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 manually tuned parameters. For batch size I tried 100 and 200. For learning rate I tried 0.1, 0.05, 0.01 and found 0.1 to be the best. For hidden layer size, I found that neither too low or two high would be good, and also the optimal hidden layer size is smaller for Sigmod than Relu. For regularization coefficient, I found 0.001 to be the best because the size of the model was large so the regularization coefficients should be relatively small.

Two-layer Relu Activation Network

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

Batch size: 200	
Learning rate: 0.1	
Hidden layer size: 100	
Regularization coefficient: 0.001	

Record the results for your best hyperparameter setting below:

Validation accuracy: 53.5	
Test accuracy: 52.4	

Two-layer Sigmoid Activation Network

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

Batch size: 200	
Learning rate: 0.1	
Hidden layer size: 75	
Regularization coefficient: 0.001	

Record the results for your best hyperparameter setting below:

Validation accuracy: 45.5	
Test accuracy: 46.86	

Three-layer Relu Activation Network

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

Batch size: 200	
Learning rate: 0.1	
Hidden layer size: 100	
Regularization coefficient: 0.001	

Record the results for your best hyperparameter setting below:

Validation accuracy: 53.5	
Test accuracy: 53.1	

Three-layer Sigmoid Activation Network

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

Batch size: 200	
Learning rate: 0.1	
Hidden layer size: 110	
Regularization coefficient: 0.01	

Record the results for your best hyperparameter setting below:

Validation accuracy: 45.5	

| Test accuracy: 45.6 | |