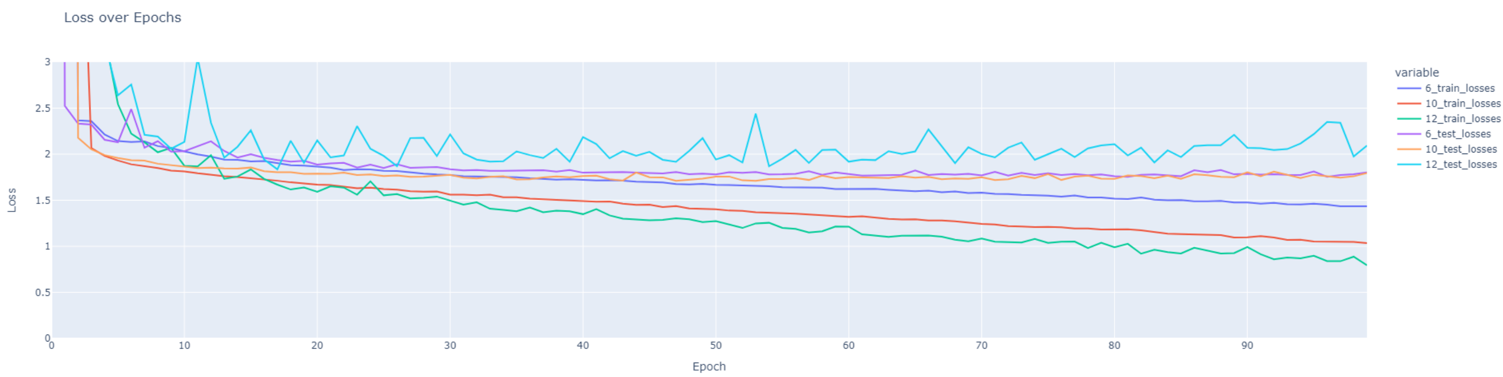




2.6 Network Width

We compare the performance of networks with width $2^6, 2^{10}, 2^{12}$. While the training improves with width, the test performance remains the same. Their **test-set accuracy is around 37% and their test-loss is around 1.8.**

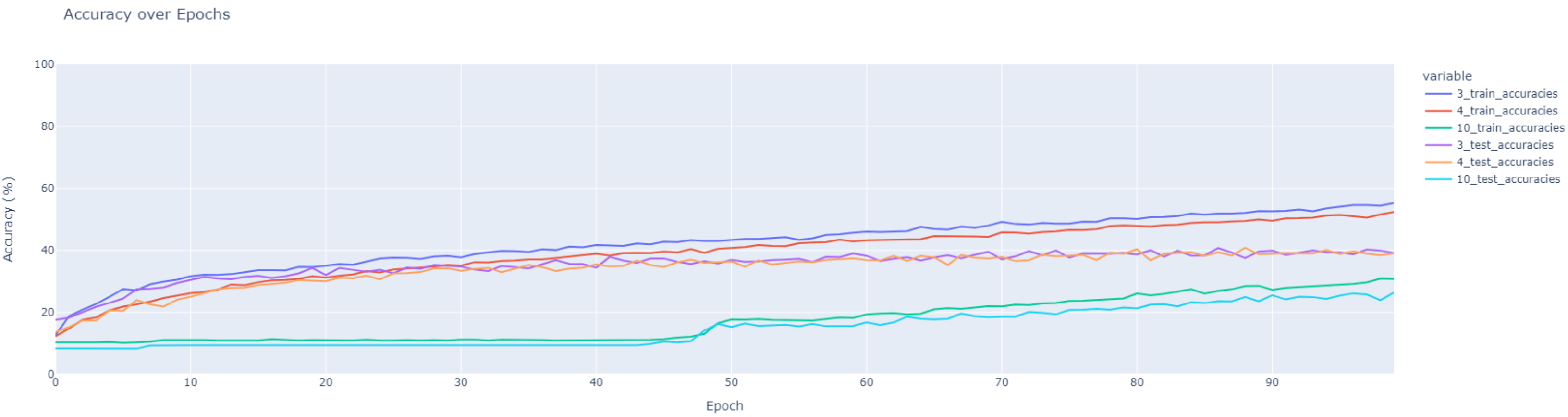


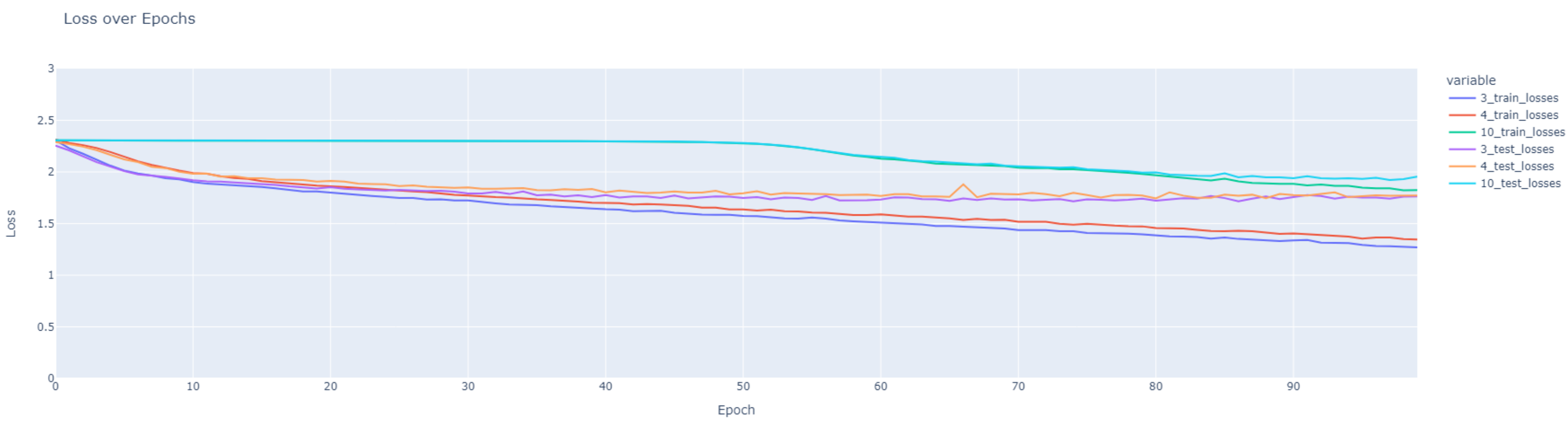


2.7 Network Depth

We compare the performance of networks with depth 3,4,10. The first two result in similar metrics. The network with depth 10 does not perform as well, which might be due to the vanishing gradients problem (as we are using gaussian initialization). The test-set accuracy and loss are presented in the following table.

Depth	Test Accuracy	Test Loss
3	40%	1.69
4	40%	1.73
10	33%	1.84





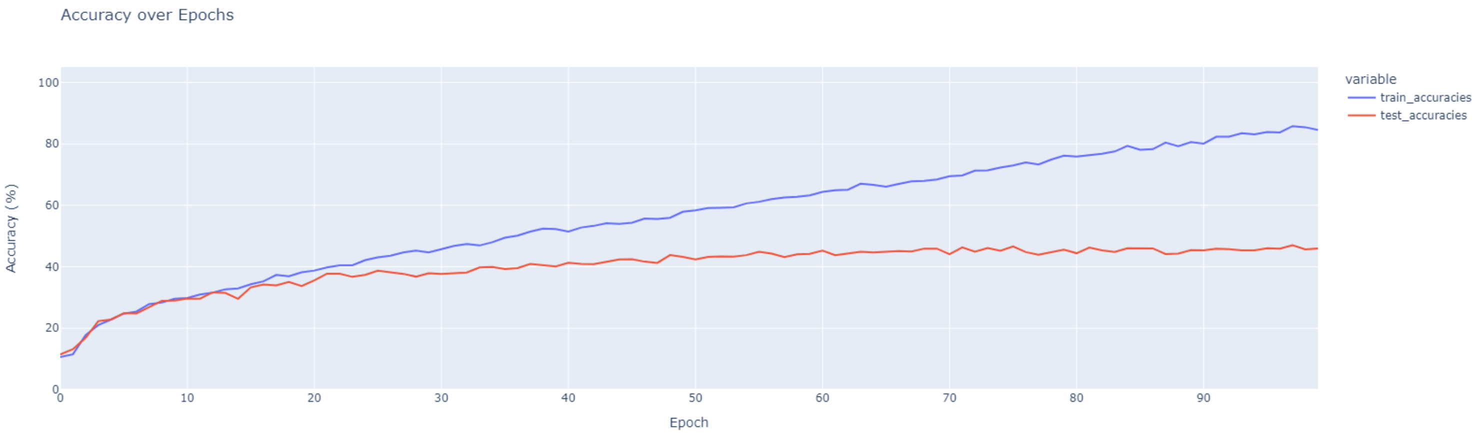
3 Convolutional Neural Network

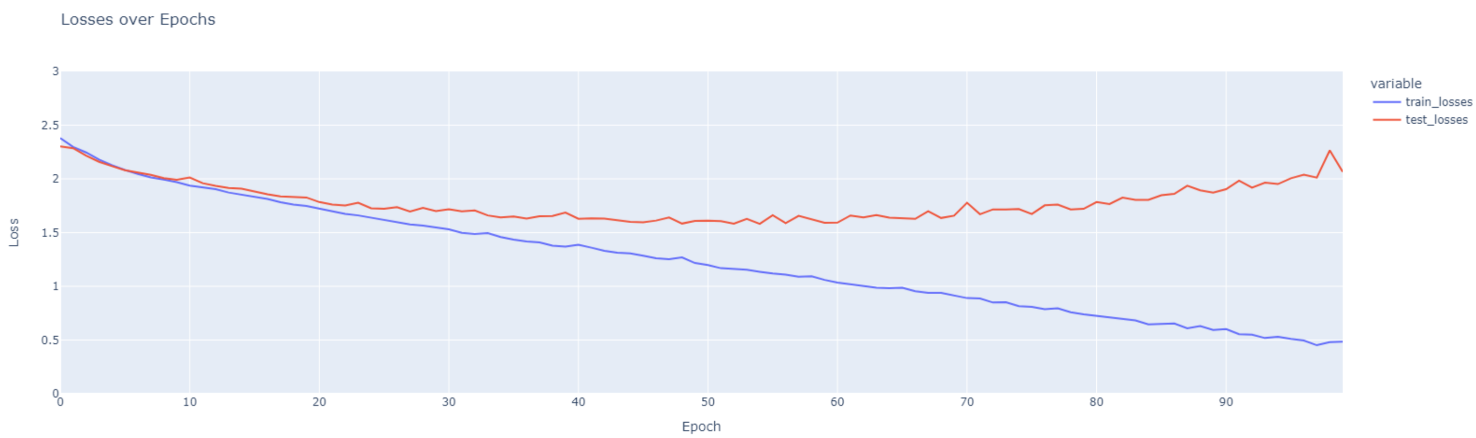
3.1 Baseline

Firstly we implemented a baseline CNN, and tuned the following hyperparatmeters - momentum, learning rate and initialization STD. We performed a grid search (of which results are documented in Q3-BASELINE-GRID-SEARCH.CSV) and found the following values to perform the best on the test set -

$$momentum = 0.9, learningRate = 0.001, initSTD = 0.1$$

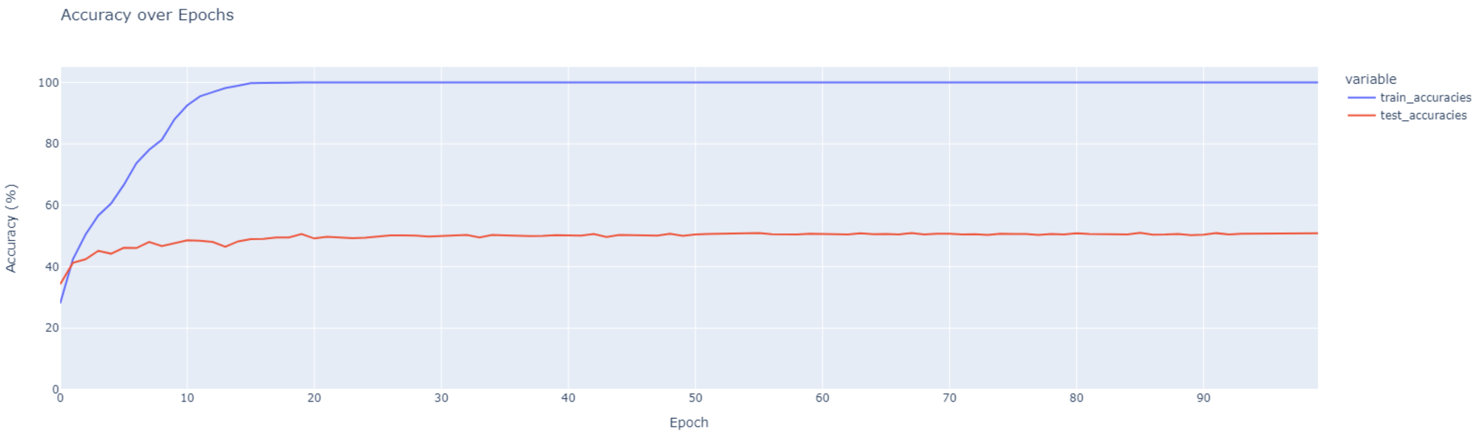
We also observed that this baseline converges to **test-set accuracy of 45% and loss of 2.6**.

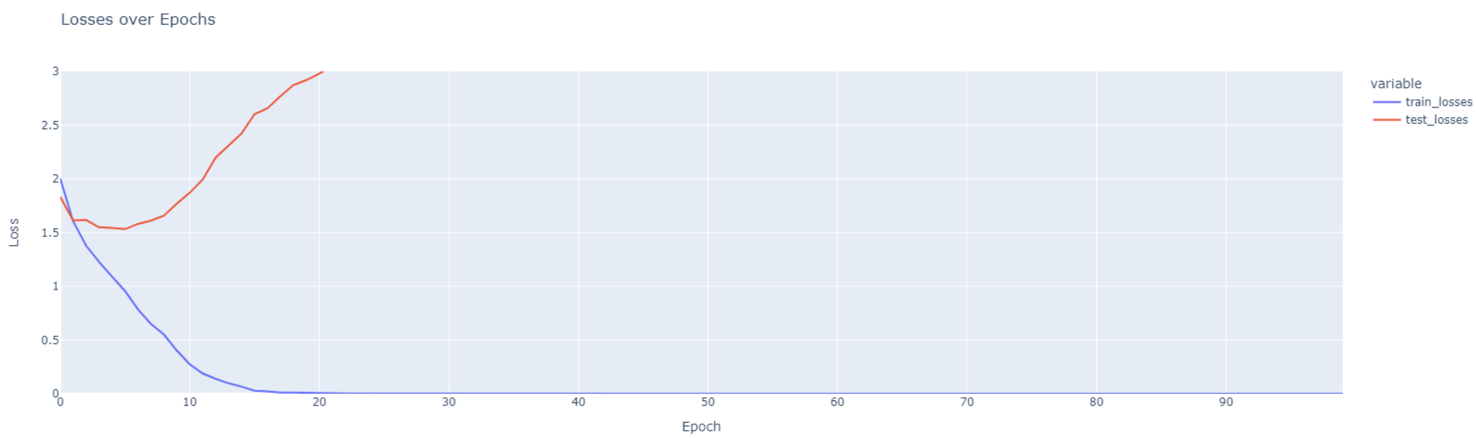




3.2 Optimization

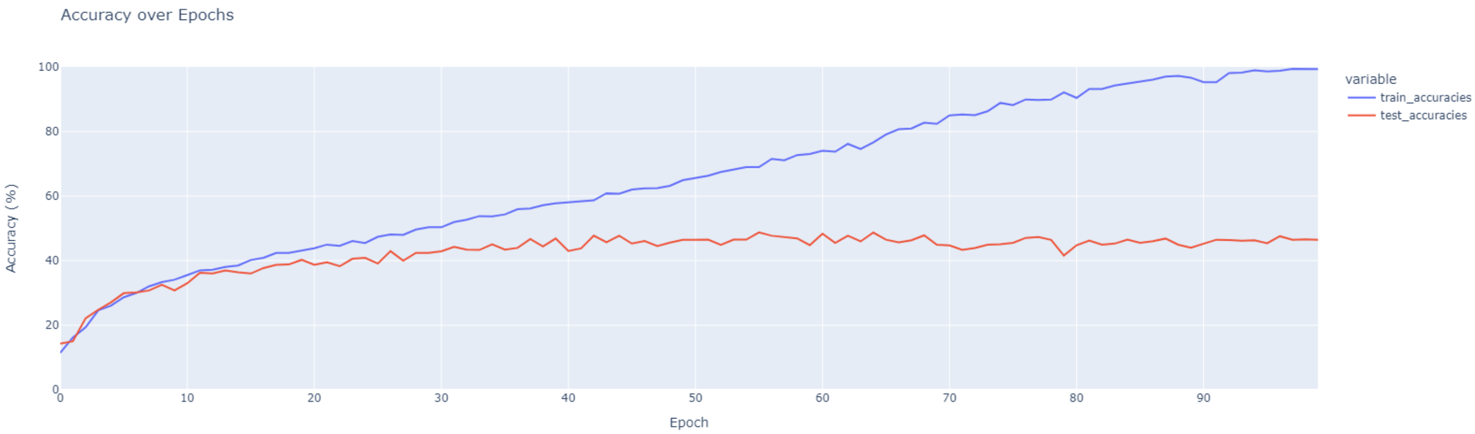
Next we test the Adam optimizer on our data. It seems to completely overfit in the training process, and it does so very quickly. **The test-set accuracy was 50% and the test loss was above 3** (as shown in the scond graph).

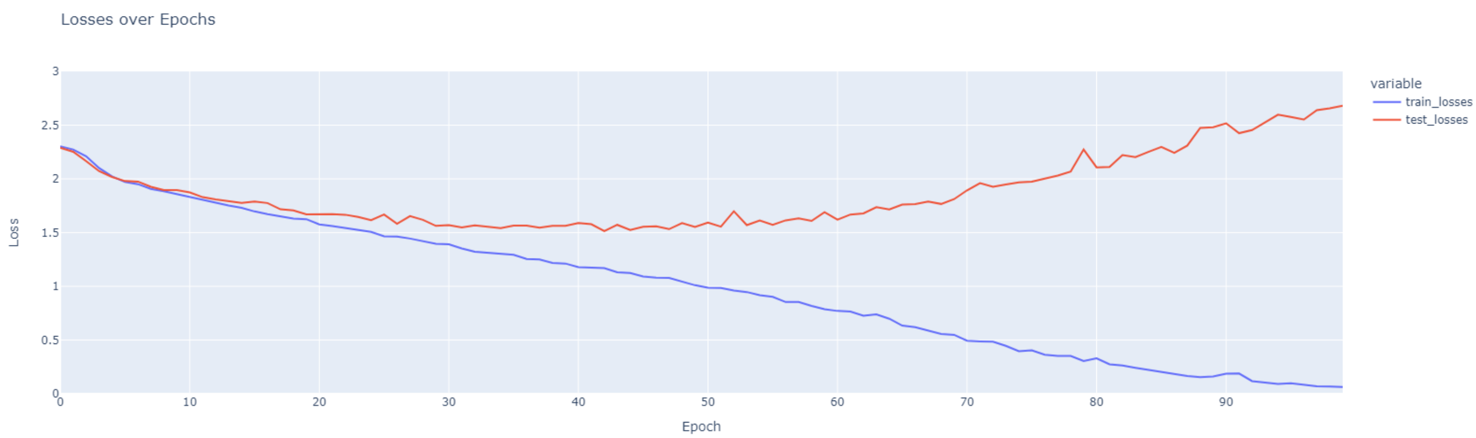




3.3 Initialization

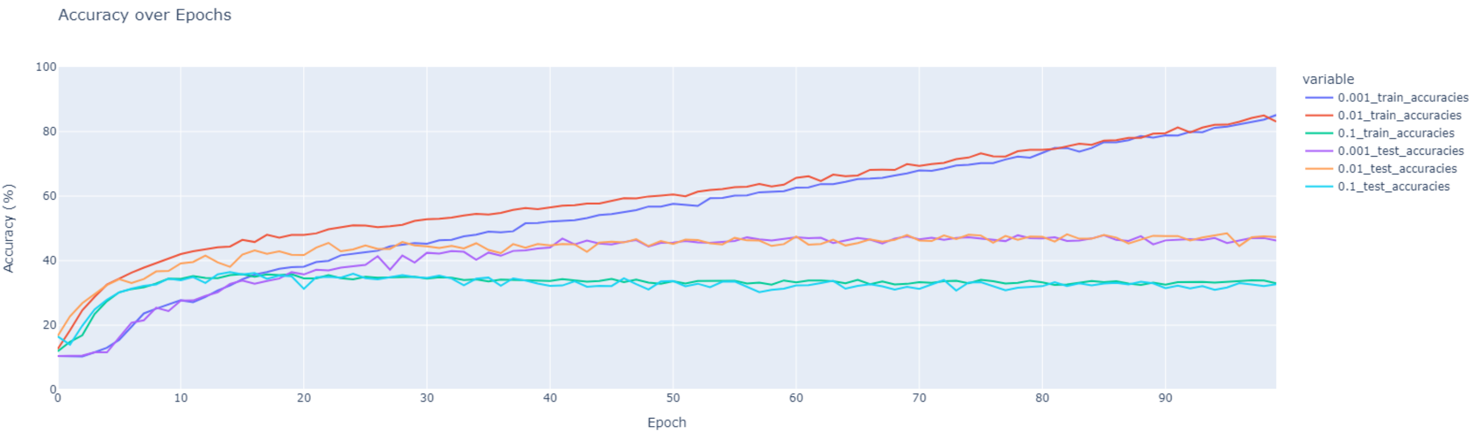
We move on to use Xavier initialization. It also seems to overfit on the training set, but the convergence rate is not as high as in the last section and the test loss is better. The process ends with **test-set accuracy of 46.75% and loss of 2.68.**

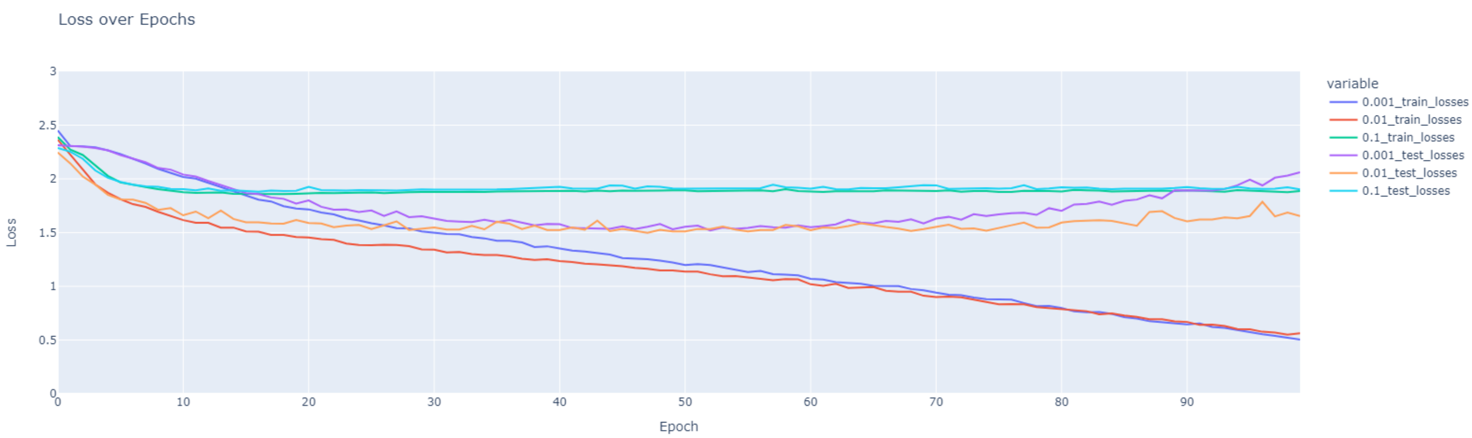




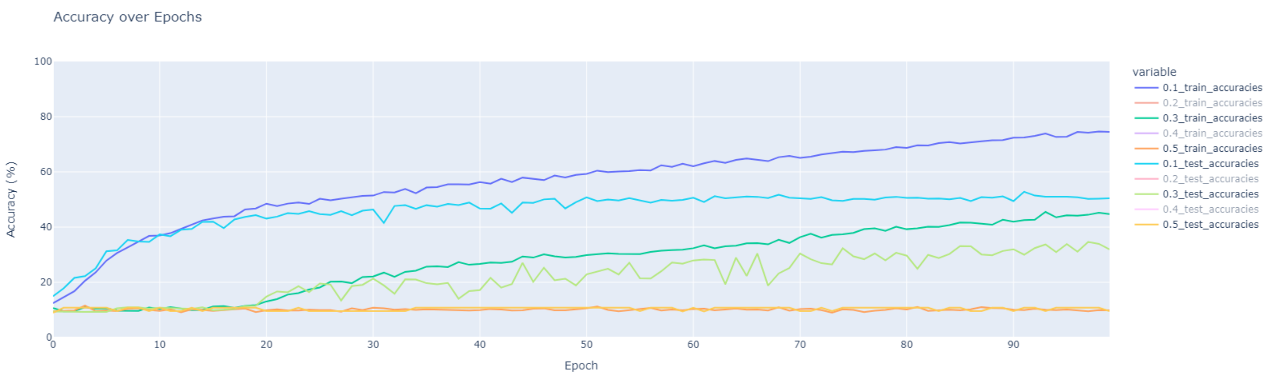
3.4 Regularization

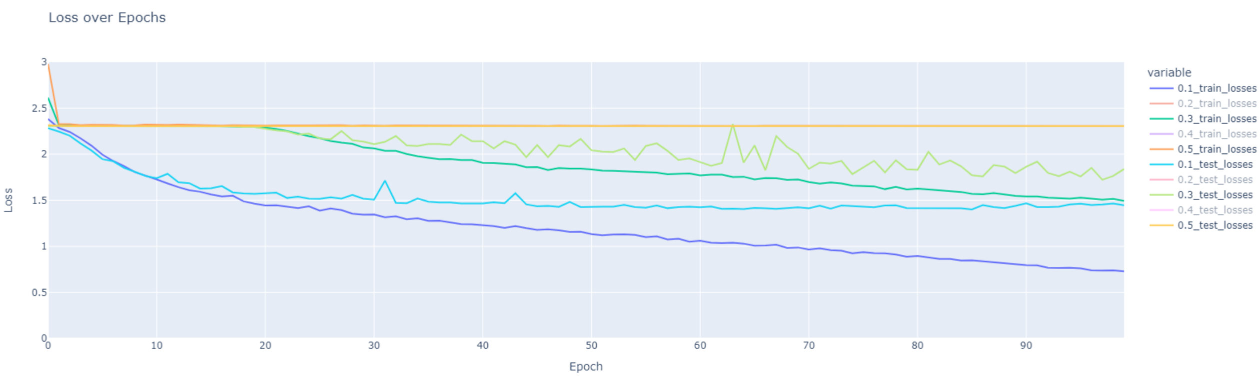
Experiments with weight decay showed lower performance for decay 0.1 and similar performance for decay 0.01, 0.001. It is noteworthy that for decay 0.1, the train accuracy and test accuracy were very similar. The optimal decay value achieved **test-set accuracy of 47% and loss of 1.68**.





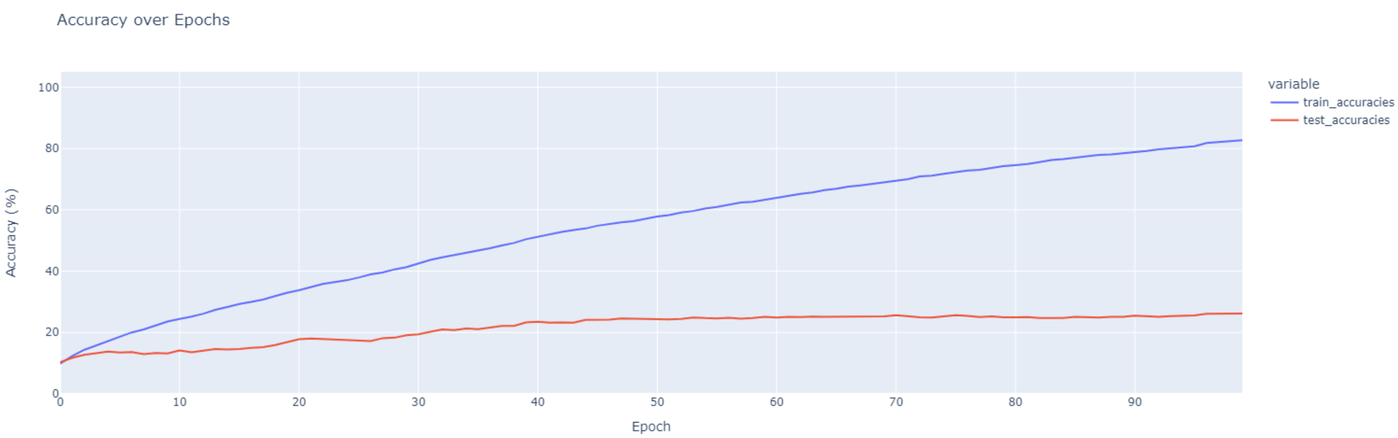
Experiments with dropout showed notable performance difference between different dropout values. Dropout of 0.1 yielded better performance than the baseline, 0.2-0.3 showed slight degredation, and 0.4-0.5 resulted in the model not learning at all. The optimal dropout value achieved **test-set accuracy of 50% and loss of 1.43**.

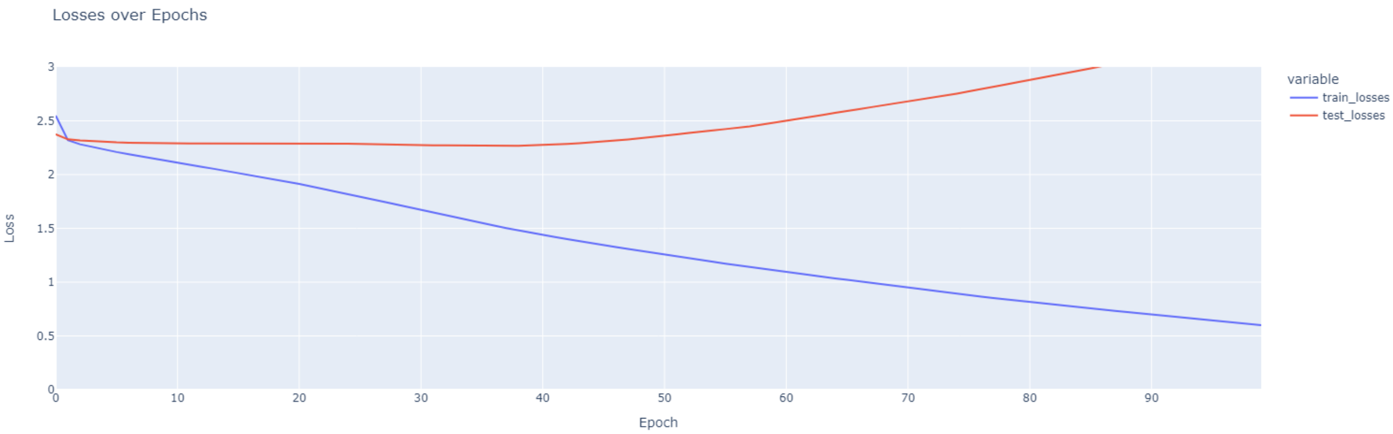

Accuracy over Epochs
Accuracy (%)
Epoch
variable
0.1_train_accuracies
0.2_train_accuracies
0.3_train_accuracies
0.4_train_accuracies
0.5_train_accuracies
0.1_test_accuracies
0.2_test_accuracies
0.3_test_accuracies
0.4_test_accuracies
0.5_test_accuracies



3.5 Preprocessing

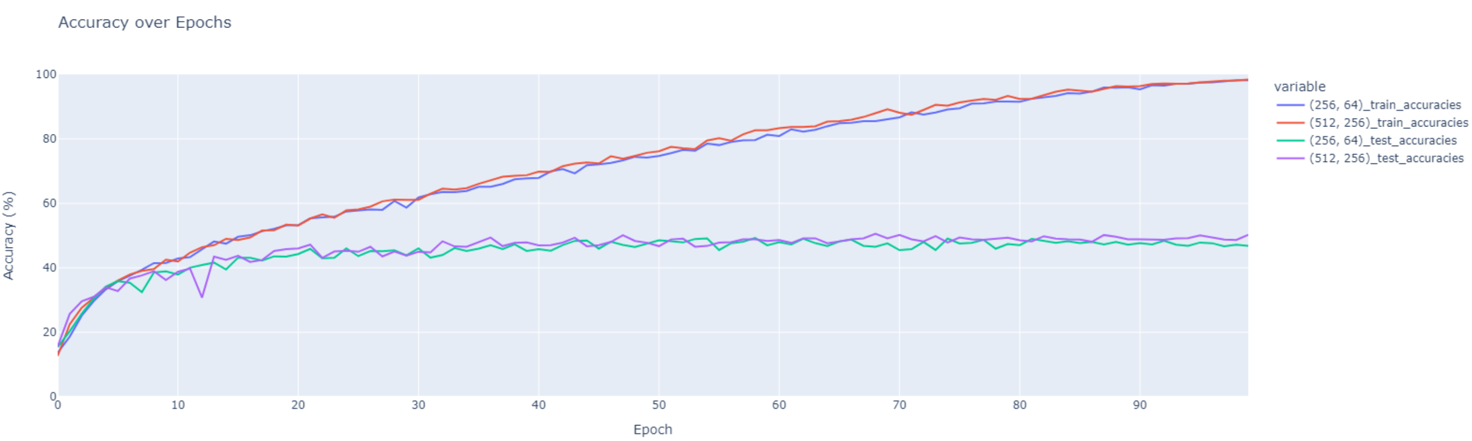
Performing PCA on the data (300 components) seems to increase the training accuracy & loss significantly, yet degrade the test accuracy & loss. We also note that the process seems more stable. The process converges to **test-set accuracy 26% and loss above 3.**





3.6 Network Width

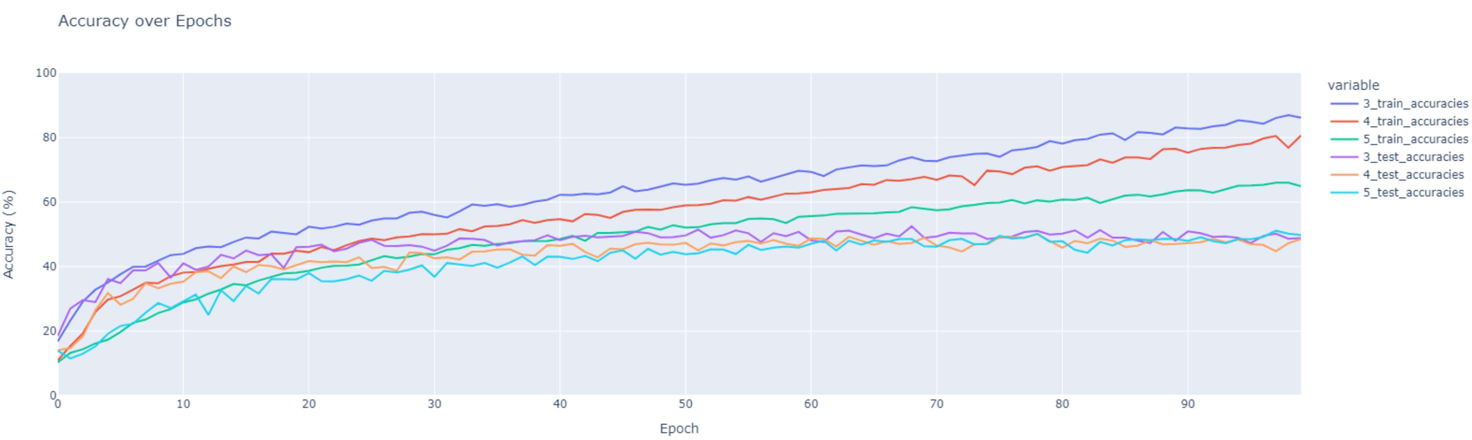
We compare the performance of networks with filter sizes (256, 64) and (512, 256). Their performance is very similar to the baseline, except the training seems to totally overfit. For both cases, **the test-set accuracy is around 48% and the loss is aroudn 2.5.**

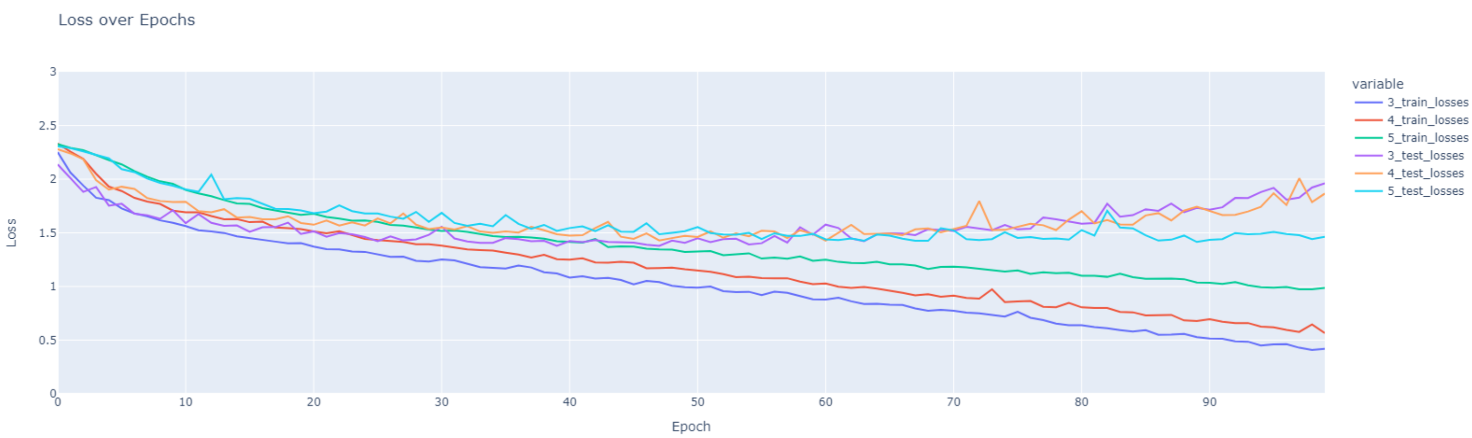


3.7 Network Depth

We compare the performance of networks with 3,4,5 convolutional layers. They have equal accuracy compared to the baseline, yet they have notably lower losses than the baseline. The test-set accuracy and loss are presented in the following table.

Depth	Test Accuracy	Test Loss
3	48%	1.96
4	48%	1.86
5	38%	1.46




Loss over Epochs
Loss
Epoch
variable
3_train_losses
4_train_losses
5_train_losses
3_test_losses
4_test_losses
5_test_losses
0
0.5
1
1.5
2
2.5
3
0
10
20
30
40
50
60
70
80
90