

Name(s):Yunxiao

NetID(s): YUNXIAO9

Team name on Kaggle leaderboard:

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

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:	0.001
Hidden layer size:	150
Regularization coefficient:	0.1

Record the results for your best hyperparameter setting below:

Validation accuracy:	0.8803
Test accuracy:	0.8744

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):

Batch size:	50
Learning rate:	0.001
Hidden layer size:	128
Regularization coefficient:	0.1

Record the results for your best hyperparameter setting below:

Validation accuracy:	0.8885
Test accuracy:	0.8805

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):

Batch size:	50
Learning rate:	0.001
Hidden layer size:	128
Regularization coefficient:	0.1
β_1	default
β_2	default

Record the results for your best hyperparameter setting below:

Validation accuracy:	0.8844
Test accuracy:	0.8812

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:	128
Learning rate:	0.001
Hidden layer size:	128
Regularization coefficient:	0.1
β_1	default
β_2	default

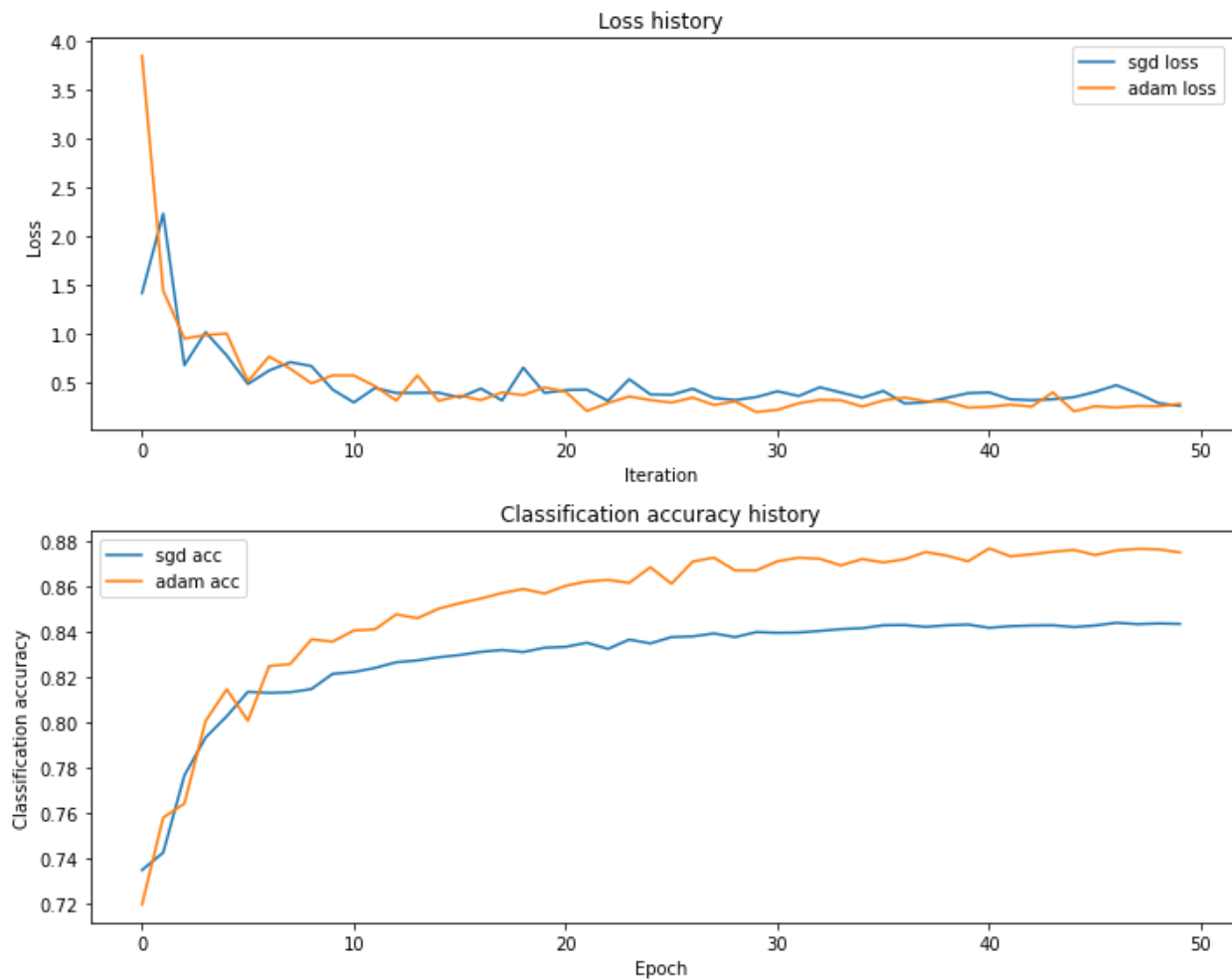
Record the results for your best hyperparameter setting below:

Validation accuracy:	0.8867
----------------------	--------

Test accuracy:	0.8826
----------------	--------

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.



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.

The adam loss diminishes faster than sgd at the beginning, and adam actually has a better fit with more epochs, if tested on validation data.