CSCE 479/879 Homework 1: Connected Architectures and Fashion MNIST

Derek DeBlieck, Sabrina Fowler, Grace Hecke, Abby Veiman

December 5, 2025

Abstract

The problem for this assignment is to classify the Fashion MNIST data set. We want to find the best architecture for classifying these images. We also want to get some practice training and comparing models with different hyperparameters. Ultimately, the 8 trials we ran resulted in about the same level of accuracy on the test set. Also, our models with regularizers performed worse than their unregularized counterparts which suggests that an L2 regularizer is not optimal for this task.

1 Introduction

For this assignment, we developed various architectures to classify the Fashion MNIST dataset. We created two main architecture structures, the first had one hidden layer with 128 nodes and the second had two hidden layers with 256 and 128 nodes respectively. We also varied our learning rates, trying 0.001 and 0.0005. We also tested the efficacy of an L2 regularizer. We tried all 8 combinations of these three binary options.

Overall, our models performed relatively similarly well, with accuracy on the test set ranging from 0.8669 to 0.8877. In general, the models using L2 regularization performed slightly less well than their unregularized counterparts.

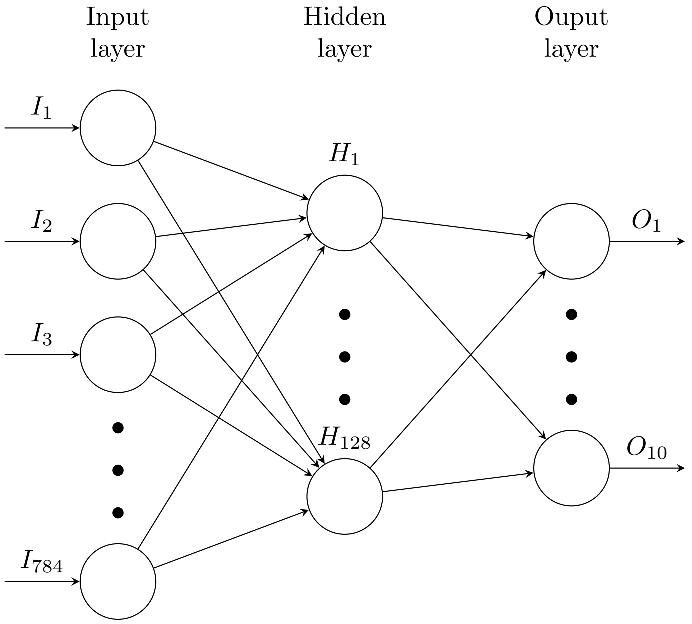
2 Problem Description

We were tasked with training a deep learning model to classify images from the fashion MNIST dataset. The data consists of images of various fashion items, labeled as t-shirt/top, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, and ankle boot. The data came from tensorflow-datasets, the training set consists of 60,000 images and the test set contains 10,000 images. In particular, the data comes to us as 28×28 arrays with pixel intensities from 0-255.

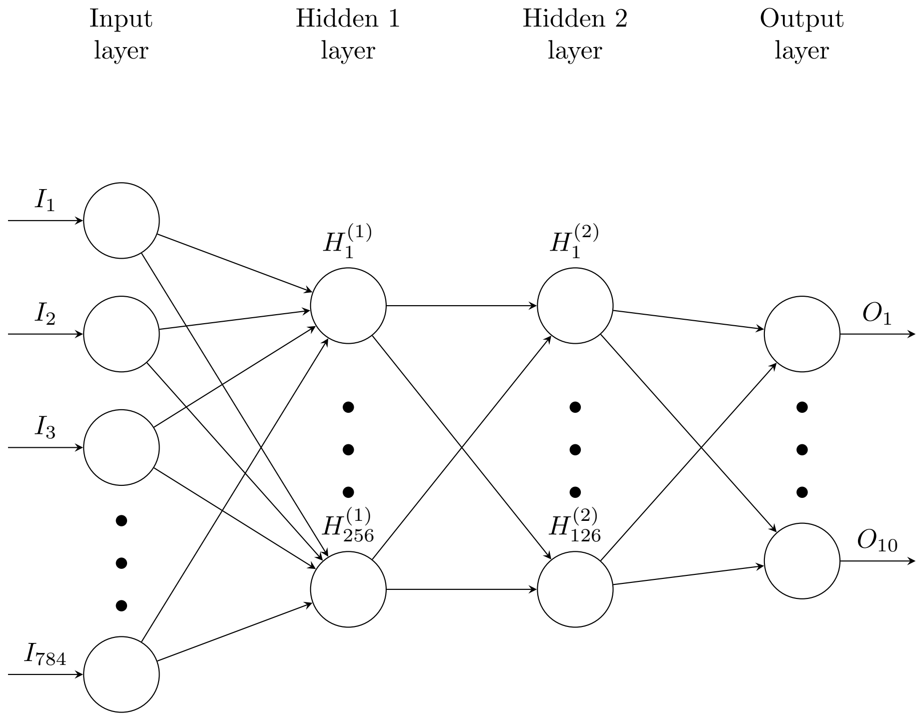
This problem is interesting because it is similar to a problem we have worked on before in our Hackathons, classifying the handwritten digit MNIST data. However these images are more complicated than the digit images which should lead to a more complex model being necessary.

3 Approaches

Our approach to this problem involved setting up a general structure for our models and then varying the number of hidden layers, learning rates, and use of regularizer. Our general structure involved rescaling the pixel intensities to fall between 0 and 1 and flattening the 28×28 arrays into a 1×784 vectors. Then if we were using a regularizer that happened at this stage. Next, these were fed into our hidden layer(s), and then to a softmax layer with 10 nodes that served as our output layer. We also used a patience parameter of 3 to perform early stopping based on the validation loss. See below for figures demonstrating our neural network architectures.



Architecture 1 with only a single hidden layer



Architecture 2 with a two hidden layers

4 Experimental Setup

We used the fashion MNIST dataset; we flattened each 28x28 image into a single vector with 784 features. We tested two different neural network architectures: the first contained only a single hidden layer with 128 nodes, while the second consisted of two hidden layers. The first of these hidden layers contained 256 nodes while the second contained 128 nodes. The Adam optimizer was used during model training, which used categorical cross-entropy as a loss function. The Adam optimizer was used during training. Additionally, we employed early stopping with a patient interval of 3 and a batch size of 32. For each architecture, we trained models using a learning rate of 0.001 and 0.0005 additionally trained a model using L_2-regularization and a model without regularization. This gives us a total of 8 trained models (2 architectures, L_2-regularization or no regularization, and 2 learning rates).

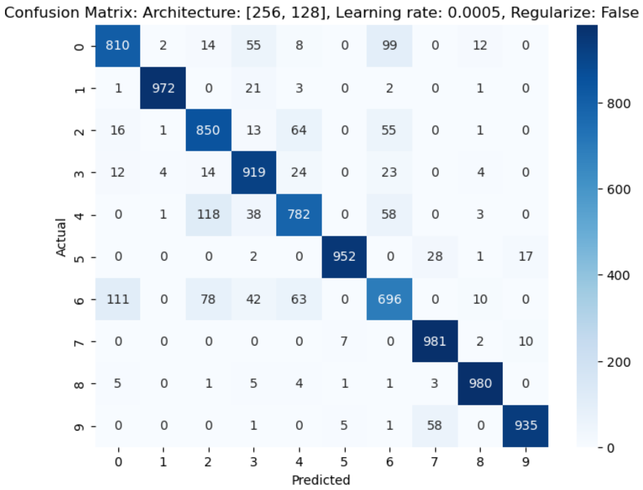
5 Experimental Results

Please see the table below for a summary of results. Based on test accuracy, we found that the architecture with two hidden layers, a smaller learning rate, and no L_2-regularization was most accurate with the smallest generalization error. However, the difference between these models was relatively small. All models performed reasonably well, which we should expect for a relatively simple classification problem.

Table 1: Summary of Model Architectures and their Performance

Hidden Layers	Learning Rate	Regularizer	Validation Accuracy	Test Accuracy	Generalization Error
128	0.001	n/a	0.8938	0.8832	0.117 ± 0.006 (95% CI)
128	0.001	L2	0.8800	0.8675	0.132 ± 0.007 (95% CI)
128	0.0005	n/a	0.9010	0.8840	0.116 ± 0.006 (95% CI)
128	0.0005	L2	0.8873	0.8695	0.131 ± 0.007 (95% CI)
256, 128	0.001	n/a	0.8940	0.8840	0.116 ± 0.006 (95% CI)
256, 128	0.001	L2	0.8802	0.8669	0.133 ± 0.007 (95% CI)
256, 128	0.0005	n/a	0.8953	0.8877	0.112 ± 0.006 (95% CI)
256, 128	0.0005	L2	0.8868	0.8716	0.128 ± 0.007 (95% CI)

Below, please see the confusion matrix for the best performing model (2 hidden layers, no L_2-regularization, and a learning rate = 0.0005)


Confusion Matrix: Architecture: [256, 128], Learning rate: 0.0005, Regularize: False

	0	1	2	3	4	5	6	7	8	9
0	810	2	14	55	8	0	99	0	12	0
1	1	972	0	21	3	0	2	0	1	0
2	16	1	850	13	64	0	55	0	1	0
3	12	4	14	919	24	0	23	0	4	0
4	0	1	118	38	782	0	58	0	3	0
5	0	0	0	2	0	952	0	28	1	17
6	111	0	78	42	63	0	696	0	10	0
7	0	0	0	0	0	7	0	981	2	10
8	5	0	1	5	4	1	1	3	980	0
9	0	0	0	1	0	5	1	58	0	935

6 Discussion

All models performed similarly on the test dataset, attaining an accuracy between 85% and 88.8%. This is probably because, by modern standards, classifying the fashion MNIST dataset is considered a relatively easy classification problem. We surmise that with a more complex dataset there would be a more noticable difference between the two main architectures we tested (that is, between a single hidden layer and two layers). Surprisingly, L_2-regularization appeared to decrease model performance; all architectures with L_2-regularization performed worse than their non-regularized counterparts. Decreasing the learning rate from 0.001 toe 0.0005 made a relatively small difference. However, in general it also increased model performance slightly, and thus if training time is not too constrained we would suggest putting a model with a training rate of 0.0005 into production (or even testing slower learning rates).

7 Conclusions

We trained and evaluated 8 different models to classify the fashion MNIST dataset. While all of these models performed relatively similarly, the best performance was achieved by using 2 hidden layers, no L_2-regularization, and a learning rate of 0.0005. Future work could include a longer patient interval, a slower learning rate, or even adding additional hidden layers to the neural network architecture.

Table 2: Contributions by team member for this assignment.

Team Member	**Contribution**
Grace Hecke	Code and LaTeX Review
Sabrina Fowler	Abstract, Sections 1-3
Derek DeBlieck	3-7
Abby Veiman	Code and LaTeX Review