

**Name(s): Chen-Kai Chang**

**NetID(s): ckchang4**

**Team name on Kaggle leaderboard: Chen-Kai Chang**

**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.*

I found that the learning rate and regularization coefficient affected the validation accuracy most so I focus mainly on them. I also added learning rate decay. Smaller batch size also increases accuracy.

### **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:	50
Learning rate:	0.001(with lr decay 0.9 per epoch)
Hidden layer size:	80
Regularization coefficient:	0.08

*Record the results for your best hyperparameter setting below:*

Validation accuracy:	0.8837
Test accuracy:	0.8810

### **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:	0.0005(with lr decay 0.95 per epoch)
Hidden layer size:	80
Regularization coefficient:	0.08

*Record the results for your best hyperparameter setting below:*

Validation accuracy:	0.8767
Test accuracy:	0.8704

### **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:	0.0001(with 0.95 lr decay every epoch)
Hidden layer size:	80
Regularization coefficient:	0.1
$\beta_1$	0.9
$\beta_2$	0.999

*Record the results for your best hyperparameter setting below:*

Validation accuracy:	0.8879
Test accuracy:	0.8837

### **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:	50
Learning rate:	0.005 (with 0.95 lr decay per epoch)
Hidden layer size:	60
Regularization coefficient:	0.03
$\beta_1$	0.9
$\beta_2$	0.99

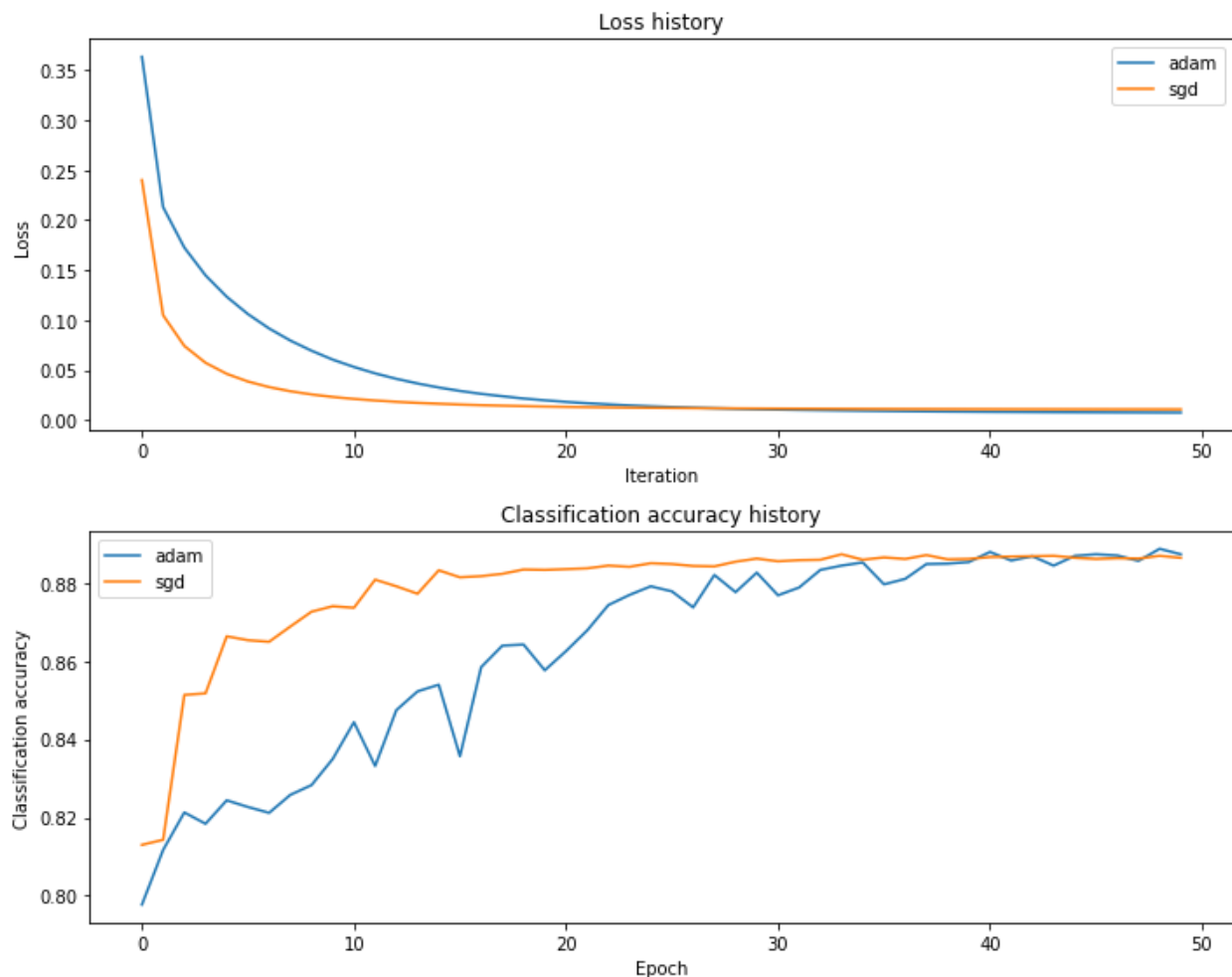
*Record the results for your best hyperparameter setting below:*

Validation accuracy:	0.8901
Test accuracy:	0.8831

## 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.



Adam seems to converge slower than finely tuned SGD, but the final accuracy is higher for Adam. SGD kind of plateaued after 30 epochs.