Thomas Jones – CS 5567-0002 – Project 2

1. While the MNIST dataset is perhaps the most frequently utilized dataset in ML courses, FashionMNIST is considered much more challenging. Explain why the samples in the dataset you used in this exercise seem harder to classify than the numerical identification tasks.
2. Utilizing the template attached, provide the calculations for map dimensions, number of weights and number of bias terms for your top performing CNN model.  
     
   Layer Activation Map Dimensions Number of Weights Number of Biases  
   Input 28x28x1 0 0  
   CONV2D (dims)  
   POOL-2  
   CONV2D (dims)  
   POOL-3  
   FC-10
3. Compare the results of your experiments for Part 1, Part 2 and Part 3 – use the values from  
   your recorded model performance to generate at least (3) meaningful figures related to your  
   results.  
   a. Display the results of the test performance for each experiment in a single graph  
   (preferred).  
   b. Provide a table or plot showing how complexity of the model contributed to challenges  
   when training both the FC and CNN implementation. (Did you overfit or stop learning?)  
   4. Discuss in a few sentences the results of your best and worst performing model.  
   a. Were larger networks (structures with more hidden nodes) worth the trade off in  
   training time?  
   b. While the performance between FC and CNN can be large, does the training time and  
   complexity of the CNN seem necessary for this task? Under what circumstances might  
   this change? What if you applied augmentations? Does complexity of a dataset  
   matter?

## Methodology

### Approach

The ultimate purpose of this project was to compare the performance of fully connected vs CNN based topologies. An experimental framework was selected where;

1. Baseline models were constructed where the maximum epoch count and learning rates were explored for both datasets. The reduced test accuracy of the fashion dataset shows the challenge of the additional complexity. See page 2 for heatmaps.
2. Based on the initial model results a target learning rate and epoch count were selected. This selection helped to limit the search space. See pages 3-X for data and diagrams.
3. Using Keras, a convolutional model was constructed. Initially, a small search space was selected using the Keras-Tuner hyperparameter library. This library reduced the code complexity found when iterating over, capturing, and reporting on results. The first execution used a GridSearch to walk through a limited search space.
4. The hyper-parameter search space was then expanded and a BayesianOptimization tuner used.

### Observations

### Fully Connected Networks

As seen from the heatmaps, the default learning rate for the Adam optimizer provided a reasonable performance. It was also found that limiting the epoch count to 7 instead of the proposed 25-50 also provided acceptable results. ***For both datasets, the target minimum accuracy was achieved with the provided topology*, *32/16, Adam optimizer with a learning rate of 0.002***. Further topologies were explored as well though as the results show, the best across the explored the best MNIST accuracy was **96%** and the best Fashion\_MNIST accuracy was **87%**. When the epoch count was increased to 25 epochs an **88%** test accuracy was achieved with a topology of **64/32/16**.

### CNN Results

### Detailed Discussion

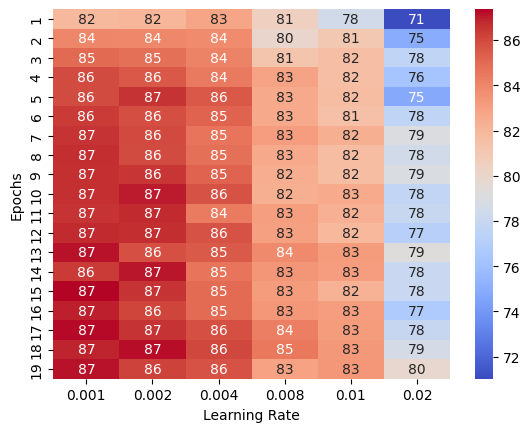
#### Learning Rate vs Epoch count analysis

Heatmaps of achieved accuracy as a function of Epochs and Learning Rate. Two hidden layers were used of 32 and 16 neurons respectively. The Adam optimizer with various learning rates were iterated over to increasing number of epochs. This gives us a view of how many epochs and what learning rate to select for further tests of the topologies.

A screenshot of a graph

Description automatically generated

*MNIST Accuracy vs Learning Rate vs Epochs*



*Fashion\_MNIST Accuracy vs Learning Rate vs Epochs*

#### Accuracy for MNIST across hyper-parameters (7 epochs)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Accuracy | Topology | Activations | Optimizer | LR | Mmt | WD |
| 96.2 | [784, 64, 32, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 95.7 | [784, 32, 16, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0.001 |
| 95.7 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 95.54 | [784, 32, 16, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 95.42 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 95.22 | [784, 64, 32, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0.001 |
| 95.06 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 94.82 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | adam | 0.002 | 0 | 0.001 |
| 94.78 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 94.7 | [784, 28, 28, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0.001 |
| 94.68 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 94.67 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 94.45 | [784, 28, 28, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 94.45 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 94.25 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | adam | 0.002 | 0 | 0.001 |
| 93.59 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 93.11 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 92.82 | [784, 64, 32, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 92.67 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 92.62 | [784, 32, 16, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 92.54 | [784, 64, 32, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0 |
| 92.39 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 92.38 | [784, 28, 28, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0 |
| 92.31 | [784, 64, 32, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0.001 |
| 92.3 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 92.29 | [784, 32, 16, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0 |
| 92.23 | [784, 28, 28, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 92.21 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 92.17 | [784, 28, 28, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 92.15 | [784, 64, 32, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 92.05 | [784, 28, 28, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0.001 |
| 91.92 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0.001 |
| 91.85 | [784, 32, 16, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 91.85 | [784, 32, 16, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0.001 |
| 91.85 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 91.73 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 91.66 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0 |
| 91.18 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0 |
| 91.14 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 89.74 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0.001 |
| 87.46 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 87.1 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 86.4 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 84.18 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 84.08 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 79.35 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 78.43 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 65.26 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 60.36 | [784, 64, 32, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 54.57 | [784, 28, 28, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 53.33 | [784, 64, 32, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 53.1 | [784, 32, 16, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 49.78 | [784, 64, 32, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 49.37 | [784, 28, 28, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 48.77 | [784, 32, 16, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 40.76 | [784, 32, 16, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 39.93 | [784, 64, 32, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0 |
| 39.47 | [784, 28, 28, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 36.08 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 31.89 | [784, 28, 28, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0 |
| 31.58 | [784, 32, 16, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0 |
| 25.84 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 24.27 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 23.81 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 19.82 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 18.52 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 12.81 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 11.7 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 11.37 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0 |
| 11.35 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 11.35 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 11.35 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 11.35 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 11.35 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 11.35 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0 |
| 11.35 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 11.35 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 11.35 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 11.35 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 9.58 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |

#### Accuracy across Fashion\_MNIST across hyper-parameters (7 epochs)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Accuracy | Topology | Activations (Hidden) | Optimizer | LR | Mmnt | WD |
| 87.07 | [784, 64, 32, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0 |
| 86.84 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | adam | 0.002 | 0 | 0 |
| 86.73 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 86.5 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0 |
| 86.49 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 86.28 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | adam | 0.002 | 0 | 0 |
| 86.23 | [784, 32, 16, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0 |
| 86.19 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 86.17 | [784, 64, 32, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 85.87 | [784, 28, 28, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0 |
| 85.79 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 85.74 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0 |
| 85.71 | [784, 64, 32, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0.001 |
| 85.65 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 85.6 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0 |
| 85.54 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 85.41 | [784, 32, 16, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 85.22 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0 |
| 85.22 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | adam | 0.002 | 0 | 0.001 |
| 85.2 | [784, 28, 28, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 85.2 | [784, 28, 28, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0.001 |
| 85.19 | [784, 32, 16, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 85.11 | [784, 32, 16, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0.001 |
| 85.04 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 84.98 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | adam | 0.002 | 0 | 0.001 |
| 84.95 | [784, 64, 32, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 84.95 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 84.86 | [784, 28, 28, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 84.76 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 84.19 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 84 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 84 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 83.02 | [784, 64, 32, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 82.99 | [784, 64, 32, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 82.91 | [784, 64, 32, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0 |
| 82.86 | [784, 28, 28, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0 |
| 82.85 | [784, 32, 16, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 82.8 | [784, 32, 16, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 82.77 | [784, 28, 28, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 82.72 | [784, 32, 16, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0 |
| 82.72 | [784, 28, 28, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 82.33 | [784, 32, 16, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0.001 |
| 82.27 | [784, 64, 32, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0.001 |
| 82.18 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 82.01 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 81.66 | [784, 28, 28, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0.001 |
| 81.35 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0 |
| 81.32 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 81.24 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 81.24 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 81.14 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0 |
| 80.02 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 79.73 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 78.41 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 76.58 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0.001 |
| 75.35 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 74.09 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0.001 |
| 73.93 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 73.4 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 72.12 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 71.57 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 71.21 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 66.92 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 66.45 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 60.62 | [784, 64, 32, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 52.58 | [784, 64, 32, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0 |
| 52.51 | [784, 64, 32, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 52.38 | [784, 32, 16, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0 |
| 51.62 | [784, 28, 28, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 51.53 | [784, 32, 16, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 51.23 | [784, 32, 16, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 50.73 | [784, 28, 28, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0 |
| 48.96 | [784, 28, 28, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 45.31 | [784, 64, 32, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 41.48 | [784, 28, 28, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 40.34 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 39.68 | [784, 32, 16, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 33.34 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 31.34 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 30.98 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 30.18 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 30.03 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 27.57 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0 |
| 27.5 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 26.98 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 26.96 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 26.78 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 23.48 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 21.16 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 19.61 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 17.62 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 15.66 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 10.36 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 10 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 10 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0 |
| 10 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |

#### Accuracy across Fashion\_MNIST across hyper-parameters (25 epochs)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Accuracy | Topology | Activations (Hidden) | Optimizer | LR | Mmnt | WD |
| 88.05 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 87.99 | [784, 64, 32, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0 |
| 87.73 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | adam | 0.002 | 0 | 0 |
| 87.5 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0 |
| 87.17 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | adam | 0.002 | 0 | 0 |
| 87.13 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 86.99 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 86.94 | [784, 64, 32, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 86.76 | [784, 28, 28, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 86.76 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0 |
| 86.68 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 86.64 | [784, 32, 16, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0 |
| 86.64 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0 |
| 86.64 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0 |
| 86.54 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 86.52 | [784, 28, 28, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 86.4 | [784, 28, 28, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0 |
| 86.39 | [784, 32, 16, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 86.36 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 86.29 | [784, 64, 32, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0 |
| 86.28 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0.001 |
| 86.27 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0 |
| 86.25 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 86.24 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | adam | 0.002 | 0 | 0.001 |
| 86.22 | [784, 64, 32, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 86.18 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | adam | 0.002 | 0 | 0.001 |
| 86.16 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 86.05 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | adam | 0.002 | 0 | 0 |
| 85.95 | [784, 64, 32, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0.001 |
| 85.91 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 85.88 | [784, 32, 16, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0 |
| 85.87 | [784, 28, 28, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0 |
| 85.76 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0 |
| 85.73 | [784, 28, 28, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0.001 |
| 85.7 | [784, 32, 16, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 85.67 | [784, 32, 16, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 85.66 | [784, 28, 28, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 85.59 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 85.57 | [784, 64, 32, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 85.53 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 85.52 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 85.42 | [784, 32, 16, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 85.37 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 85.34 | [784, 28, 28, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0.001 |
| 85.34 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 85.3 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 85.23 | [784, 32, 16, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0.001 |
| 85.21 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 85.2 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 85.13 | [784, 32, 16, 10] | [sigmoid, relu] | adam | 0.002 | 0 | 0.001 |
| 85.12 | [784, 64, 32, 10] | [sigmoid, relu] | asgd | 0.02 | 0 | 0.001 |
| 84.98 | [784, 64, 32, 10] | [sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 84.92 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 84.84 | [784, 28, 28, 10] | [sigmoid, tanh] | adam | 0.002 | 0 | 0.001 |
| 84.73 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 84.48 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 84.18 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 84.11 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0 |
| 83.9 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | adam | 0.002 | 0 | 0.001 |
| 83.64 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0 |
| 82.2 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | asgd | 0.02 | 0 | 0.001 |
| 81.19 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 80.7 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | asgd | 0.02 | 0 | 0.001 |
| 74.39 | [784, 64, 32, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 73.94 | [784, 32, 16, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 73.73 | [784, 64, 32, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 73.14 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | asgd | 0.02 | 0 | 0.001 |
| 73.06 | [784, 32, 16, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 72.92 | [784, 64, 32, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0 |
| 72.61 | [784, 64, 32, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 71.83 | [784, 28, 28, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 71.82 | [784, 28, 28, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 71.79 | [784, 28, 28, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0 |
| 71.31 | [784, 28, 28, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 70.67 | [784, 32, 16, 10] | [sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 69.71 | [784, 32, 16, 10] | [sigmoid, relu] | sgd | 0.001 | 0.1 | 0 |
| 66.64 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 64.48 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 62.82 | [784, 28, 28, 28, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 61.12 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 57.66 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 55.12 | [784, 64, 32, 16, 10] | [relu, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 54.65 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0 |
| 52.17 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 43.98 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 41.89 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0.001 |
| 38.66 | [784, 64, 32, 16, 8, 10] | [relu, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 35.59 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 32.12 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 30.15 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 29.03 | [784, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 28.88 | [784, 28, 28, 28, 28, 10] | [relu, sigmoid, tanh, relu] | sgd | 0.001 | 0.1 | 0 |
| 27.58 | [784, 64, 32, 16, 10] | [sigmoid, sigmoid, tanh] | sgd | 0.001 | 0.1 | 0 |
| 26.06 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0 |
| 25.43 | [784, 28, 28, 28, 28, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |
| 22.06 | [784, 64, 32, 16, 8, 10] | [sigmoid, sigmoid, tanh, tanh] | sgd | 0.001 | 0.1 | 0.001 |