Assignment 1: Standard Practices in Supervised Deep Learning

Dana Arad 311183321

Amit Elyasi 316291434

Part 1 - Setup and Baseline

Linear kernel

```
train accuracy with linear:  0.9862
test accuracy with linear:  0.294
```

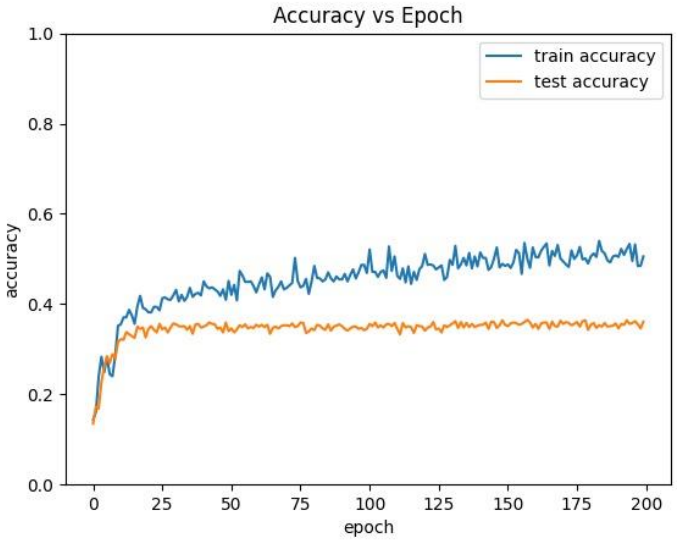
RBF kernel

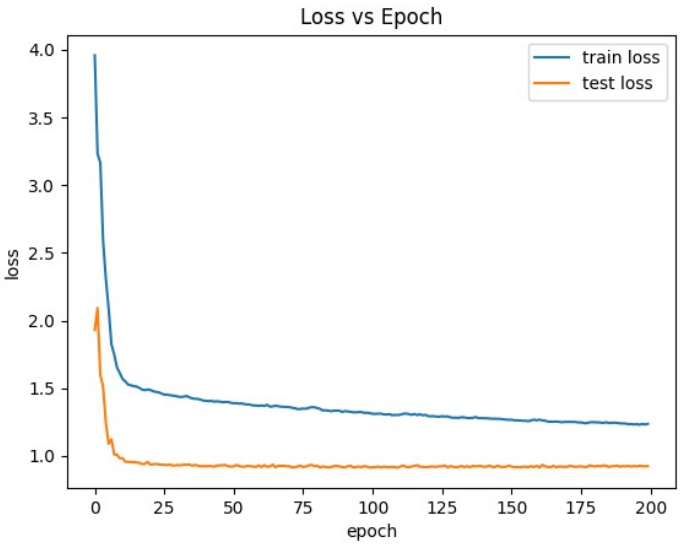
```
train accuracy with rbf:  0.719
test accuracy with rbf:  0.418
```

Part 2 - Feed Forward Neural Network

1. Baseline

We performed a grid search over the learning rate (0.1, 0.01, 0.001), the momentum (0.9, 0.5, 0.1), and the standard deviation (1.0, 0.5, 0.1).





```
(/vol/scratch/danaarad1/env) danaarad1@c-001:/vol/scratch/danaarad1$ python q2.py
running with lr:0.1, m:0.9, sd:1.0
Running on the GPU
Epoch 0, Train Accuracy: 9.955752212389381%, test accuracy: 10.8% , TrainLoss: 42.0096549987793 , Testloss: 22.809783935546875
Epoch 100, Train Accuracy: 21.01769911504425%, test accuracy: 16.7% , TrainLoss: 4.139572064081828 , Testloss: 2.75341534614563
Epoch 200, Train Accuracy: 23.893805309734514%, test accuracy: 18.5% , TrainLoss: 3.897568702697754 , Testloss: 2.686087131500244
running with lr:0.1, m:0.9, sd:0.5
Running on the GPU
Epoch 0, Train Accuracy: 11.283185840707963%, test accuracy: 10.100000000000001% , TrainLoss: 27.176001866658527 , Testloss: 14.633440017700195
Epoch 100, Train Accuracy: 14.048672566371682%, test accuracy: 14.099999999999998% , TrainLoss: 4.535558144251506 , Testloss: 2.74735164642334
Epoch 200, Train Accuracy: 9.8451327433628 32%, test accuracy: 11.4% , TrainLoss: 4.621268033981323 , Testloss: 2.7738311290740967
running with lr:0.1, m:0.9, sd:0.1
Running on the GPU
Epoch 0, Train Accuracy: 9.070796460176991%, test accuracy: 14.299999999999999% , TrainLoss: 10.646092573801676 , Testloss: 11.838658332824707
Epoch 100, Train Accuracy: 9.292035398230089%, test accuracy: 11.4% , TrainLoss: 4.621056079864502 , Testloss: 2.772688865661621
Epoch 200, Train Accuracy: 10.287610619469026%, test accuracy: 11.4% , TrainLoss: 4.621012608210246 , Testloss: 2.772667169570923
running with lr:0.1, m:0.5, sd:1.0
Running on the GPU
Epoch 0, Train Accuracy: 9.513274336283185%, test accuracy: 9.8% , TrainLoss: 47.272833506266274 , Testloss: 19.508663177490234
Epoch 100, Train Accuracy: 28.98230088495575%, test accuracy: 24.2% , TrainLoss: 6.132649819056193 , Testloss: 3.1774048805236816
Epoch 200, Train Accuracy: 30.752212389380528%, test accuracy: 25.5% , TrainLoss: 5.083605686823527 , Testloss: 3.5400896072387695
running with lr:0.1, m:0.5, sd:0.5
Running on the GPU
Epoch 0, Train Accuracy: 9.734513274336283%, test accuracy: 10.100000000000001% , TrainLoss: 22.402713139851887 , Testloss: 10.421036720275879
Epoch 100, Train Accuracy: 23.893805309734514%, test accuracy: 22.5% , TrainLoss: 3.806590216064453 , Testloss: 2.409684896469116
Epoch 200, Train Accuracy: 19.24778761061947%, test accuracy: 17.5% , TrainLoss: 3.9929820696512857 , Testloss: 2.552320957183838
running with lr:0.1, m:0.5, sd:0.1
Running on the GPU
Epoch 0, Train Accuracy: 10.06637168141593%, test accuracy: 11.3% , TrainLoss: 9.832101980845133 , Testloss: 9.0056734085083
Epoch 100, Train Accuracy: 9.955752212389381%, test accuracy: 10.9% , TrainLoss: 4.622044722239177 , Testloss: 2.773178815841675
Epoch 200, Train Accuracy: 10.06637168141593%, test accuracy: 10.9% , TrainLoss: 4.621514558792114 , Testloss: 2.772897481918335
running with lr:0.1, m:0.1, sd:1.0
Running on the GPU
Epoch 0, Train Accuracy: 6.8584070796460175%, test accuracy: 5.8999999999999995% , TrainLoss: 45.854614893595375 , Testloss: 19.628219604492188
Epoch 100, Train Accuracy: 20.6858407079646%, test accuracy: 18.3% , TrainLoss: 4.963893175125122 , Testloss: 3.1800975799560547
Epoch 200, Train Accuracy: 21.792035398230087%, test accuracy: 20.200000000000003% , TrainLoss: 4.364752372105916 , Testloss: 2.868678331375122
running with lr:0.1, m:0.1, sd:0.5
Running on the GPU
Epoch 0, Train Accuracy: 5.641592920353982%, test accuracy: 7.6% , TrainLoss: 19.627463817596436 , Testloss: 6.790407180786133
Epoch 100, Train Accuracy: 26.991150442477874%, test accuracy: 20.9% , TrainLoss: 5.916882832845052 , Testloss: 3.532999038696289
running with lr:0.1, m:0.5, sd:1.0
Running on the GPU
Epoch 0, Train Accuracy: 9.513274336283185%, test accuracy: 9.8% , TrainLoss: 47.272833506266274 , Testloss: 19.508663177490234
Epoch 100, Train Accuracy: 28.98230088495575%, test accuracy: 24.2% , TrainLoss: 6.132649819056193 , Testloss: 3.1774048805236816
Epoch 200, Train Accuracy: 30.752212389380528%, test accuracy: 25.5% , TrainLoss: 5.083605686823527 , Testloss: 3.5400896072387695
running with lr:0.01, m:0.9, sd:0.5
Running on the GPU
Epoch 0, Train Accuracy: 4.756637168141593%, test accuracy: 4.9% , TrainLoss: 24.462753613789875 , Testloss: 11.469594955444336
Epoch 100, Train Accuracy: 28.097345132743364%, test accuracy: 23.1% , TrainLoss: 2.472739060719808 , Testloss: 1.7602057456970215

Epoch 200, Train Accuracy: 35.73008849557522%, test accuracy: 25.4% , TrainLoss: 1.987736741701762 , Testloss: 1.5325467586517334
```

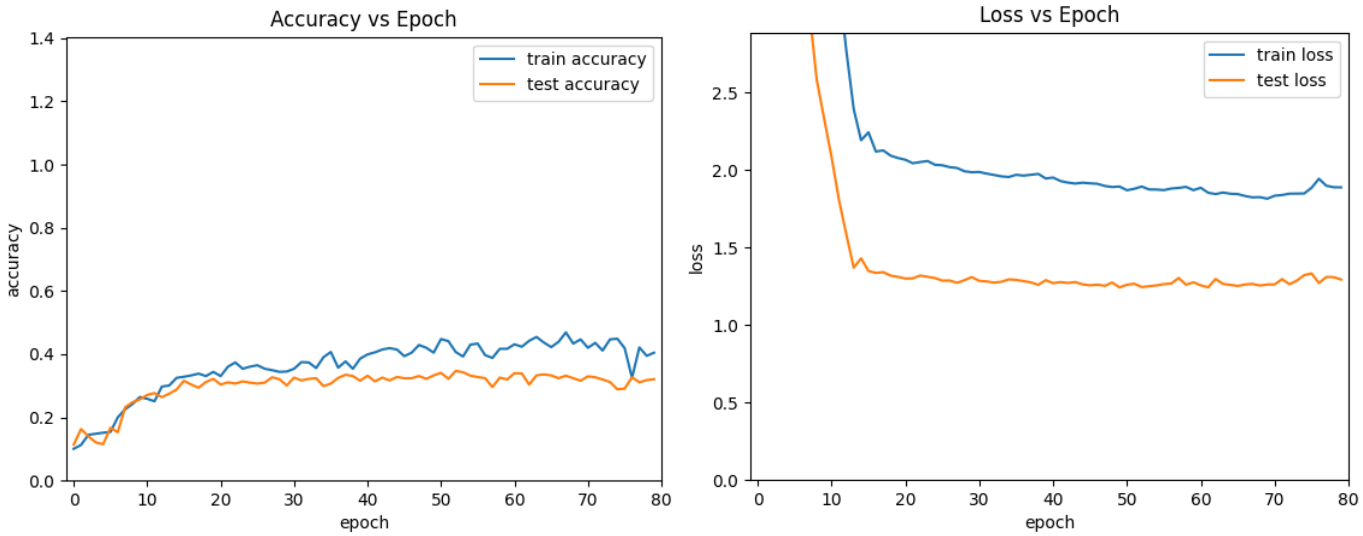
The best result was obtained with a learning rate of 0.01, momentum of 0.9 and parameters were initialized randomly by sampling from a zero-mean Gaussian distribution with a standard deviation of 0.1:

```
running with lr:0.01, m:0.9, sd:0.1
Running on the GPU
Epoch 0, Train Accuracy: 12.5%, test accuracy: 13.8% , TrainLoss: 4.570571978886922 , Testloss: 1.8519604206085205
Epoch 100, Train Accuracy: 45.575221238938053%, test accuracy: 32.2% , TrainLoss: 1.3452197114626567 , Testloss: 0.9543548226356506
Epoch 200, Train Accuracy: 51.43805309734514%, test accuracy: 32.300000000000004% , TrainLoss: 1.2466452320416768 , Testloss: 0.9502150416374207
```

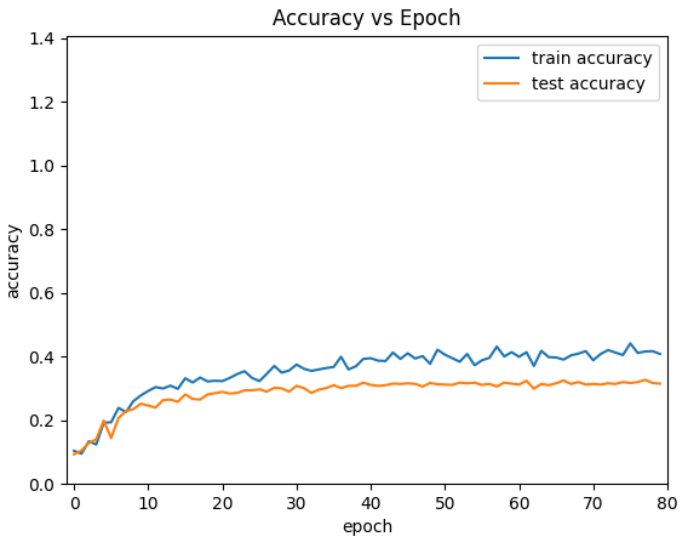
2. Optimization

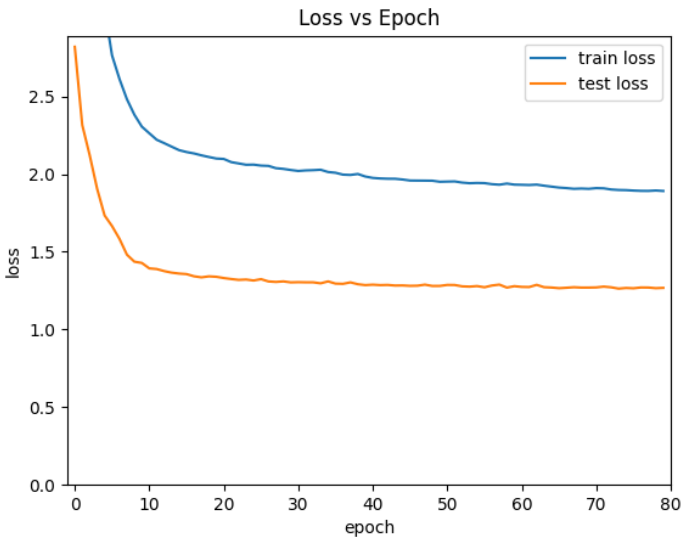
Using the base line from the previous question, we added the use of an optimizer and compared the performance with SGD optimizer vs ADAM optimizer:

Adam optimizer:



SGD optimizer:



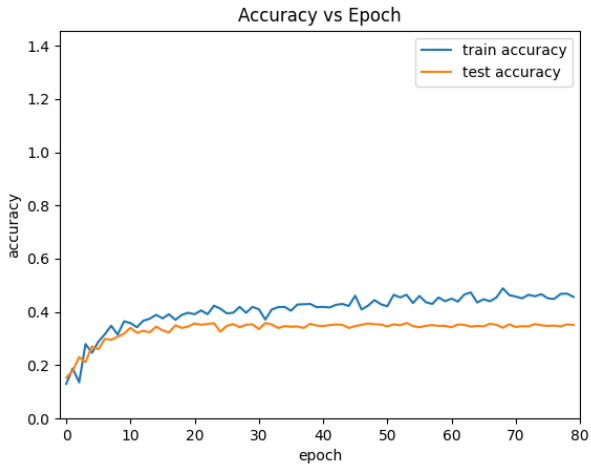


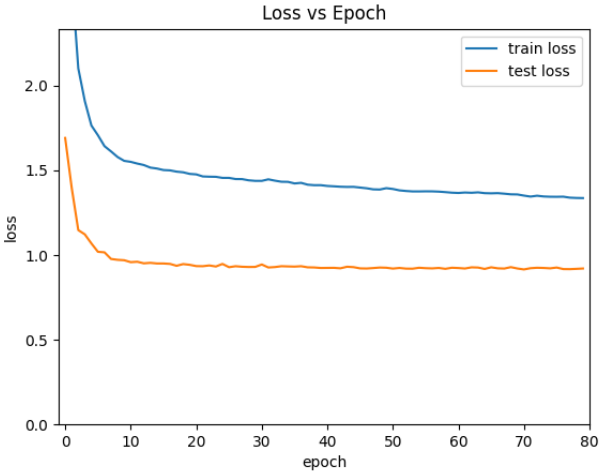

```
Adam: lr=0.01, momentum=0.9 sd=0.1
Running on the GPU
Epoch 0, Train Accuracy: 9.402654867256636%, test accuracy: 10.9% , TrainLoss: 8.169473648071289 , Testloss: 10.945428848266602
Epoch 100, Train Accuracy: 43.80530973451327%, test accuracy: 31.4% , TrainLoss: 1.402742902437846 , Testloss: 1.0189645290374756
Epoch 200, Train Accuracy: 53.31858407079646%, test accuracy: 30.3% , TrainLoss: 1.2516236702601116 , Testloss: 1.087253212928772
SGD: lr=0.01, momentum=0.9 sd=0.1
Running on the GPU
Epoch 0, Train Accuracy: 11.283185840707963%, test accuracy: 10.0% , TrainLoss: 5.271224578221639 , Testloss: 2.417876958847046
Epoch 100, Train Accuracy: 45.57522123893805%, test accuracy: 33.4% , TrainLoss: 1.3627516825993855 , Testloss: 0.9619688391685486
Epoch 200, Train Accuracy: 47.01327433628318%, test accuracy: 34.8% , TrainLoss: 1.2586273749669392 , Testloss: 0.9460850954055786
```

Using the baseline hyper-parameters, it appears the training accuracy is better when using Adam optimizer, but the better test accuracy was obtained when using SGD. After seeing these results we also found online discussions that support the claim that in some cases SGD generalizes better then ADAM, which fit our results.

3. Initialization

Using Xavier initialization:



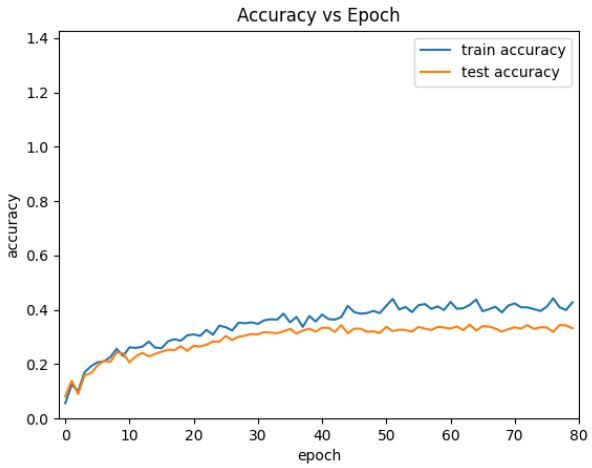


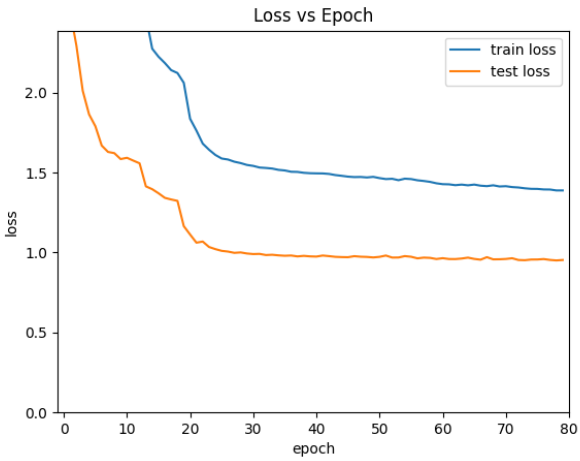

```
Xavier: SGD lr=0.01, momentum=0.9 sd=0.1
Running on the GPU
Epoch 0, Train Accuracy: 14.15929203539823%, test accuracy: 10.9% , TrainLoss: 3.8164892196655273 , Testloss: 1.6594510078430176
Epoch 100, Train Accuracy: 45.90707964601771%, test accuracy: 35.0% , TrainLoss: 1.3191478252410889 , Testloss: 0.918856680393219
Epoch 200, Train Accuracy: 51.991150442477874%, test accuracy: 35.699999999999996% , TrainLoss: 1.2317133943239849 , Testloss: 0.9285532236099243
Normal init: SGD lr=0.01, momentum=0.9 sd=0.1
Running on the GPU
Epoch 0, Train Accuracy: 9.513274336283185%, test accuracy: 8.4% , TrainLoss: 5.432631174723308 , Testloss: 2.7607860565185547
Epoch 100, Train Accuracy: 42.36725663716815%, test accuracy: 33.6% , TrainLoss: 1.8952709039052327 , Testloss: 1.2452442646026611
Epoch 200, Train Accuracy: 48.783185840707965%, test accuracy: 34.5% , TrainLoss: 1.7841605345408122 , Testloss: 1.2256819009780884
```

As expected, the test accuracy is higher using Xavier initialization. Both experiments ran for the same number of epochs, so we can therefore say the convergence time is better with Xavier initialization.

4. Regularization

Using the best configuration (wd=0, dropout=0.005)






```
Dropout=0.005, Weight Decay=0: Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the GPU
Epoch 200, Train Accuracy: 52.986725663716825%, test accuracy: 35.5% , TrainLoss: 1.2738404075304668 , Testloss: 0.9302968978881836
Dropout=0.005, Weight Decay=0.001: Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
```

When preforming grid search over the dropout and weight decay with values of (0, 0.001, 0.005, 0.009), it appears the best accuracy obtained with dropout rate of 0.005 and weight decay of 0.

The runtime did not vary much between configurations (was about a minute for each), but it seems the best runtime was obtained when using 0.009 for both dropout and weight decay.

The full results:

```
Dropout=0, Weight Decay=0:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 59.51327433628318%, test accuracy: 38.1% , TrainLoss: 1.0091150204340618 , Testloss: 0.8901153802871704
runtime: 55.32379174232483

Dropout=0, Weight Decay=0.001:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 61.17256637168141%, test accuracy: 39.5% , TrainLoss: 1.0102195938428242 , Testloss: 0.8646039366722107
runtime: 67.1025607585907

Dropout=0, Weight Decay=0.005:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 55.64159292035398%, test accuracy: 38.2% , TrainLoss: 1.0739732384681702 , Testloss: 0.8615325093269348
runtime: 56.15059542655945

Dropout=0, Weight Decay=0.009:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 51.216814159292035%, test accuracy: 38.1% , TrainLoss: 1.157835304737091 , Testloss: 0.849243700504303
runtime: 59.690314054489136

Dropout=0.001, Weight Decay=0:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 58.075221238938056%, test accuracy: 39.7% , TrainLoss: 1.046691815058390 4 , Testloss: 0.8793240189552307
runtime: 54.570722818374634
```

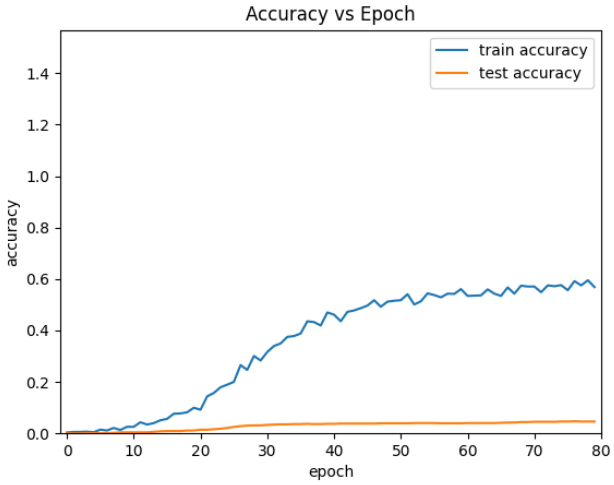
```
Dropout=0.001, Weight Decay=0.001:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 62.057522123893804%, test accuracy: 37.7% , TrainLoss: 1.0190854668617249 , Testloss: 0.894524335861206
runtime: 55.74453854560852

Dropout=0.001, Weight Decay=0.005:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 55.862831858407085%, test accuracy: 40.0% , TrainLoss: 1.0782677133878071 , Testloss: 0.8585054278373718
runtime: 57.04437971115112

Dropout=0.001, Weight Decay=0.009:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 50.33185840707964%, test accuracy: 39.4% , TrainLoss: 1.543711006641388 , Testloss: 1.086220145225525
runtime: 55.31844472885132

Dropout=0.005, Weight Decay=0:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 59.51327433628318%, test accuracy: 39.800000000000004% , TrainLoss: 0.9755876859029134 , Testloss: 0.8921000361442566
runtime: 55.571391582489014

Dropout=0.005, Weight Decay=0.001:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 63.495575221238944%, test accuracy: 38.0% , TrainLoss: 1.3326427936553955 , Testloss: 1.0503088235855103
runtime: 57.095712423324585
```

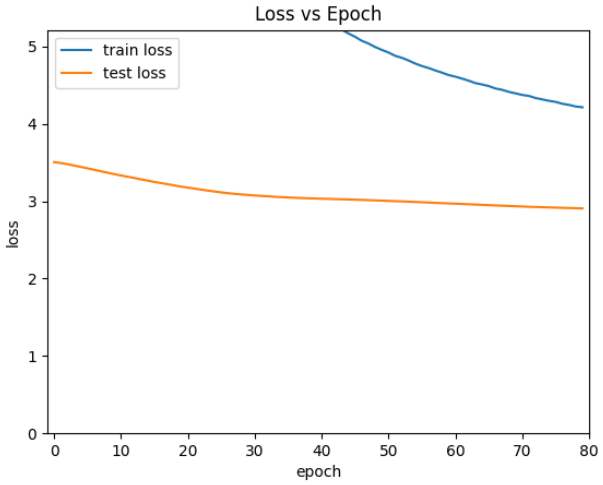
```
Dropout=0.005, Weight Decay=0.005:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 55.309734513274336%, test accuracy: 39.1% , TrainLoss: 1.1156868934631348 , Testloss: 0.8664318919181824
runtime: 58.422532081604004

Dropout=0.005, Weight Decay=0.009:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 57.63274336283186%, test accuracy: 37.4% , TrainLoss: 1.1322887142499287 , Testloss: 0.8525092005729675
runtime: 64.70559811592102

Dropout=0.009, Weight Decay=0:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 57.96460176991151%, test accuracy: 39.5% , TrainLoss: 0.991459846496582 , Testloss: 0.8773033022880554
runtime: 59.30212712287903

Dropout=0.009, Weight Decay=0.001:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 57.63274336283186%, test accuracy: 37.2% , TrainLoss: 1.3581884304682414 , Testloss: 1.0918428897857666
runtime: 59.30420231819153

Dropout=0.009, Weight Decay=0.005:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 55.530973451327434%, test accuracy: 37.0% , TrainLoss: 1.4602909485499065 , Testloss: 1.0915788412094116
runtime: 49.426220655441284
```

```
Dropout=0.009, Weight Decay=0.009:
 Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the CPU
Epoch 200, Train Accuracy: 55.64159292035398%, test accuracy: 41.199999999999996% , TrainLoss: 1.1600195566813152 , Testloss: 0.848567545413971
runtime: 46.805907249450684
```

5. Preprocessing

With preforming whitening:





Best results using whitening:

```
Best epoch 77, Train Accuracy: 59.85%, test accuracy: 0.3% , TrainLoss: 4.19 , Testloss: 4.67
runtime: 68.67
```

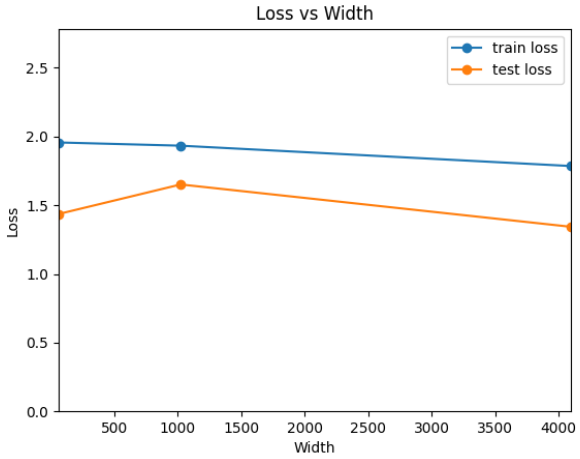
Best results without whitening:

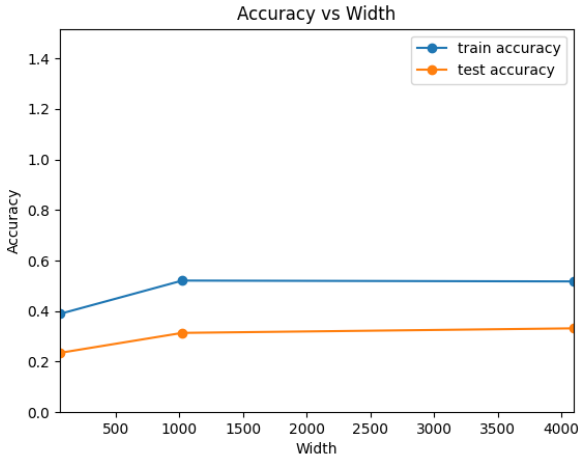
```
Best epoch 76, Train Accuracy: 44.58%, test accuracy: 34.0% , TrainLoss: 1.4 , Testloss: 0.96
runtime: 30.82
```

Using whitening increased the training accuracy significantly, to almost 60%. However, the test accuracy decreased to less than 1% which means the model preformed even worse than a random classifier. This combination suggests that in this case adding the whitening caused the model to overfit the training data and that this model will not generalize well.

In addition, using whitening almost doubles the training runtime.

6. Network Width

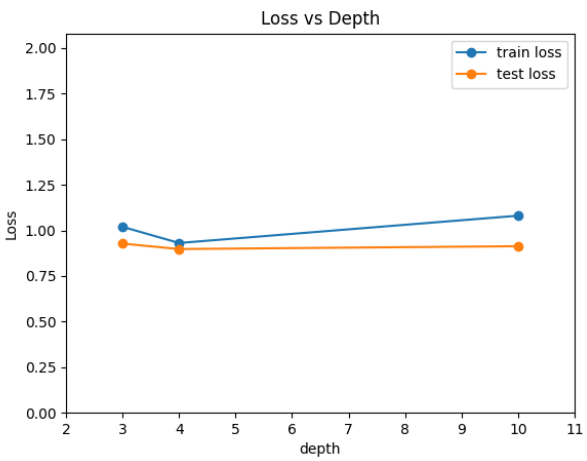


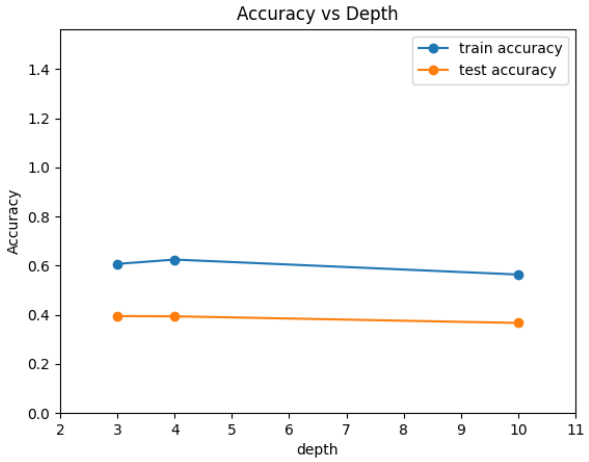



```
width 64: Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1, dropout=0.005, wd=0
Running on the CPU
Epoch 200, Train Accuracy: 38.93805309734513%, test accuracy: 23.5% , TrainLoss: 1.956958810488383 , Testloss: 1.436587929725647
runtime: 84.30295276641846

width 1024: Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1, dropout=0.005, wd=0
Running on the CPU
Epoch 200, Train Accuracy: 52.10176991150443%, test accuracy: 31.4% , TrainLoss: 1.9335975646972656 , Testloss: 1.6512720584869385
runtime: 1345.0870671272278

width 4096: Xavier, SGD, lr=0.01, momentum=0.9 sd=0.1, dropout=0.005, wd=0
Running on the CPU
Epoch 200, Train Accuracy: 51.76991150442478%, test accuracy: 33.2% , TrainLoss: 1.7854748169581096 , Testloss: 1.3433334827423096
runtime: 71770.35615372658
```

It seems that when increasing the width up to 2^{10} the loss increases as well as the accuracy. increasing the width even more does not cause much more improvement in accuracy, but the loss decreases. Overall, increasing the width improves the results but with the tradeoff of very long runtime. As we saw in class “there are no free lunches” and so we need to find a correct balance between the width and the runtime somewhere in the middle.

7. Network Depth






```
depth 3:
Running on the CPU
Epoch 200, Train Accuracy: 55.75221238938053%, test accuracy: 37.1% , TrainLoss: 1.2764645616213481 , Testloss: 1.0531295537948608
runtime: 61.76734519004822

depth 4:
Running on the CPU
Epoch 200, Train Accuracy: 64.71238938053098%, test accuracy: 37.5% , TrainLoss: 0.8528250753879547 , Testloss: 1.0006681680679321
runtime: 62.07352876663208

depth 10:
Running on the CPU
Epoch 200, Train Accuracy: 51.32743362831859%, test accuracy: 36.4% , TrainLoss: 1.1178983648618062 , Testloss: 0.9344442486763
runtime: 85.44237017631531
```

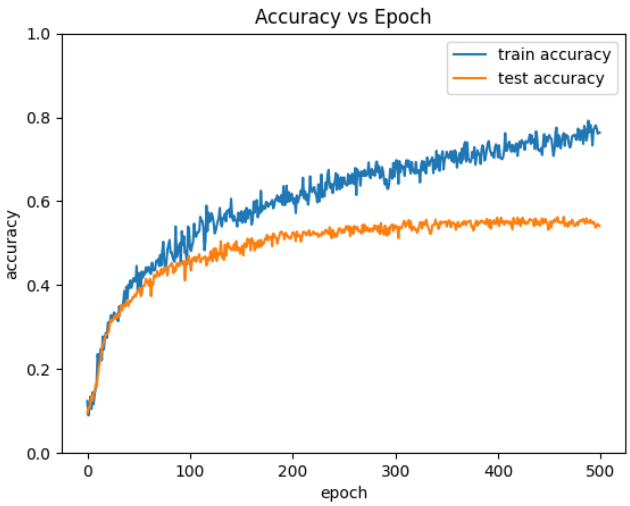
Again, it apears the best configuration is somewhere in the middle: the loss decreses when we move from 3 layers to 4, but when increasing the depth farther it seems we get higher loss again. The accuracy behaves quite the same: it increases as the depth increases up to 4, but decreases when increasing the depth farther. The runtime does not change much between experiments. These results suggest that the network depth is not the most dominant hyper-parameter out of the parameters we examined.

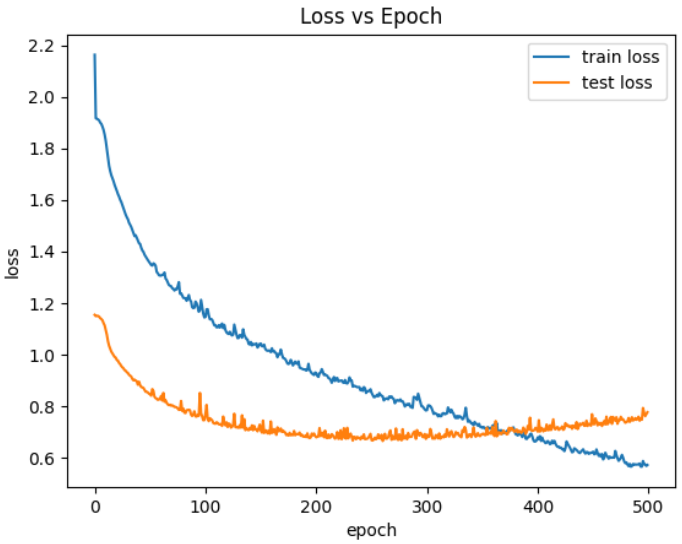
Part 3 - Convolutional Neural Network

1. Baseline

We performed a grid search over the learning rate (0.1, 0.01, 0.001), the momentum (0.9, 0.5, 0.1, 0), and the standard deviation (1.0, 0.5, 0.25, 0.1). Our best configuration was learning rate of 0.01, momentum of 0.9 and standard deviation of 0.1. we achieved 56.3% test accuracy and 0.737 test loss.

```
running with lr:0.01, m:0.9, sd:0.1
Running on the GPU
Best Epoch: 464, Train Accuracy: 0.7621681415929203, test accuracy: 0.563 , TrainLoss: 0.5974976619084676 , Testloss: 0.7367131114006042
Last Epoch: 499, Train Accuracy: 0.7632743362831859, test accuracy: 0.541 , TrainLoss: 0.5730489691098531 , Testloss: 0.7775813937187195
running with lr:0.01, m:0.5, sd:1.0
Running on the GPU
Best Epoch: 42, Train Accuracy: 0.08960176991150443, test accuracy: 0.113 , TrainLoss: 1.9320207436879475 , Testloss: 1.154518961906433
Last Epoch: 499, Train Accuracy: 0.09955752212389381, test accuracy: 0.113 , TrainLoss: 1.9173962672551472 , Testloss: 1.1539257764816284
running with lr:0.01, m:0.5, sd:0.5
Running on the GPU
Best Epoch: 30, Train Accuracy: 0.09513274336283185, test accuracy: 0.119 , TrainLoss: 1.9218629995981853 , Testloss: 1.153422474861145
Last Epoch: 499, Train Accuracy: 0.10287610619469026, test accuracy: 0.114 , TrainLoss: 1.9132225910822551 , Testloss: 1.1515049934387207
running with lr:0.01, m:0.5, sd:0.25
Running on the GPU
Best Epoch: 324, Train Accuracy: 0.1172566371681416, test accuracy: 0.117 , TrainLoss: 1.9158640305201213 , Testloss: 1.1522531509399414
Last Epoch: 499, Train Accuracy: 0.11836283185840708, test accuracy: 0.113 , TrainLoss: 1.915035565694173 , Testloss: 1.1524174213409424
running with lr:0.01, m:0.5, sd:0.1
Running on the GPU
Best Epoch: 490, Train Accuracy: 0.5630530973451328, test accuracy: 0.5 , TrainLoss: 1.0628163814544678 , Testloss: 0.7337731122970581
Last Epoch: 499, Train Accuracy: 0.5707964601769911, test accuracy: 0.485 , TrainLoss: 1.040474275747935 , Testloss: 0.7381711602210999
```



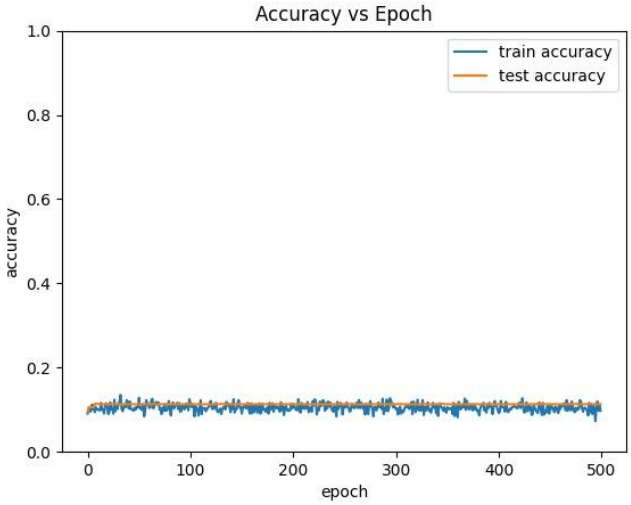


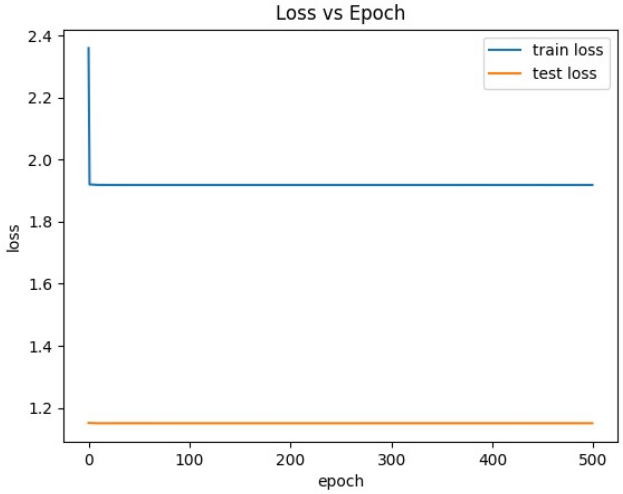
2. Optimization

```
Adam: lr=0.01, momentum=0.9 sd=0.1
Running on the GPU
Best Epoch: 7/500, Train Accuracy: 9.4%, test accuracy: 11.4% , TrainLoss: 1.919 , Testloss: 1.151
SGD: lr=0.01, momentum=0.9 sd=0.1
Running on the GPU
Best Epoch: 287/500, Train Accuracy: 66.81%, test accuracy: 55.7% , TrainLoss: 0.75 , Testloss: 0.693
```

We compared ADAM and SGD using the best configuration from article 1. We got relatively good results for SGD but significantly lower accuracy and higher loss for ADAM. This configuration was obtained by maximizing the accuracy for SGD so it is possible that there are much better configurations for when using ADAM.

The graphs below show the accuracy vs. epoch and loss vs. epoch when using ADAM optimizer. The same graphs for SGD are in article 1.

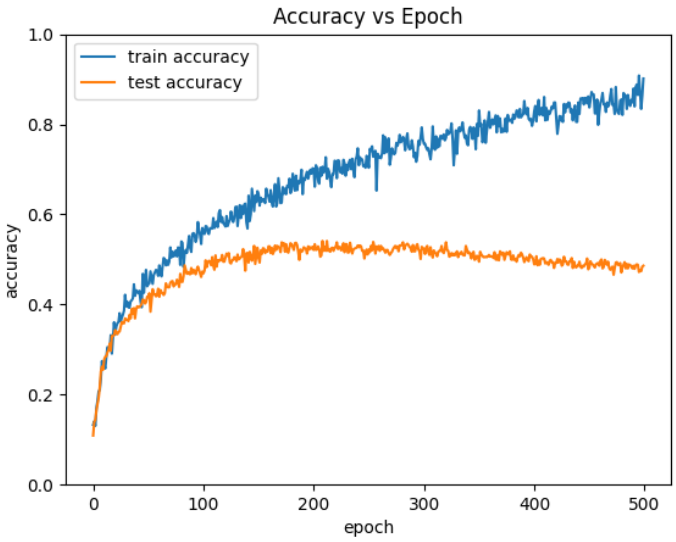


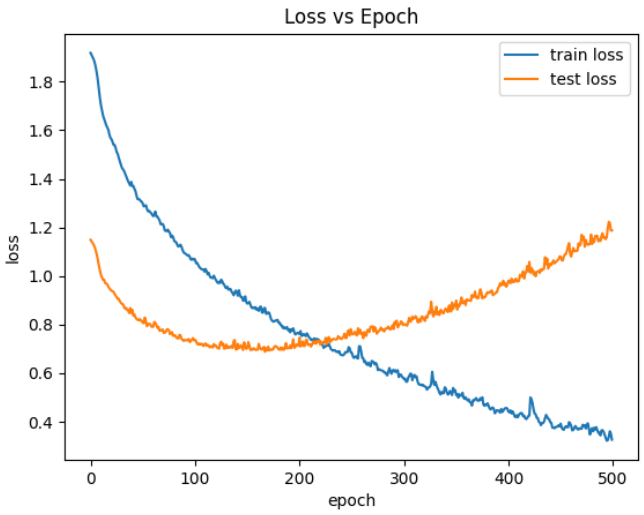


3. Initialization

Using the same configuration, we compared Xavier initialization and initialization with normal distribution. The results with Xavier were very similar, but slightly worst, to the baseline. The convergence time is also similar, with slightly shorter convergence time for Xavier (with Xavier, the best results is achieved in epoch number 250, whereas the baseline achieved best results in epoch number 287)

```
Xavier: SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the GPU
Best Epoch: 250/500, Train Accuracy: 68.03%, Test accuracy: 55.8% , Train loss: 0.733 , Test loss: 0.714
Last Epoch: 499, Train Accuracy: 82.3%, Test accuracy: 51.1% , Train loss: 0.451 , Test loss: 0.982
```

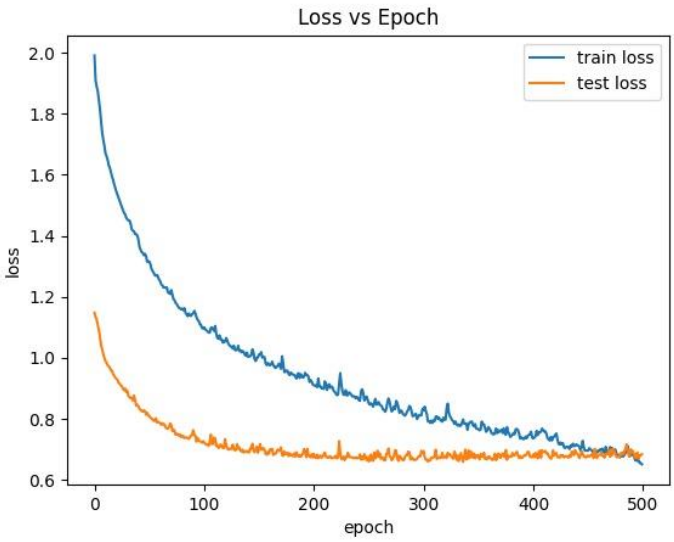


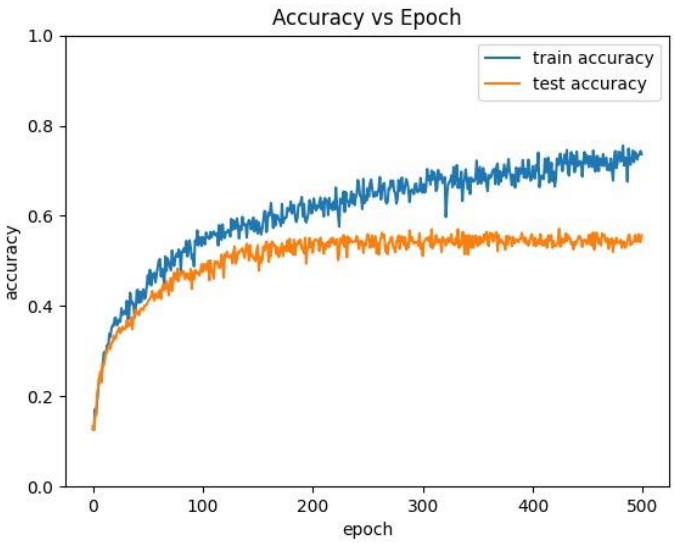


4. Regularization

Using regularization slightly improved our results (57.1% test accuracy compared to 56.3% in the baseline). It seems that regularization slows down the convergence rate – our best epoch was epoch number 424, compared to 287 in the baseline. The runtime did not change much when using different dropout and weight decay values.

```
dropout: 0.001, weight decay: 0.009:
Running on the GPU
Best Epoch: 424/500, Train Accuracy: 71.79%, test accuracy: 57.1% , TrainLoss: 0.716 , Testloss: 0.673
runtime: 115
```

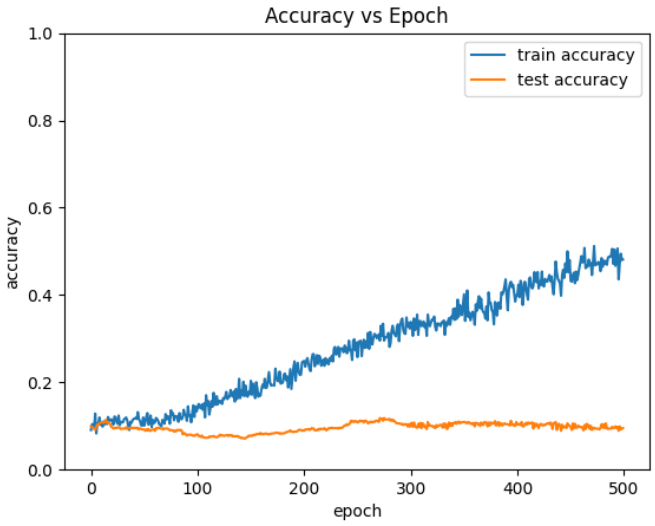


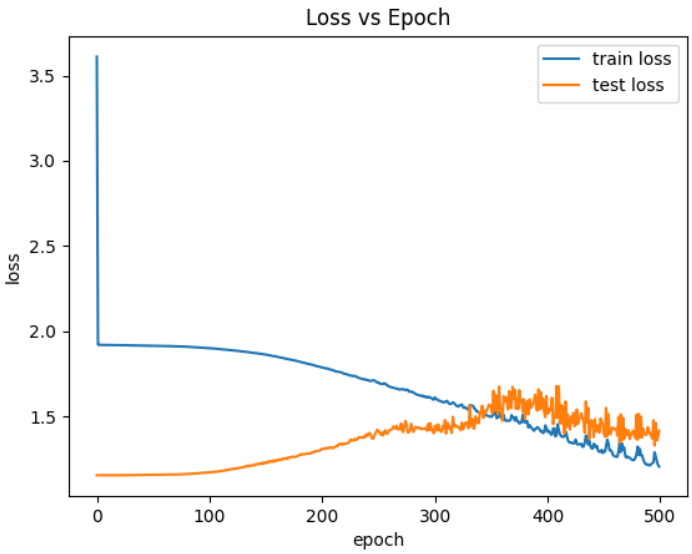


5. Preprocessing

Adding the whitening decreased the overall performance of the model (comparing best epoch) – both train and test accuracy are lower when using the whitening. The test accuracy dropped significantly, which suggests overfitting (similarly to the whitening in question 2). Runtime almost doubles when using whitening.

```
Without whitening: SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the GPU
Best Epoch: 427/500, Train Accuracy: 70.8%, Test accuracy: 52.6% , Train loss: 0.715 , Test loss: 0.777
Last Epoch: 499, Train Accuracy: 75.77%, Test accuracy: 51.4% , Train loss: 0.647 , Test loss: 0.808
runtime: 287

With PCA whitening: SGD, lr=0.01, momentum=0.9 sd=0.1
Running on the GPU
Best Epoch: 149/500, Train Accuracy: 27.1%, Test accuracy: 14.2% , Train loss: 1.713 , Test loss: 1.158
Last Epoch: 499, Train Accuracy: 76.77%, Test accuracy: 9.1% , Train loss: 0.631 , Test loss: 1.261
runtime: 458
```

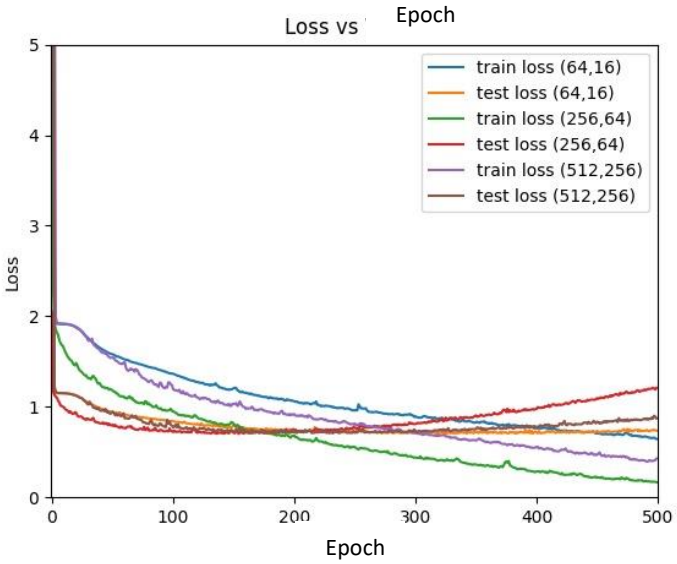


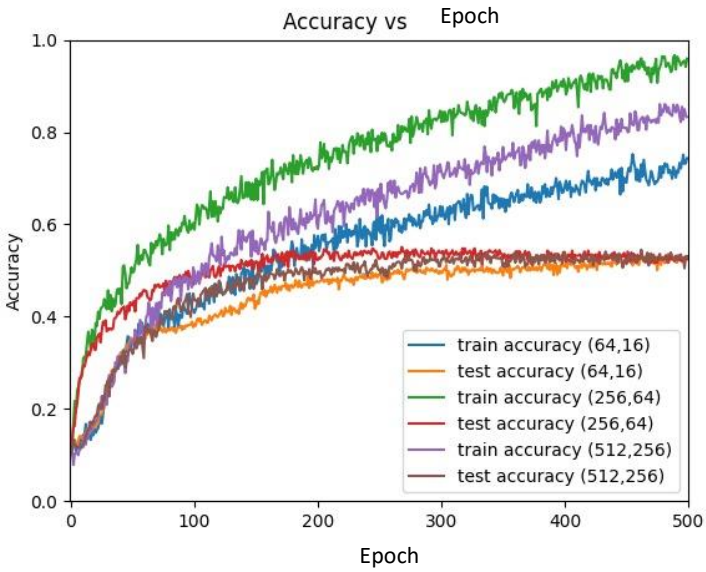


6. Network Width

The widest network preformed the best, with test accuracy of 56.2% and test loss of 0.705. It is expected that a wider network will be able to learn more information since it has more parameters. It also seems that the widest network converges first, as the best epoch is epoch number 204, where in the other settings it is closer to 400.

```
width (64,16): SGD, lr=0.01, momentum=0.9 sd=0.1, dropout=0, wd=0
Best Epoch: 378/500, Train Accuracy: 73.34%, Test accuracy: 55.0% , Train loss: 0.642 , Test loss: 0.738
Last Epoch: 499, Train Accuracy: 81.19%, Test accuracy: 53.4% , Train loss: 0.473 , Test loss: 0.822

width (256,64): SGD, lr=0.01, momentum=0.9 sd=0.1, dropout=0, wd=0
Best Epoch: 390/500, Train Accuracy: 84.62%, Test accuracy: 53.5% , Train loss: 0.448 , Test loss: 0.851
Last Epoch: 499, Train Accuracy: 88.5%, Test accuracy: 52.2% , Train loss: 0.321 , Test loss: 0.976


width (512,256): SGD, lr=0.01, momentum=0.9 sd=0.1, dropout=0, wd=0
Best Epoch: 204/500, Train Accuracy: 71.68%, Test accuracy: 56.2% , Train loss: 0.682 , Test loss: 0.705
Last Epoch: 499, Train Accuracy: 99.0%, Test accuracy: 52.0% , Train loss: 0.092 , Test loss: 1.189
```





7. Network Depth

The best results were obtained when the network had 2 convolution layers. It seems that a higher number of convolution layers did not only decreased the test accuracy but also decreased the training accuracy – this means that the deeper networks did not overfit the training data, they overall preformed worse than shallower networks.

```
2 convolution layers: SGD, lr=0.01, momentum=0.9 sd=0.1, dropout=0, wd=0
Running on the GPU
Best Epoch: 379/500, Train Accuracy: 69.8%, Test accuracy: 54.1% , Train loss: 0.754 , Test loss: 0.74
Last Epoch: 499, Train Accuracy: 76.11%, Test accuracy: 51.5% , Train loss: 0.629 , Test loss: 0.78
3 convolution layers: SGD, lr=0.01, momentum=0.9 sd=0.1, dropout=0, wd=0
Running on the GPU
Best Epoch: 275/500, Train Accuracy: 76.33%, Test accuracy: 53.3% , Train loss: 0.538 , Test loss: 0.764
Last Epoch: 499, Train Accuracy: 97.23%, Test accuracy: 49.9% , Train loss: 0.112 , Test loss: 1.416
4 convolution layers: SGD, lr=0.01, momentum=0.9 sd=0.1, dropout=0, wd=0
Running on the GPU
Best Epoch: 110/500, Train Accuracy: 62.39%, Test accuracy: 49.1% , Train loss: 0.886 , Test loss: 0.74
Last Epoch: 499, Train Accuracy: 100.0%, Test accuracy: 45.2% , Train loss: 0.005 , Test loss: 2.647
5 convolution layers: SGD, lr=0.01, momentum=0.9 sd=0.1, dropout=0, wd=0
Running on the GPU
Best Epoch: 25/500, Train Accuracy: 10.4%, Test accuracy: 11.6% , Train loss: 1.919 , Test loss: 1.148
Last Epoch: 499, Train Accuracy: 11.62%, Test accuracy: 11.3% , Train loss: 1.918 , Test loss: 1.151
```