Name(s): Won Woo Lyu, Albert Yoo
NetID(s): wlyu2, alberty3
Team name on Kaggle leaderboard: WWL

For each of the sections below, your reported test accuracy should approximately match the accuracy reported on Kaggle.

Briefly describe the hyperparameter tuning strategies you used in this assignment. Then record your optimal hyperparameters and test/val performance for the four different network types.

Two-layer Network Trained with SGD

Batch size	Learning Rate	Hidden Layer Size	Regularization coefficient	Highest Validation Accuracy
50	1e-2	100	0.1	0.535
100	1e-2	100	0.1	0.537
150	1e-2	100	0.1	0.533
200	1e-2	100	0.1	0.51

Batch size	Learning Rate	Hidden Layer Size	Regularization coefficient	Highest Validation Accuracy
100	1e-1	100	0.1	0.542
100	1e-2	100	0.1	0.537
100	1e-3	100	0.1	0.474
100	1e-4	100	0.1	0.372

Batch size	Learning Rate	Hidden Layer Size	Regularization	Highest

			coefficient	Validation Accuracy
100	1e-1	30	0.1	0.515
100	1e-1	60	0.1	0.531
100	1e-1	90	0.1	0.547
100	1e-1	120	0.1	0.553
100	1e-1	150	0.1	0.556
100	1e-1	180	0.1	0.55

Batch size	Learning Rate	Hidden Layer Size	Regularization coefficient	Highest Validation Accuracy
100	1e-1	150	0	0.547
100	1e-1	150	0.0001	0.544
100	1e-1	150	0.001	0.558
100	1e-1	150	0.01	0.531
100	1e-1	150	0.1	0.556

Best hyperparameters (if you changed any of the other default hyperparameters like initialization method, etc. please note that as well):

We set the learning rate to 1e-2, hidden layer size to 100, and regularization coefficient to 0.1 and tried batch sizes which ranged from 50 to 200 in increments of 50. We found that the batch size of 100 gave the highest validation accuracy. With the batch size fixed at 100, we tried learning rates 1e-1, 1e-2, 1e-3, and 1e-4 and found that the learning rate of 1e-1 gave the highest validation accuracy. With the batch size fixed at 100 and the learning rate fixed at 1e-1, we tried hidden layer sizes which ranged from 30 to 180 in increments of 30. We found that the hidden layer size of 150 gave the highest validation accuracy. With the batch size fixed at 100, the learning rate fixed at 1e-1, and the hidden layer size fixed at 150, we tried regularization coefficients 0, 0.0001, 0.001, 0.01, and 0.1. We found that the regularization coefficient of 0.001 gave the highest validation accuracy.

Based on what we tried, the 2sgd neural network with a batch size of 100, a learning rate of 1e-1, a hidden layer size of 150, and regularization coefficient of 0.001 gave the highest validation accuracy.

Batch size:	100
Learning rate:	1e-1
Hidden layer size:	150
Regularization coefficient:	0.001

Record the results for your best hyperparameter setting below:

Validation accuracy:	0.558
Test accuracy:	0.50000

Three-layer Network Trained with SGD

Best hyperparameters (if you changed any of the other default hyperparameters like initialization method, etc. please note that as well):

Since we found out optimized parameters that lead to the highest accuracy, we are using same parameter for three-layer SGD.

Based

Batch size:	100
Learning rate:	1e-1
Hidden layer size:	150
Regularization coefficient:	0.001

Record the results for your best hyperparameter setting below:

Validation accuracy:	0.545
Test accuracy:	0.50770

Two-layer Network Trained with Adam

Best hyperparameters (if you changed any of the other default hyperparameters like initialization method, etc. please note that as well):

Using the same parameters as SGD, we had to adjust the learning rate down to 1e-2 and 1e-3 because an error occurred when the learning rate of 1e-1 was used. The 2 layer Adam network had a higher validation accuracy when the learning rate was 1e-3 then when the learning rate was 1e-2.

Record the results for your best hyperparameter setting below:

Batch size:	100
Learning rate:	1e-3
Hidden layer size:	150
Regularization coefficient:	0.001
β_1	0.9
β_2	0.999

Validation accuracy:	0.55
Test accuracy:	0.52840

Three-layer Network Trained with Adam

Best hyperparameters (if you changed any of the other default hyperparameters like initialization method, etc. please note that as well):

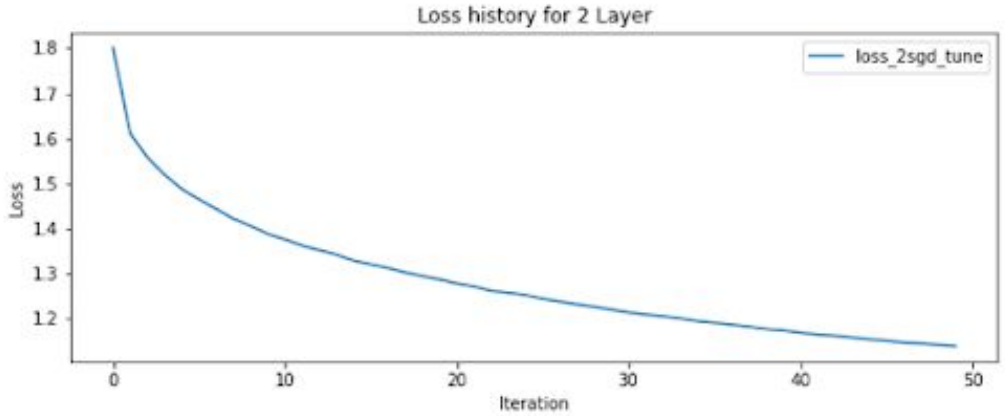
Batch size:	100
Learning rate:	1e-3
Hidden layer size:	150
Regularization coefficient:	0.001
β_1	0.9
β_2	0.999

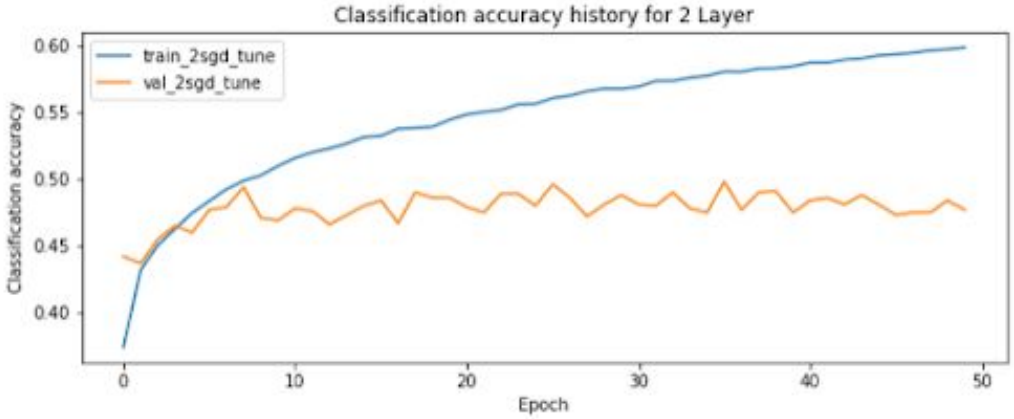
Record the results for your best hyperparameter setting below:

Validation accuracy:	0.553
Test accuracy:	0.53320

Hyperparameter Tuning

In the pdf of Ipython notebook, overall graphs for each sgd and adam show that there is a large gap between training and validation accuracy. This tells that there is an overfitting due to a large model without much regularization. To check this, we modified 2 layers of sgd to not overfit by changing the hidden_size to 25 and regularization to 0.1. Below is new graph after tuning, and there is much less gap between train accuracy and valid accuracy.





Comparison of SGD and Adam

Attach two plots, one of the training loss for each epoch and one of the validation accuracy for each epoch. Both plots should have a line for SGD and Adam. Be sure to add a title, axis labels, and a legend.

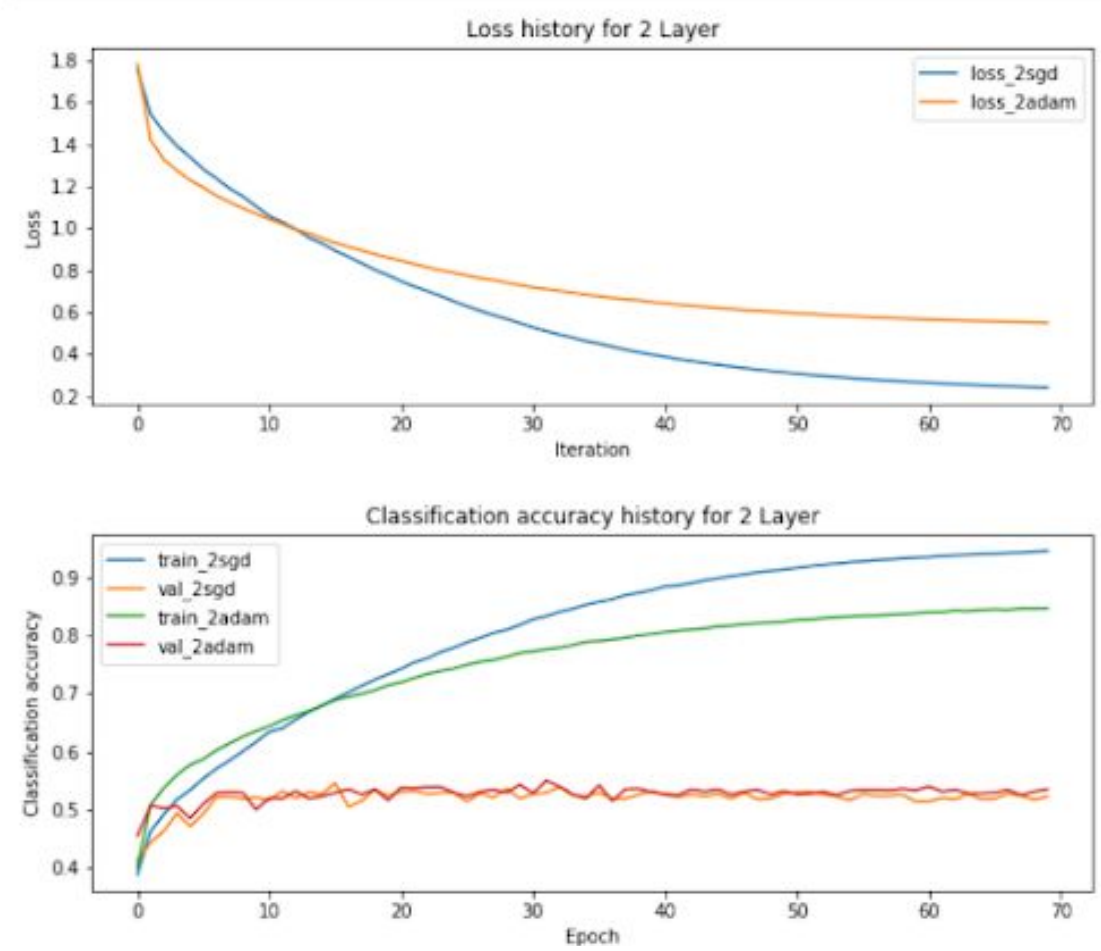


Figure 1. 2 Layer SGD and Adam

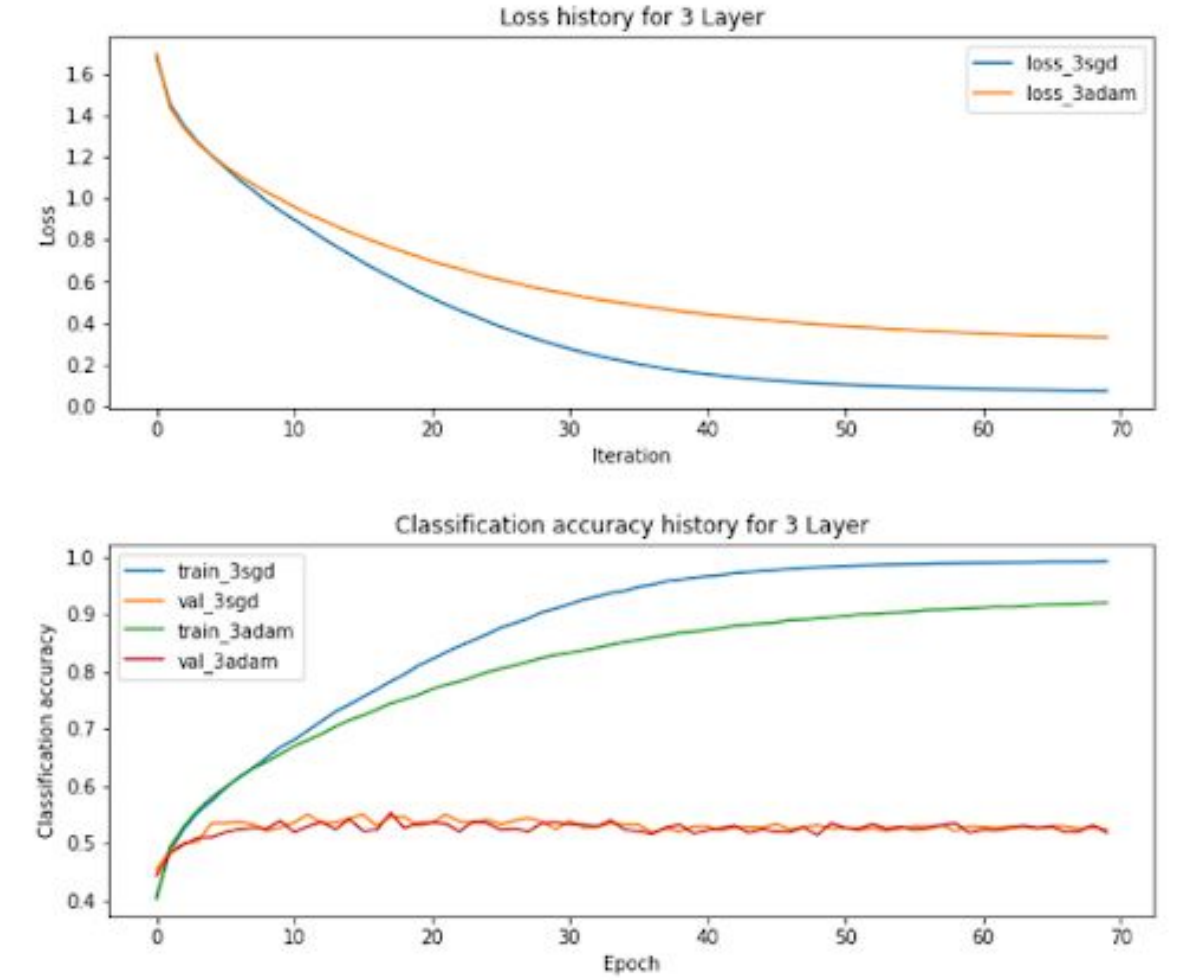


Figure 2. 3 Layer SGD and Adam

Compare the performance of SGD and Adam on training times and convergence rates. Do you notice any difference? Note any other interesting behavior you observed as well.

Interesting part is that SGD has much less loss compared with adam; however, the valid accuracy seems to be similar between the two of them. Also, the Adam network takes more epochs to reduce its loss and increase its training accuracy compared to the SGD network. Having lower train accuracy by looking at the graph, Adam still had similar or higher test accuracy that can be seen from the graph and our data.