

# CS7015 : DEEP LEARNING

## PROGRAMMING ASSIGNMENT 1

*Team Members :*

Madhura Pande  
(CS17S031)  
Aakriti Budhraj  
(CS18S009)

*Course Instructor:*

Dr. Mitesh Khapra  
Dept. of CSE, Indian  
Institute of Technology,  
Madras

March 2, 2019



# 1 Comparison of different number of hidden layers and neurons(Q1-Q4)

## 1.1 One Hidden layer

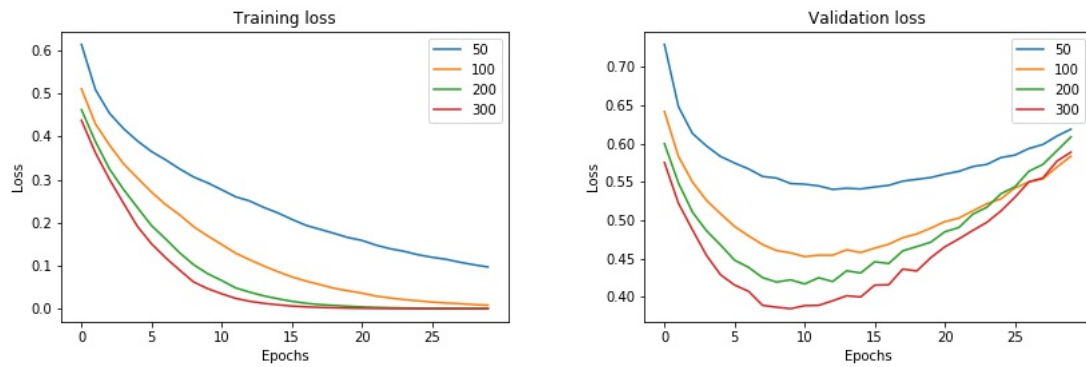


Figure 1: Loss vs Epochs curves for varying no. of neurons in 1 hidden layer

## 1.2 Two Hidden layers

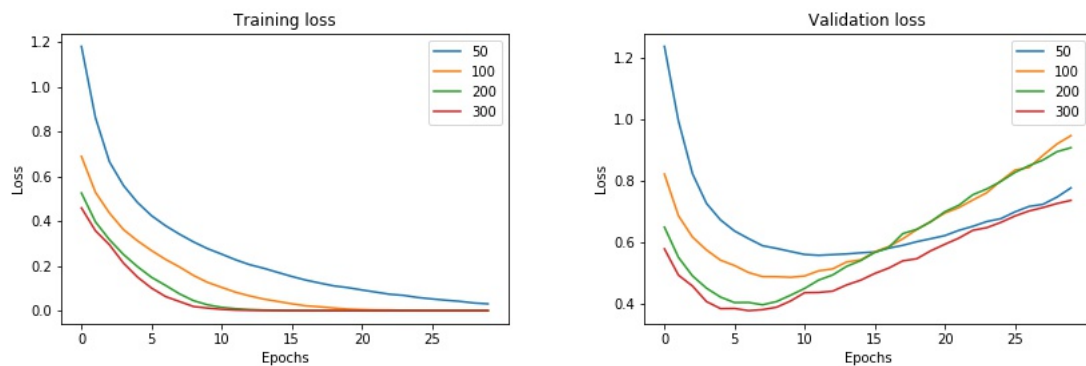


Figure 2: Loss vs Epochs curves for varying no. of neurons in 2 hidden layers

### 1.3 Three Hidden layers

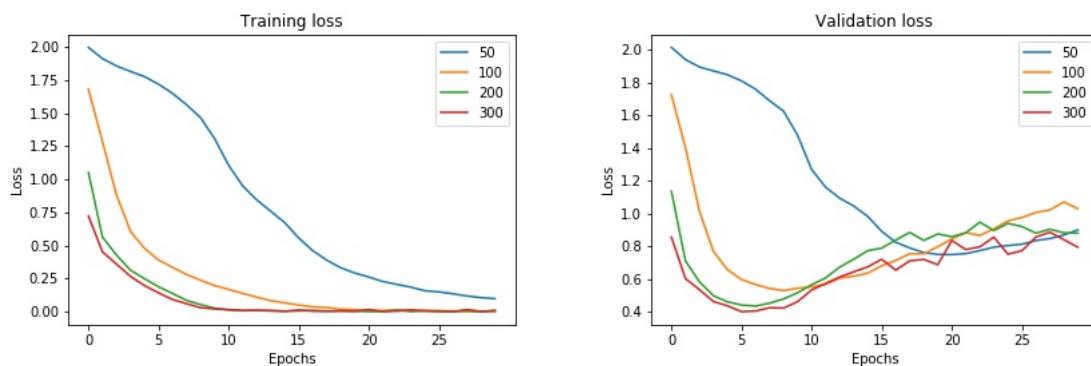


Figure 3: Loss vs Epochs curves for varying no. of neurons in 3 hidden layers

### 1.4 Four Hidden layers

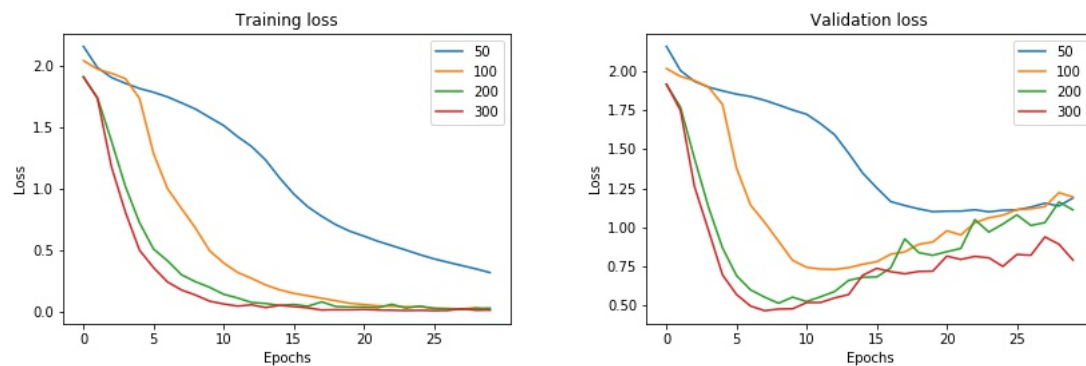


Figure 4: Loss vs Epochs curves for varying no. of neurons in 4 hidden layers

## 2 Comparing various optimization algorithms

### (GD, Momentum, NAG, Adam)

Train loss and validation loss for various Optimizers.

Hidden Layers: 4

Num of Neurons in each layer: 300

Activation: Sigmoid

Loss: Cross Entropy

Batch size: 20

**Observations:** We ran this experiment for 50 epochs and observed that Momentum gradient descent had most oscillations. Nesterov had lesser and Adam had negligible. This can be verified from the loss plot of training as well.

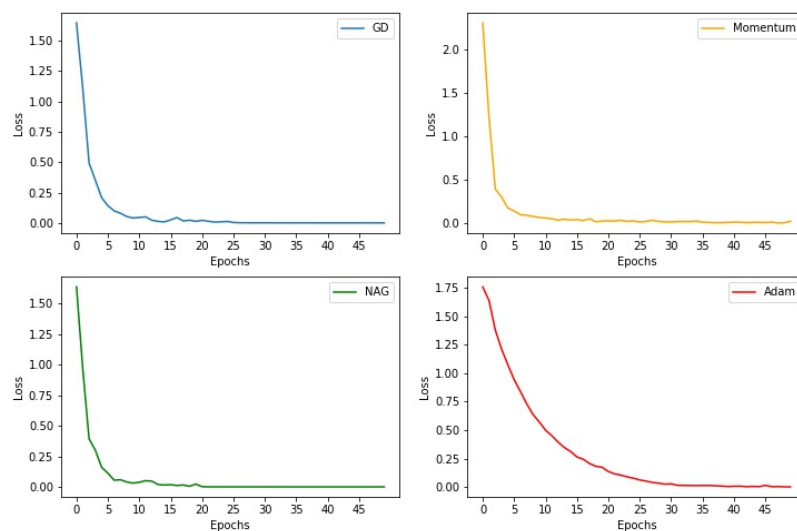


Figure 5: Training Loss vs Epochs curves for different optimizers

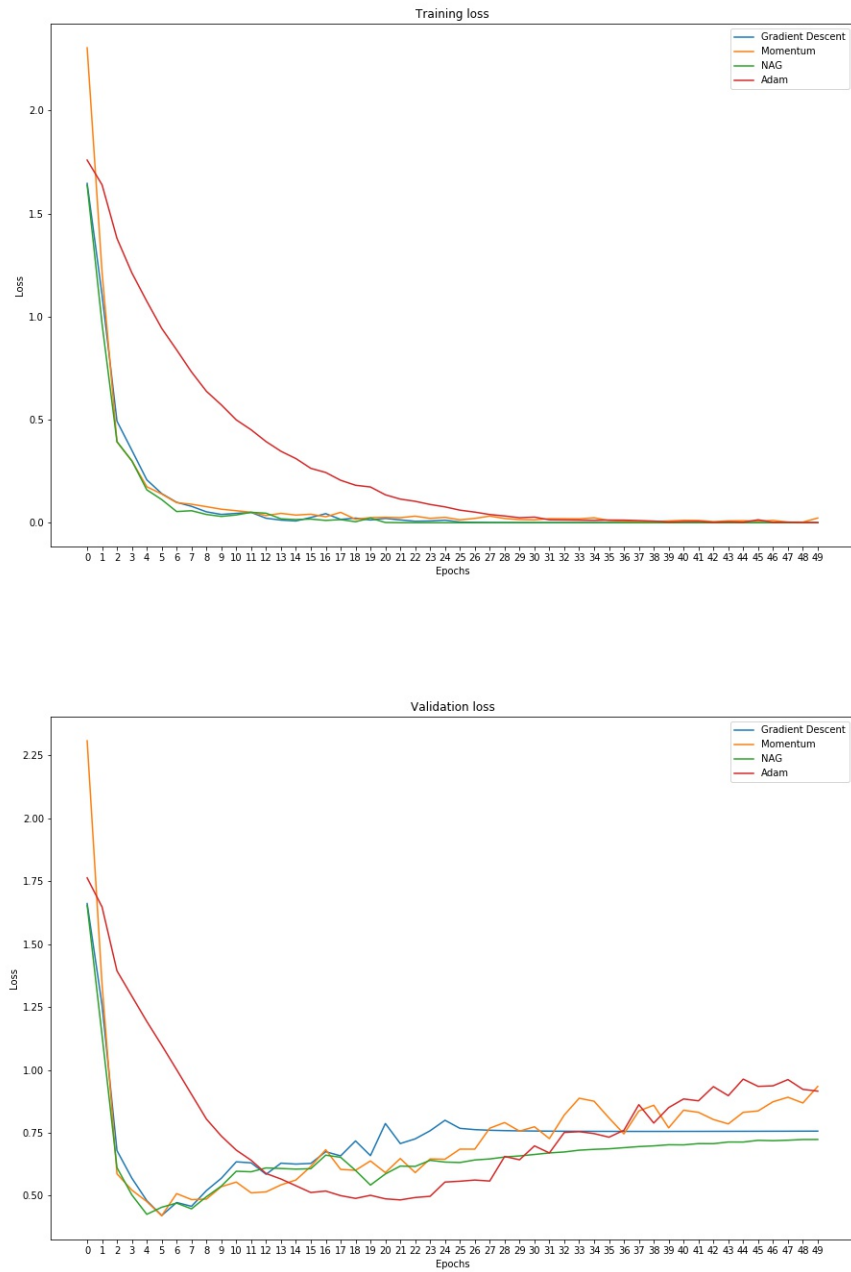


Figure 6: Validation Loss vs Epochs curves for different optimizers

### 3 Comparison of different activation functions

Train loss and validation loss for *sigmoid* and *tanh* activation functions.

Hidden Layers: 2

Num of Neurons in each layer: 100

Loss: Cross Entropy

Batch size: 20

Optimizer: Adam

**Observations:** We ran this experiment for 30 epochs and observed that *tanh* activation function performed better than *sigmoid* function. This can be observed from the Train and Validation loss plots below.

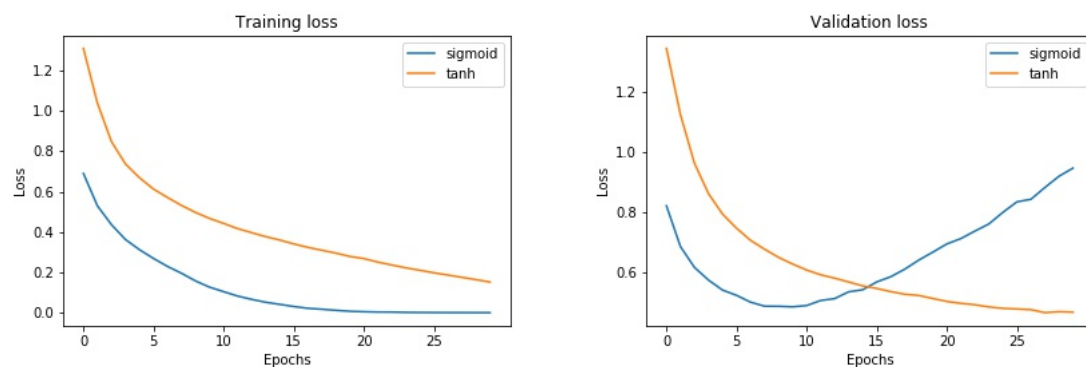


Figure 7: Loss vs Epochs curves for *sigmoid* and *tanh* activation functions

## 4 Comparison of cross entropy and squared error loss functions

Train loss and validation loss for *sigmoid* and *tanh* activation functions.

Hidden Layers: 2

Num of Neurons in each layer: 100

Loss: Cross Entropy

Batch size: 20

Optimizer: Adam

**Observations:** Cross-Entropy loss turned out to be a better choice as evident from the plot as well.

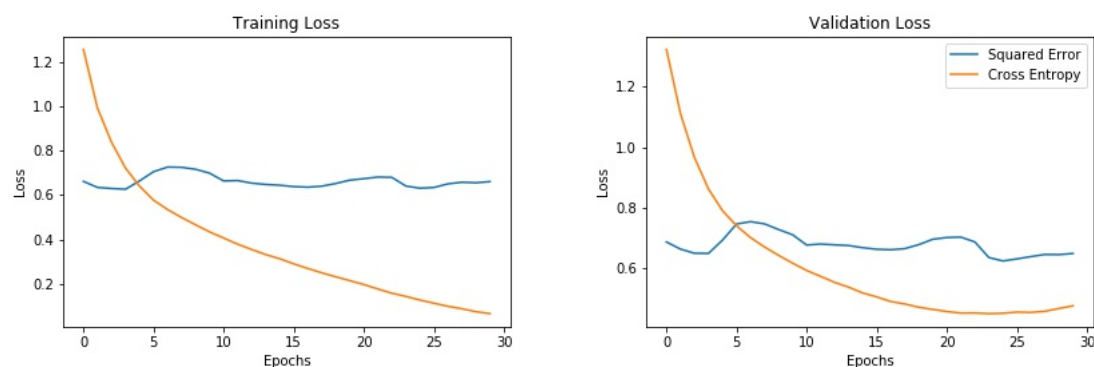


Figure 8: Cross-entropy loss vs Squared error loss

## 5 Comparison of different batch sizes

Below are the train loss and validation loss for various batch sizes.

Hidden Layers: 2

Num of Neurons in each layer: 100

Activation: Sigmoid

Loss: Cross Entropy

Optimizer: Adam

**Observations:** We ran this experiment for 15 epochs and observed that batch sizes 100 and 1000 gave smoother curves as compared to batch sizes

1(Stochastic Gradient Descent) and 20. More the batch size, more closer the gradients are to the true gradient. This can be verified from the loss plot of training as well.

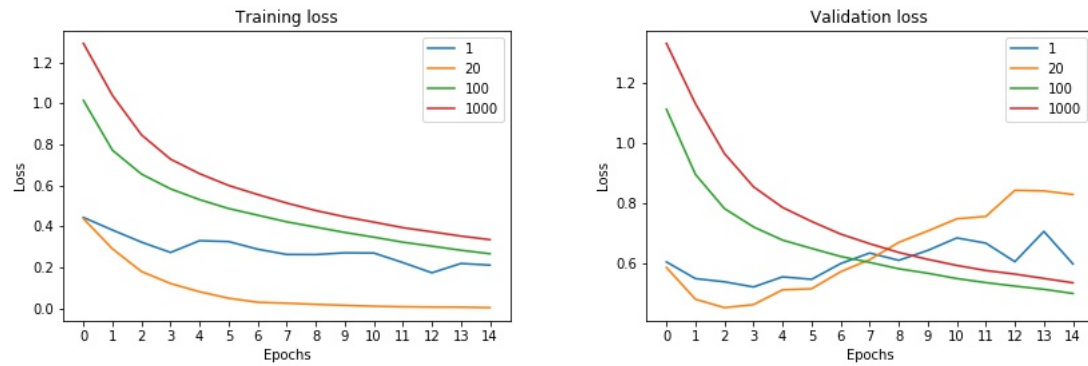


Figure 9: Loss vs Epochs curves for different batch sizes