

# Programming Assignment 1 - Report

Anshul Kumar (ME15B082)

Ashutosh Raj (ME15B086)

2<sup>nd</sup> March 2019

## 1 Answer 1:

Hidden Layers : 1

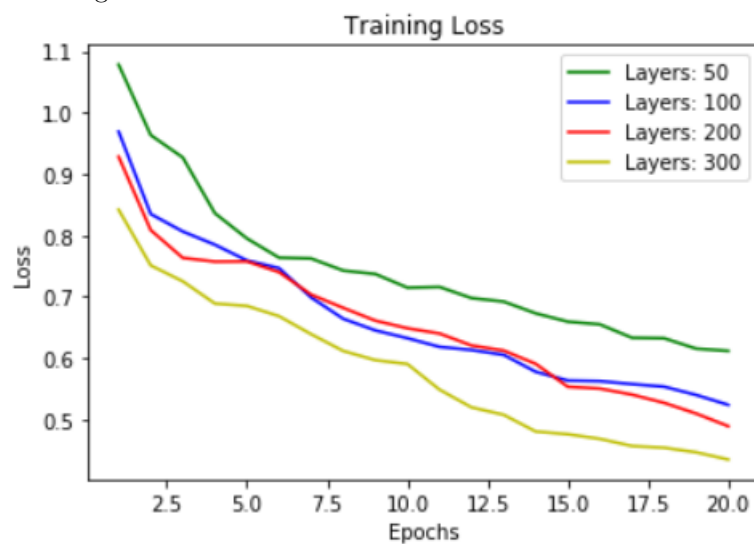
Optimizer : Adam

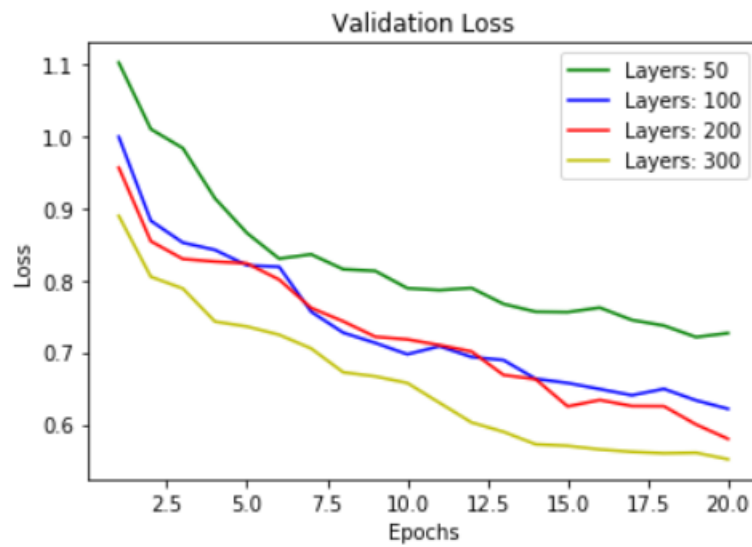
Batch Size : 20

Activation : Sigmoid

Initial Learning Rate : 0.001

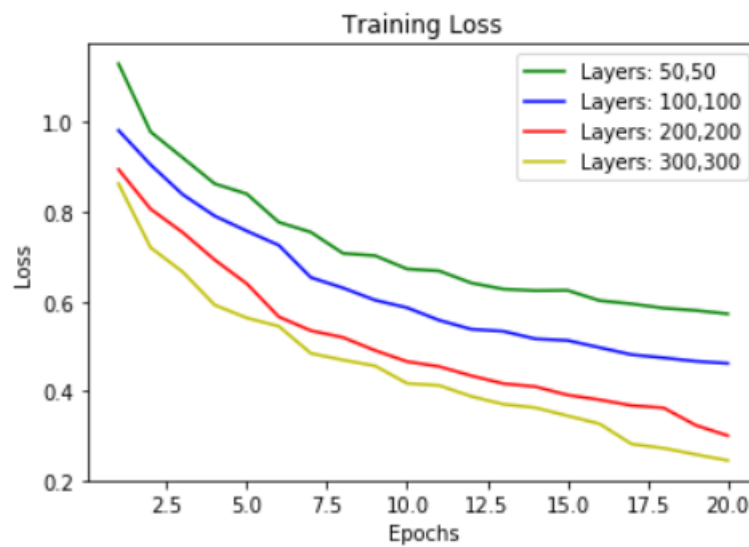
Annealing : True

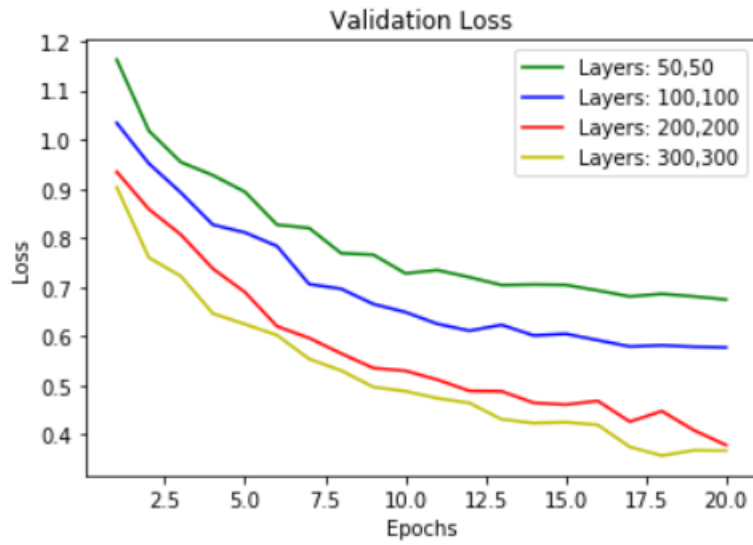




## 2 Answer 2:

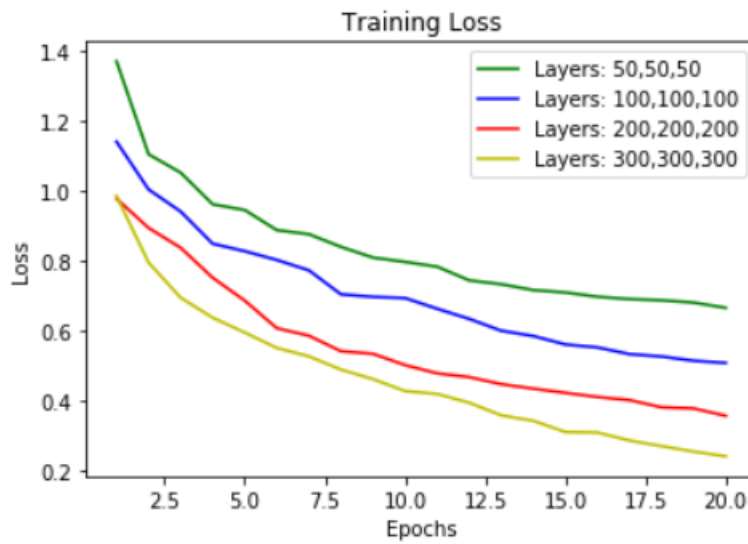
Hidden Layers : 2  
Optimizer : Adam  
Batch Size : 20  
Activation : Sigmoid  
Initial Learning Rate : 0.001  
Annealing : True

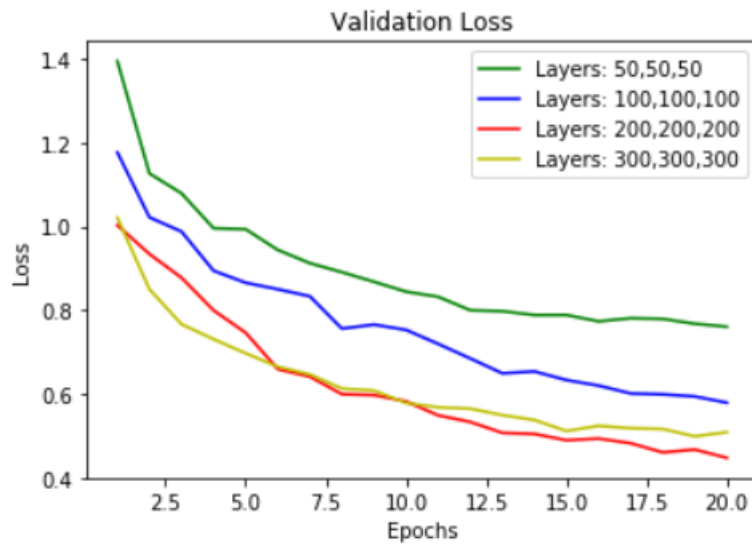




### 3 Answer 3:

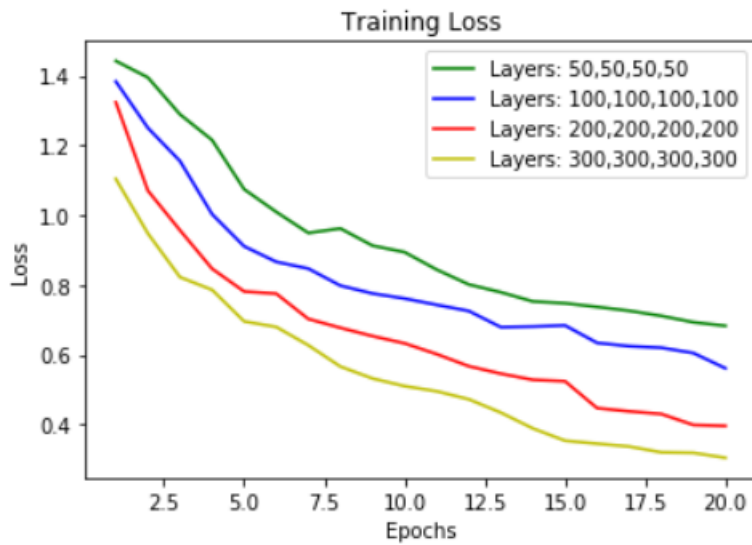
Hidden Layers : 3  
 Optimizer : Adam  
 Batch Size : 20  
 Activation : Sigmoid  
 Initial Learning Rate : 0.001  
 Annealing : True

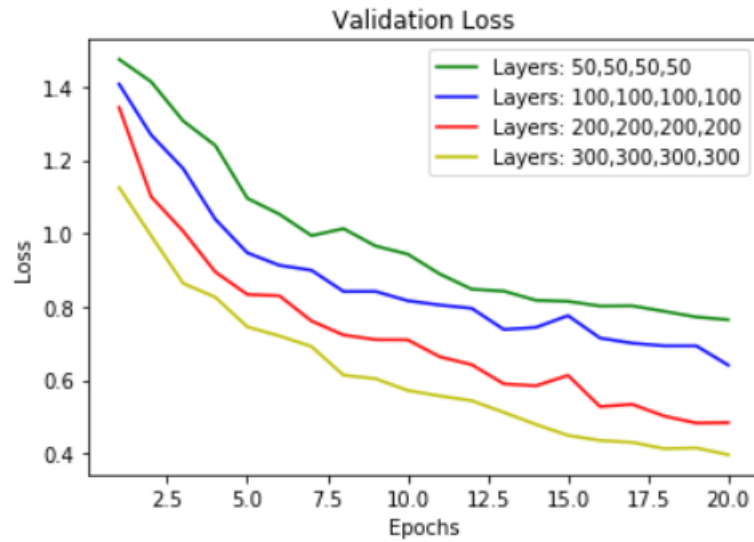




#### 4 Answer 4:

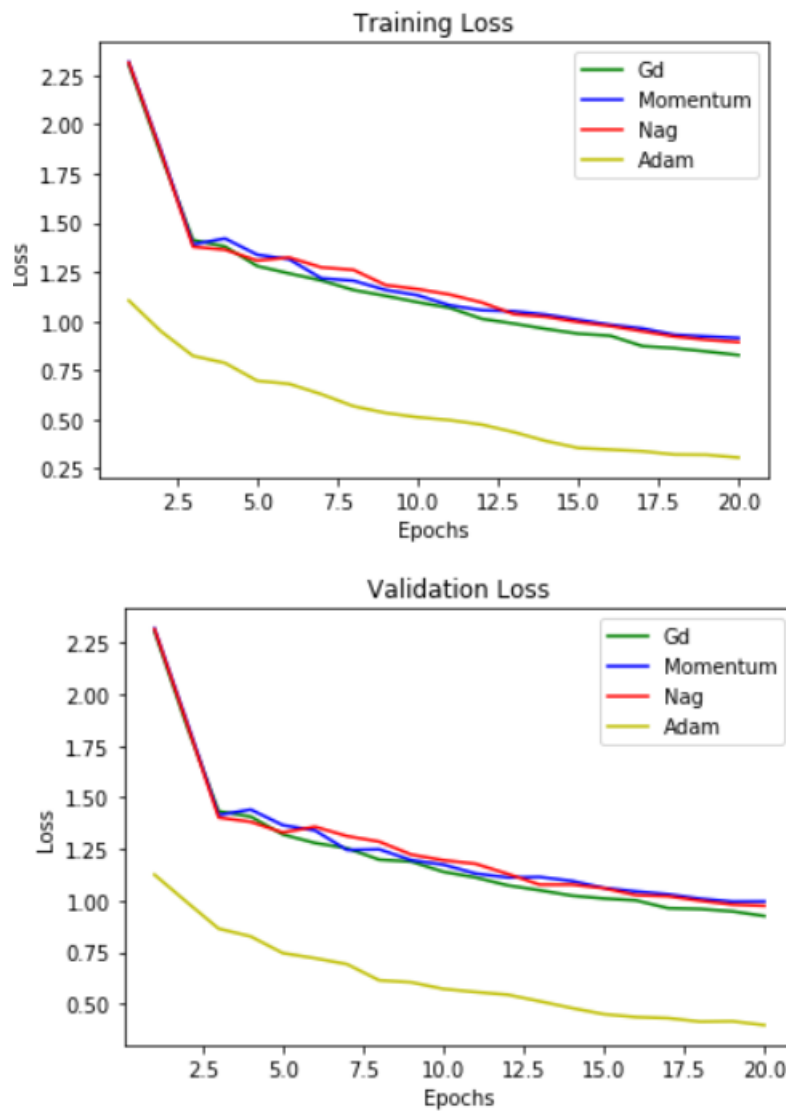
Hidden Layers : 4  
 Optimizer : Adam  
 Batch Size : 20  
 Activation : Sigmoid  
 Initial Learning Rate : 0.001  
 Annealing : True





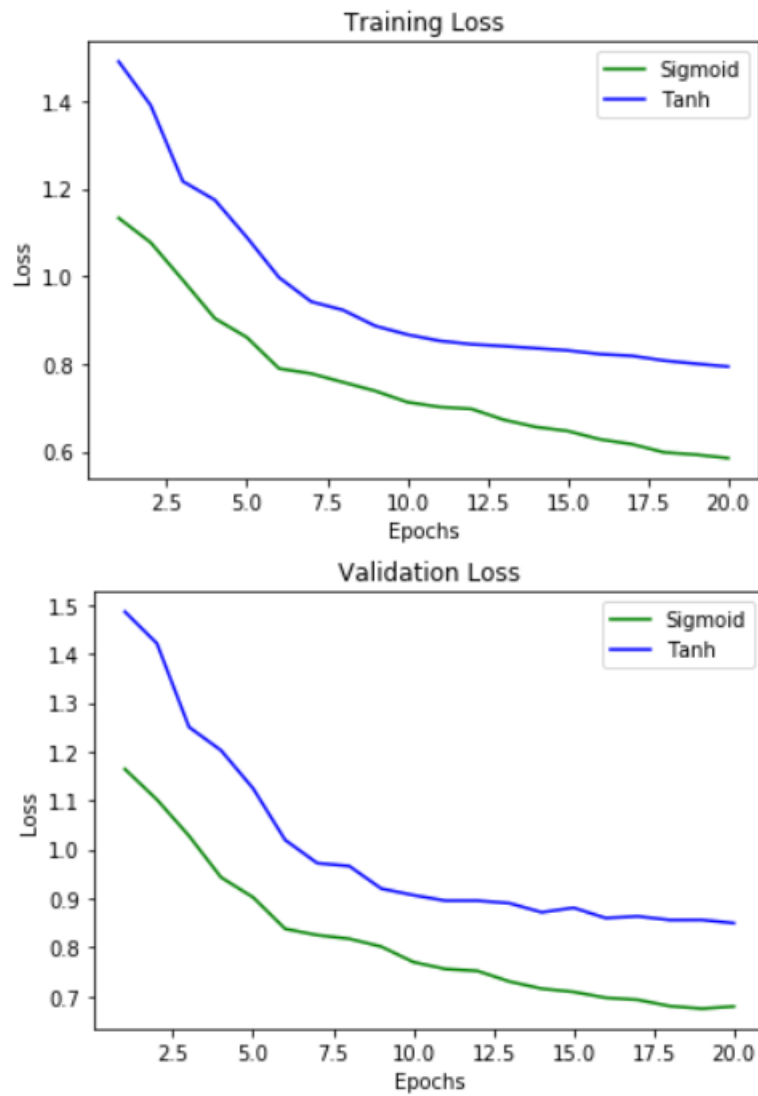
## 5 Answer 5:

Hidden Layers : 4  
Neurons in each layer : 300  
Optimizer : Adam  
Loss : Cross Entropy  
Activation : Sigmoid  
Batch Size : 20  
Initial Learning Rate for Adam : 0.001  
Initial Learning Rate for GD : 0.001  
Initial Learning Rate for NAG : 0.0001  
Gamma NAG : 0.9  
Initial Learning Rate for Momentum : 0.0001  
Gamma Momentum : 0.9  
Annealing : True



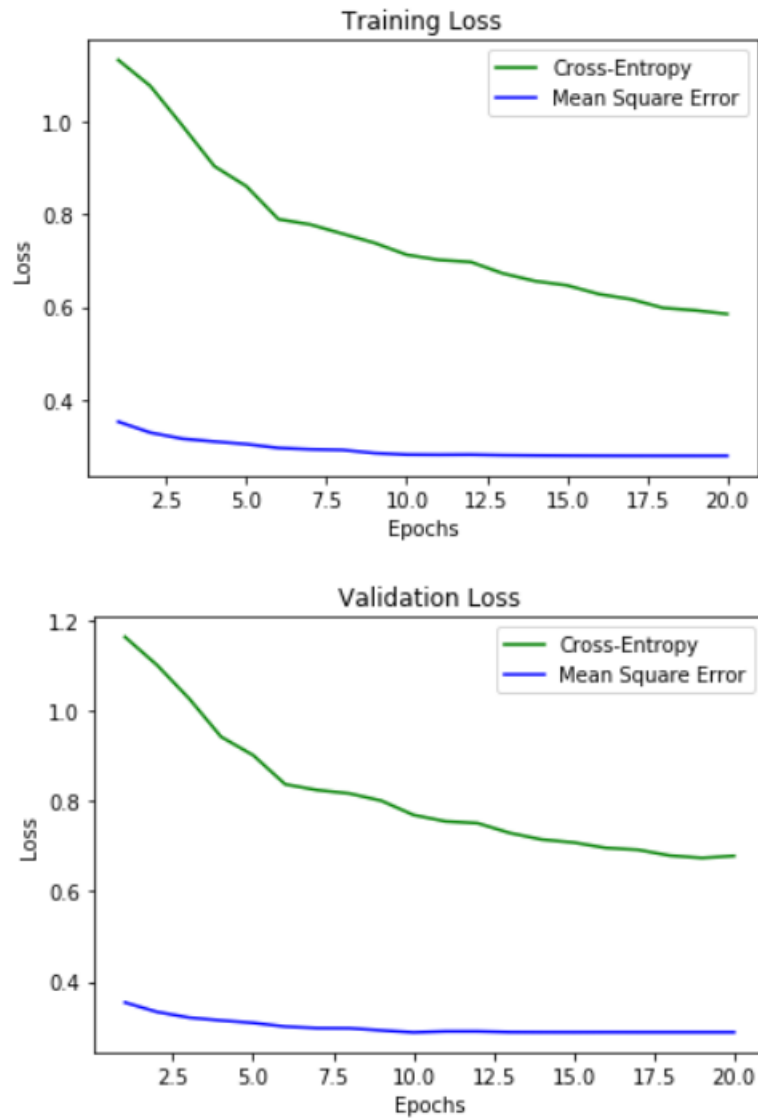
## 6 Answer 6:

Hidden Layers : 2  
 Neurons in each layer : 100 Optimizer : Adam  
 Batch Size : 20  
 Loss : Cross Entropy  
 Initial Learning Rate : 0.001  
 Annealing : True



## 7 Answer 7:

Hidden Layers : 2  
Neurons in each layer : 100  
Optimizer : Adam  
Batch Size : 20  
Activation : Sigmoid  
Initial Learning Rate : 0.001  
Annealing : True

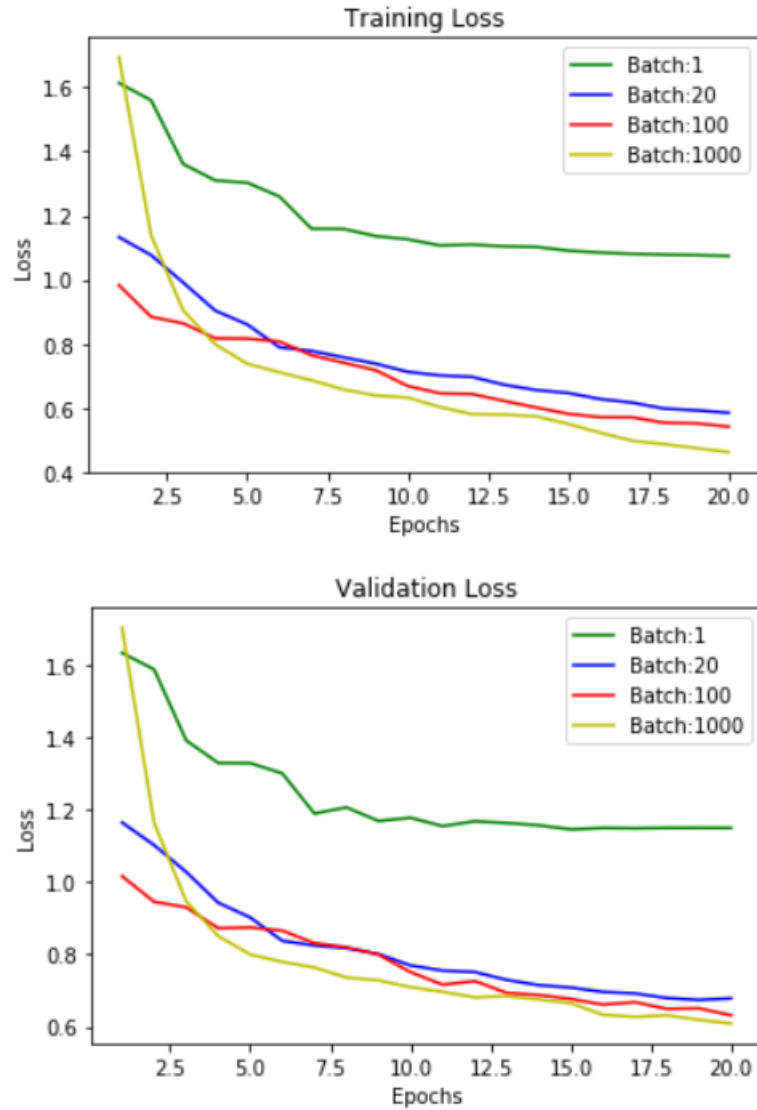


## 8 Answer 8:

Hidden Layers : 2  
Neurons in each layer : 100  
Optimizer : Adam  
Loss : Cross Entropy  
Activation : Sigmoid  
Initial Learning Rate : 0.001



Annealing : True



## 9 Note

The results used in the report are obtained **without** using PCA, Dropout and L2 regularization.

However, for improving the model's performance on the Kaggle dataset, we have used PCA for dimensionality reduction, and have used Dropout and L2 regularization to prevent overfitting.