Name(s): Yunfan Yang
NetID(s): yunfany3
Team name on Kaggle leaderboard: yunfany3

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):
I've adjusted the batch size from 200 to 10, learning rate from 1e-3 to 1e-2, Hidden layer size from 20 to 60, and Regularization coefficient from 0.01 to 0.0001

Batch size:	batch_size = 20
Learning rate:	learning_rate = 1.4e-3
Hidden layer size:	hidden_size = 60
Regularization coefficient:	regularization = 0.0001

Record the results for your best hyperparameter setting below:

Validation accuracy:	0.93078
Test accuracy:	0.867

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):
I've adjusted the batch size from 200 to 10, learning rate from 1e-3 to 1e-2, Hidden layer size from 20 to 60, and Regularization coefficient from 0.01 to 0.0001

Batch size:	batch_size = 20

Learning rate:	learning_rate = 2e-3
Hidden layer size:	hidden_size = 60
Regularization coefficient:	regularization = 0.0001

Record the results for your best hyperparameter setting below:

Validation accuracy:	0.93088
Test accuracy:	0.8729

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):
I've adjusted the batch size from 200 to 10, learning rate from 1e-3 to 1e-2, Hidden layer size from 20 to 60, and Regularization coefficient from 0.01 to 0.0001

Batch size:	20
Learning rate:	0.001
Hidden layer size:	60
Regularization coefficient:	0.0001
β_1	0.9
β_2	0.999

Record the results for your best hyperparameter setting below:

Validation accuracy:	0.8582
Test accuracy:	0.8592

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):
I've adjusted the batch size from 200 to 10, learning rate from 1e-3 to 1e-2, Hidden layer size from 20 to 60, and Regularization coefficient from 0.01 to 0.0001

Batch size:	20
Learning rate:	0.001
Hidden layer size:	60
Regularization coefficient:	0.0001
β_1	0.9
β_2	0.999

Record the results for your best hyperparameter setting below:

Validation accuracy:	0.8663
Test accuracy:	0.8602

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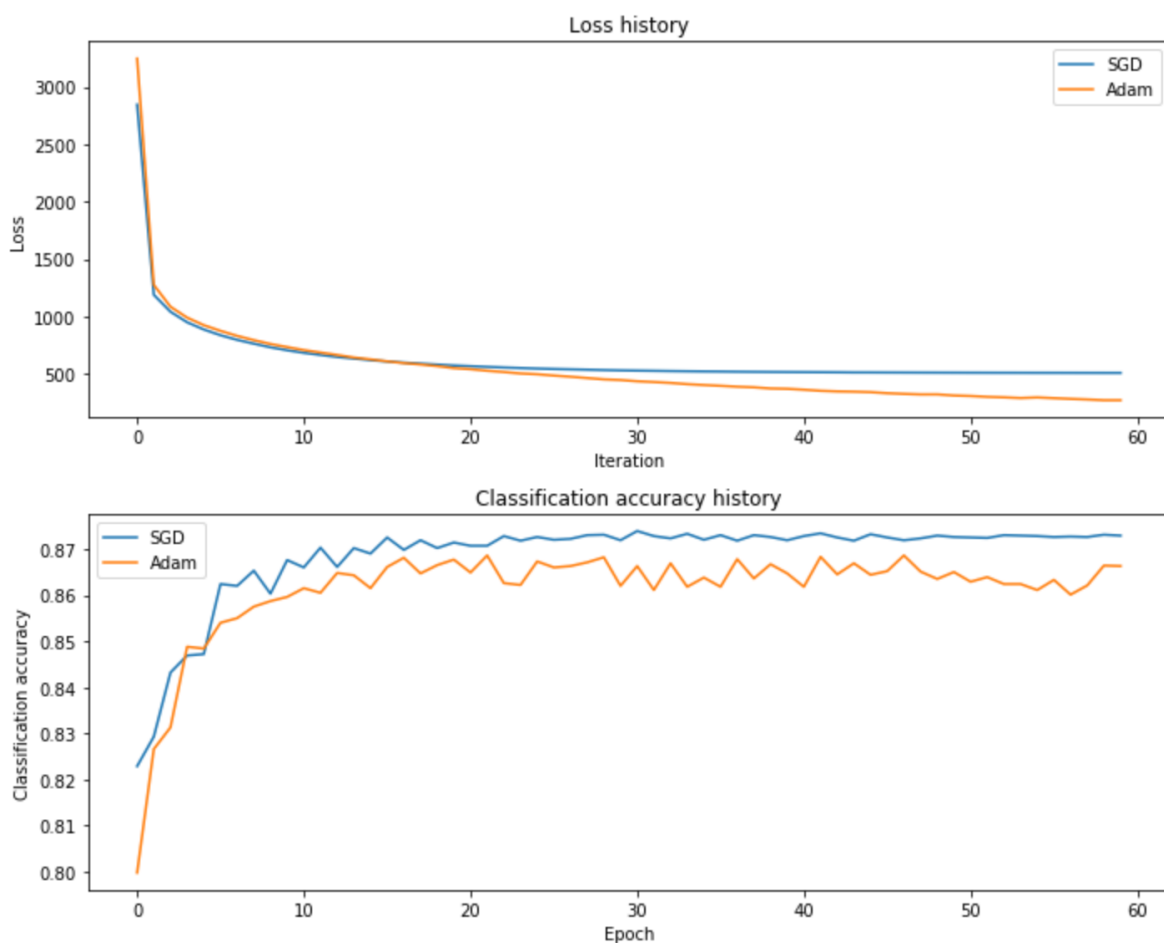
Compare the performance of SGD and Adam on training times and convergence rates. Do you notice any difference? Note any other interesting behavior you observed as well.

Looks like accuracy using Adam is much more unstable than accuracy using SGD. Besides, SGD is faster in training time and converges a little bit faster. However Adam achieves a lower loss finally.