CS7015-Deep Learning

Programming Assignment 1

Jeshuren Chelladurai (CS17M017)
Ajith Kumar M (CS17M009)

February 17, 2018

Note: All the experiments in the report are performed using the following settings,(unless specified)

- Loss Function: Cross Entropy
- Activation function in the hidden layers: Sigmoid
- Initial Learning Rate: 0.02
- Batch-size: 20
- Optimizer: ADAM
- Annealing schedule: Halve the learning rate every two epochs
- # epochs: 20
- Regularization : L2

1. Single Hidden Layer Performance

1.1 Training Data

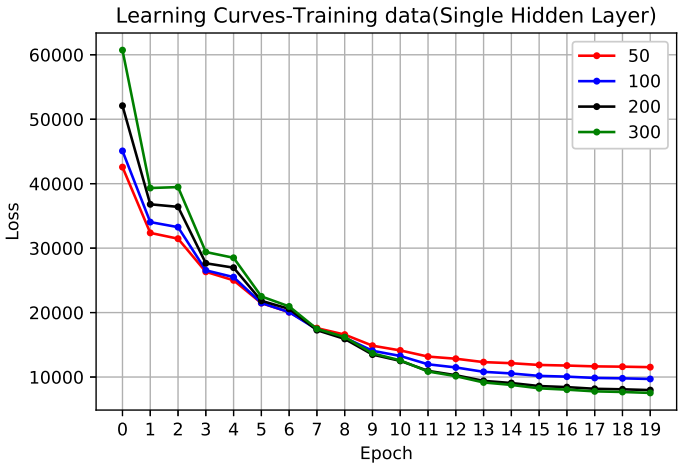

Figure 1: Single hidden layer learning curves for training data over different hidden layer sizes

1.2 Validation data

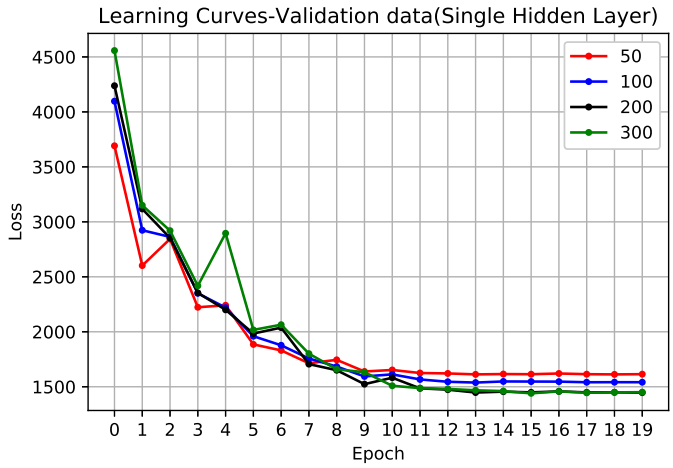

Figure 2: Single hidden layer learning curves for validation data over different hidden layer sizes

2. Two Hidden Layers Performance

2.1 Training Data

Figure 3: Two hidden layers learning curves for training data over different hidden layer sizes

2.2 Validation data

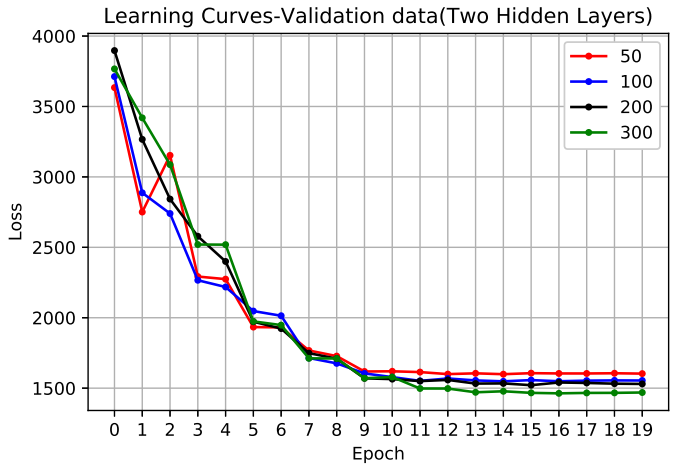

Figure 4: Two hidden layers learning curves for validation data over different hidden layer sizes

3. Three Hidden Layers Performance

3.1 Training Data

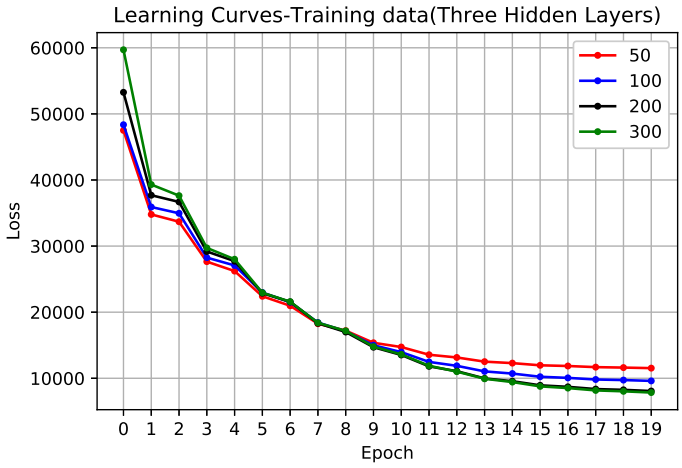

Figure 5: Three hidden layers learning curves for training data over different hidden layer sizes

3.2 Validation data

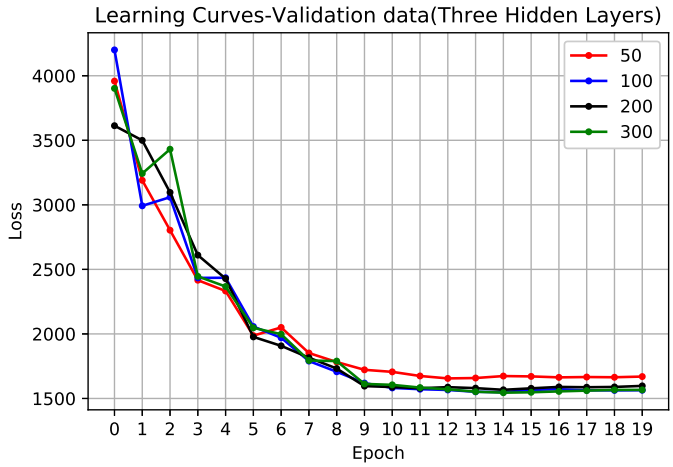

Figure 6: Three hidden layers learning curves for validation data over different hidden layer sizes

4. Four Hidden Layers Performance

4.1 Training Data

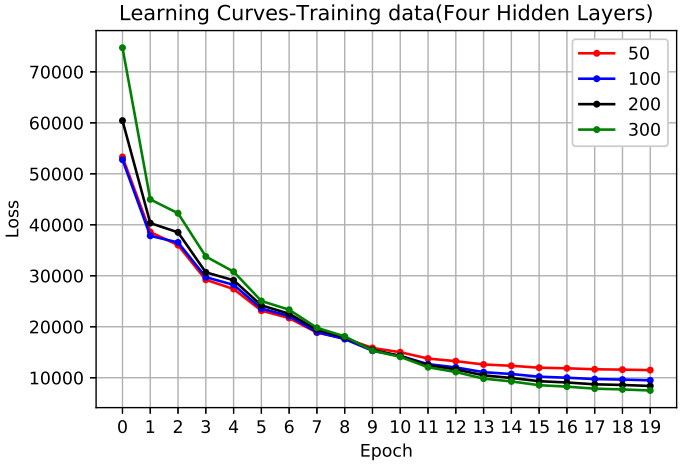

Figure 7: Four hidden layers learning curves for training data over different hidden layer sizes

4.2 Validation data

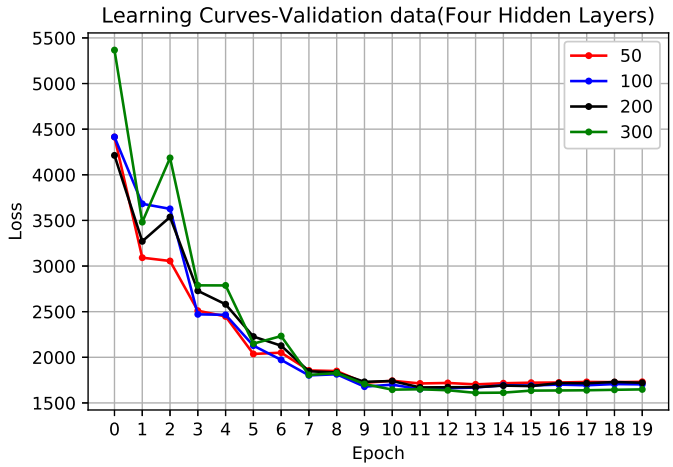

Figure 8: Four hidden layers learning curves for validation data over different hidden layer sizes

5. Performance Of Different Optimizers

This experiment was carried with 3 hidden layers and each layer having 300 neurons.

5.1 Training Data

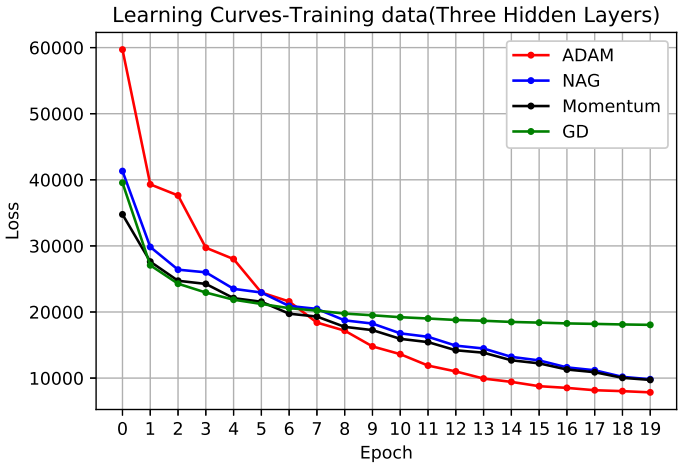

Figure 9: learning curves for training data over different optimization algorithms

5.2 Validation data

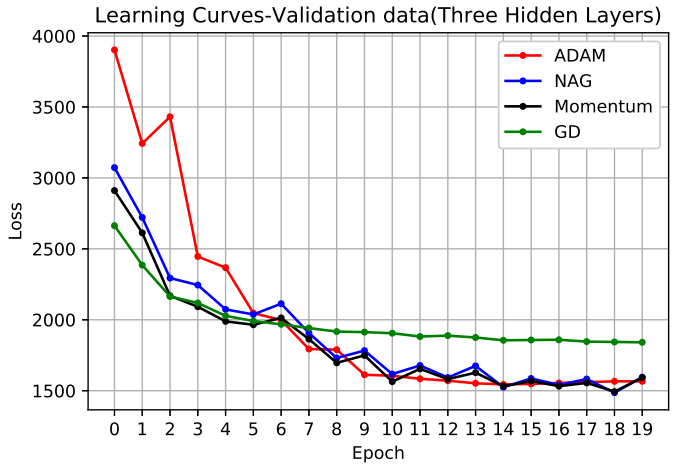

Figure 10: learning curves for validation data over different optimization algorithms

6. Performance Of Different Activation Functions

This experiment was carried with 2 hidden layers and each layer having 100 neurons.

6.1 Training Data

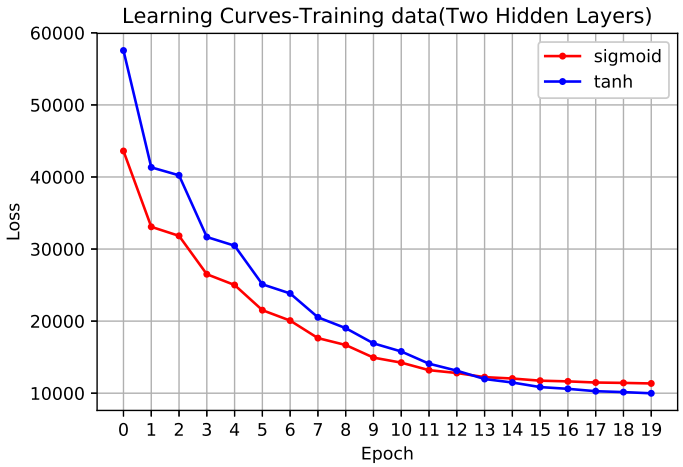

Figure 11: learning curves for training data over different activation functions

6.2 Validation data

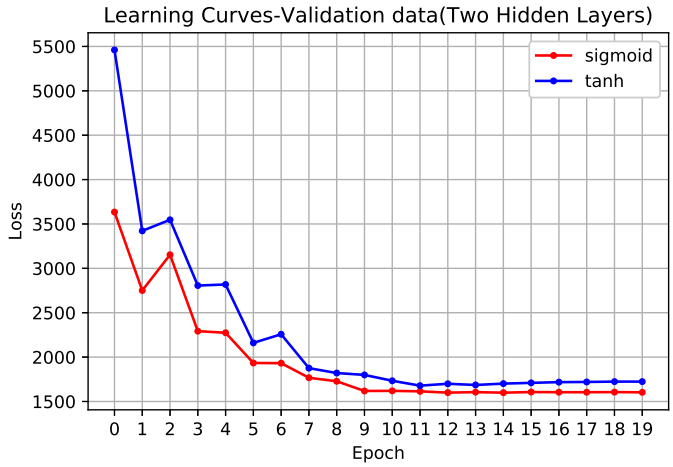

Figure 12: learning curves for validation data over different activation functions

7. Performance Of Different Loss Functions

This experiment was carried with 2 hidden layers and each layer having 100 neurons.

7.1 Training Data

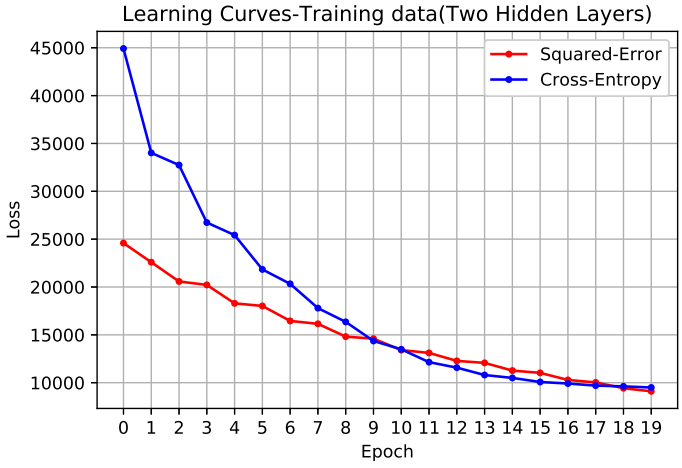

Figure 13: learning curves for training data over different loss functions

7.2 Validation data

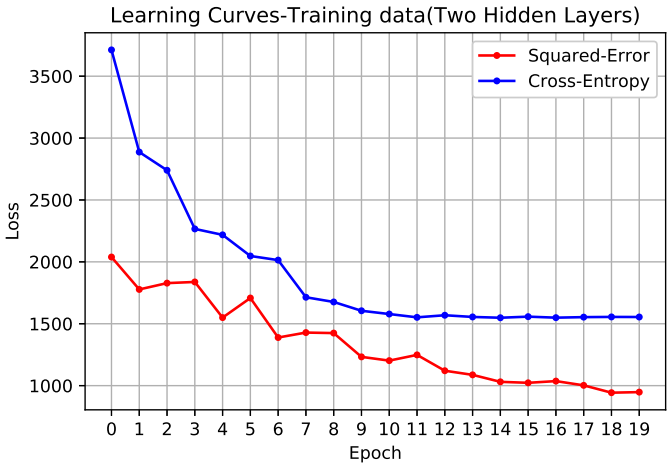

Figure 14: learning curves for validation data over different loss functions

8. Performance Of Different Batch Sizes

This experiment was carried with 2 hidden layers and each layer having 100 neurons.

8.1 Training Data

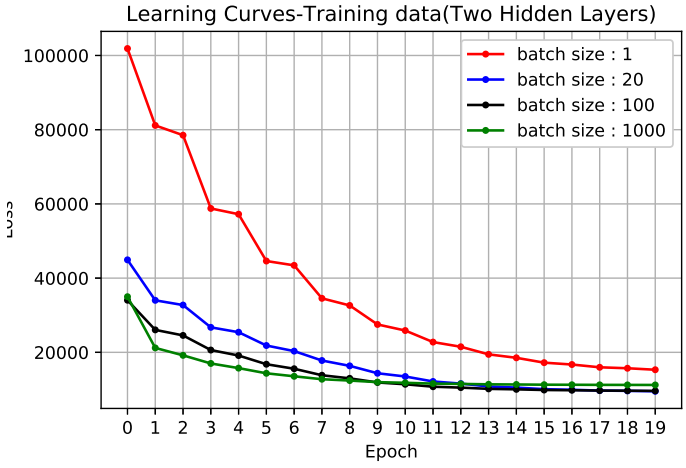

Figure 15: learning curves for training data over different batch sizes

8.2 Validation data

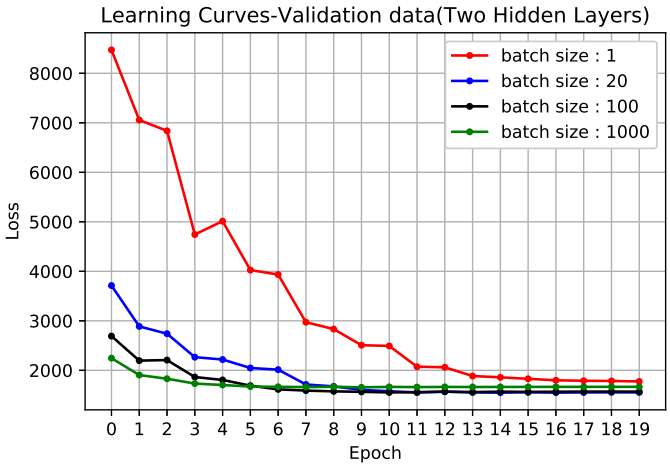

Figure 16: learning curves for validation data over different batch sizes