CS7015 : Programming Assignment 1

Sandesh Katta(EE15B096), Manish Reddy(EE15B070)

March 2, 2019

1 Pre Processing

- MinMax scalar : since data is sparse and bounded we used minmax scalar to bound inputs between 0 and 1
- PCA

2 Difficulties Faced

- Divide by zero error with softmax function : in Softmax function when inputs are larger negative values denominator becomes zero. We avoided this by mean centering inputs which doesn't change output values.
- Divide by zero error with Gradient of cross Entropy : while finding gradients of cross entropy loss if predicted values are small may result in this error. we avoided this by combining cross entropy and softmax layers which resulted in cleaner expressions

3 Experiments

3.1 One Hidden layer

PCA components = Input features = 50 , Output classes = 10
Number of hidden layers = 1 , Sizes of hidden layers = 50/100/200/300 (same size for each hidden layer)
Loss function = Cross entropy , Activation function = Sigmoid
Optimization algorithm = Adam
Learning rate,η = 0.001 , Batch size = 20

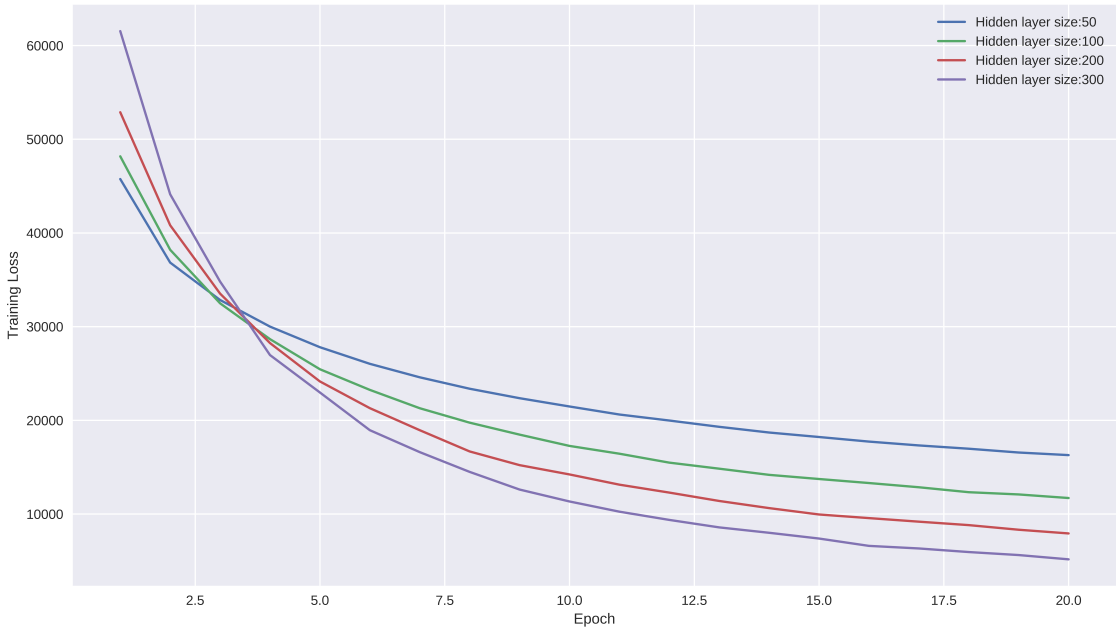

Figure 1: Training Loss for 1 hidden layer

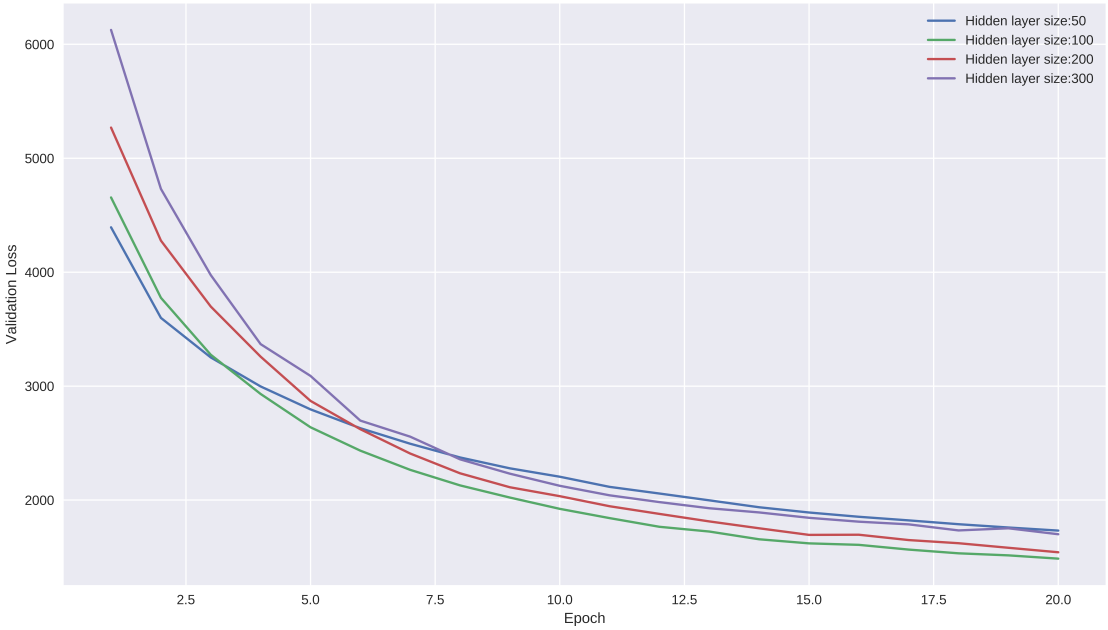

Figure 2: Validation Loss for 1 hidden layer

3.2 Two Hidden layers

PCA components = Input features = 50 , Output classes = 10
Number of hidden layers = 2 , Sizes of hidden layers = 50/100/200/300 (same size for each hidden layer)
Loss function = Cross entropy , Activation function = Sigmoid
Optimization algorithm = Adam

Learning rate,$\eta = 0.001$, Batch size = 20

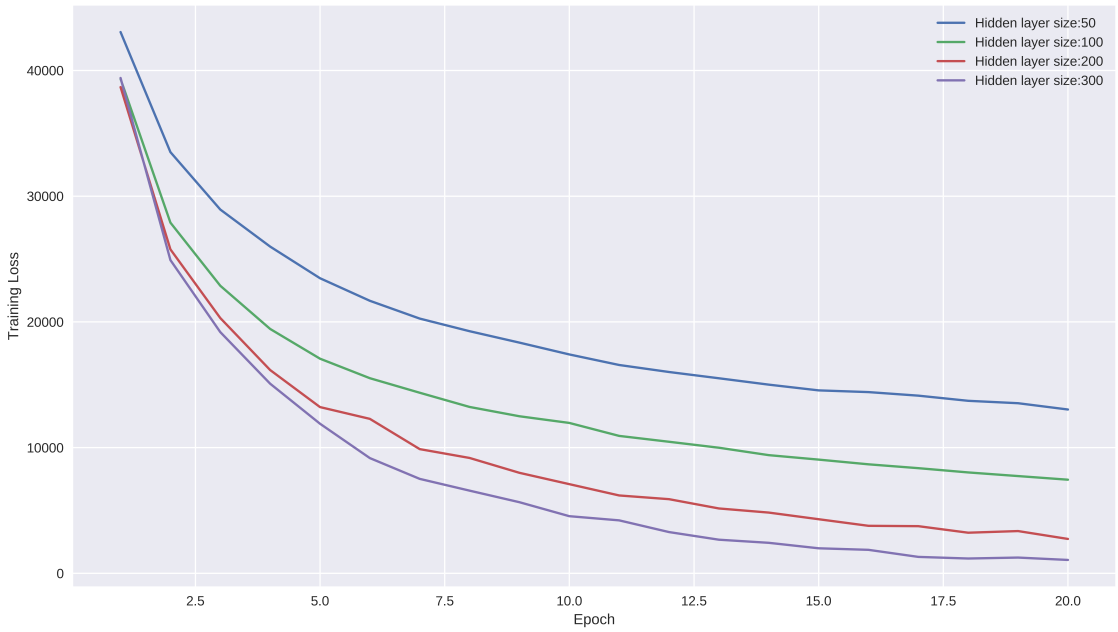

Figure 3: Training Loss for 2 hidden layer

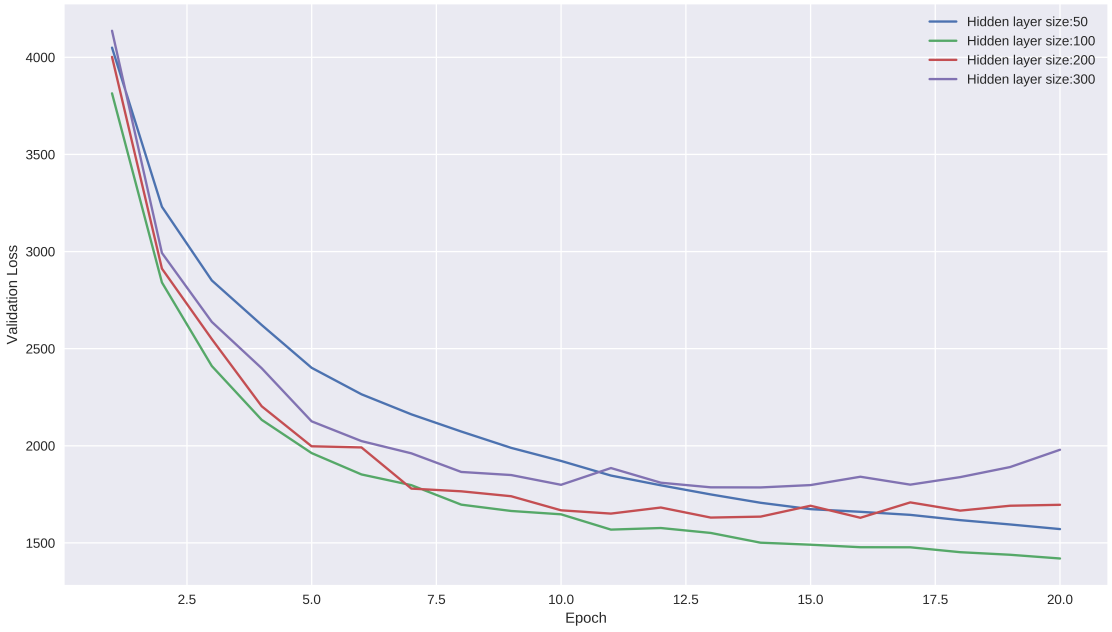

Figure 4: Validation Loss for 2 hidden layer

3.3 Three Hidden layers

PCA components = Input features = 50 , Output classes = 10
Number of hidden layers = 3 , Sizes of hidden layers = 50/100/200/300 (same size for each hidden layer)
Loss function = Cross entropy , Activation function = Sigmoid
Optimization algorithm = Adam
Learning rate,$\eta = 0.001$, Batch size = 20

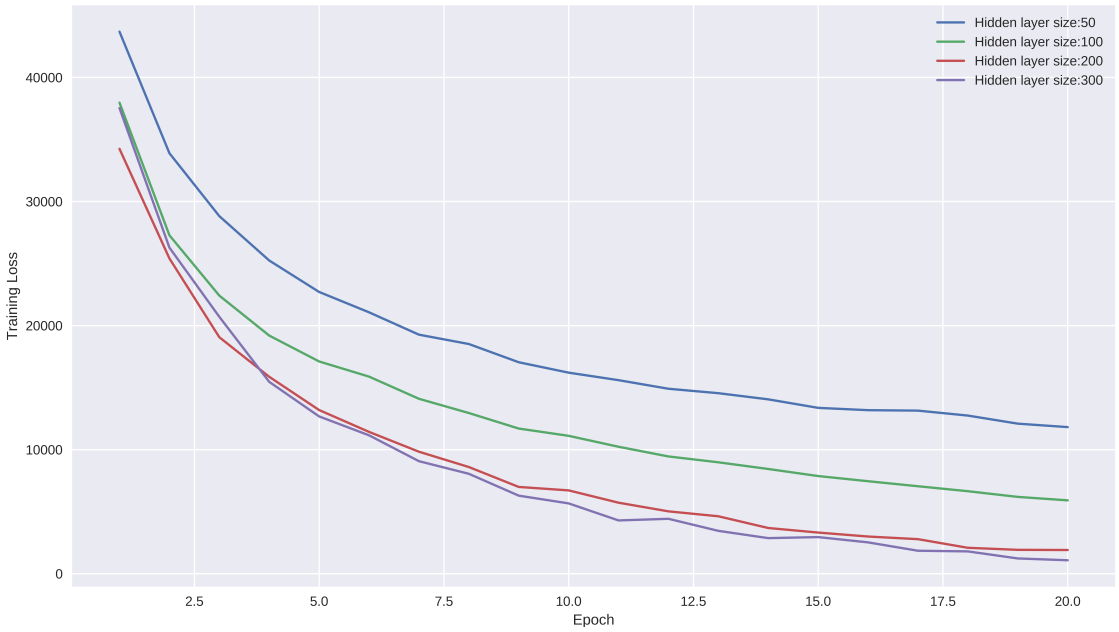

Figure 5: Training Loss for 3 hidden layer

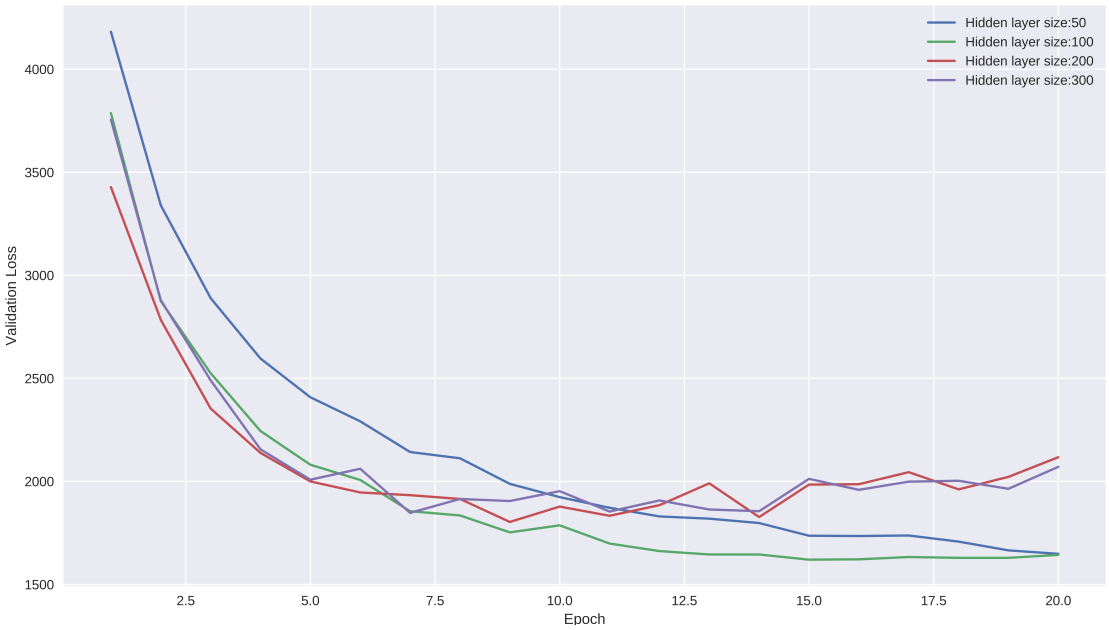

Figure 6: Validation Loss for 3 hidden layer

3.4 Four Hidden layers

PCA components = Input features = 50 , Output classes = 10
Number of hidden layers = 4 , Sizes of hidden layers = 50/100/200/300 (same size for each hidden layer)
Loss function = Cross entropy , Activation function = Sigmoid
Optimization algorithm = Adam
Learning rate,$\eta = 0.001$, Batch size = 20

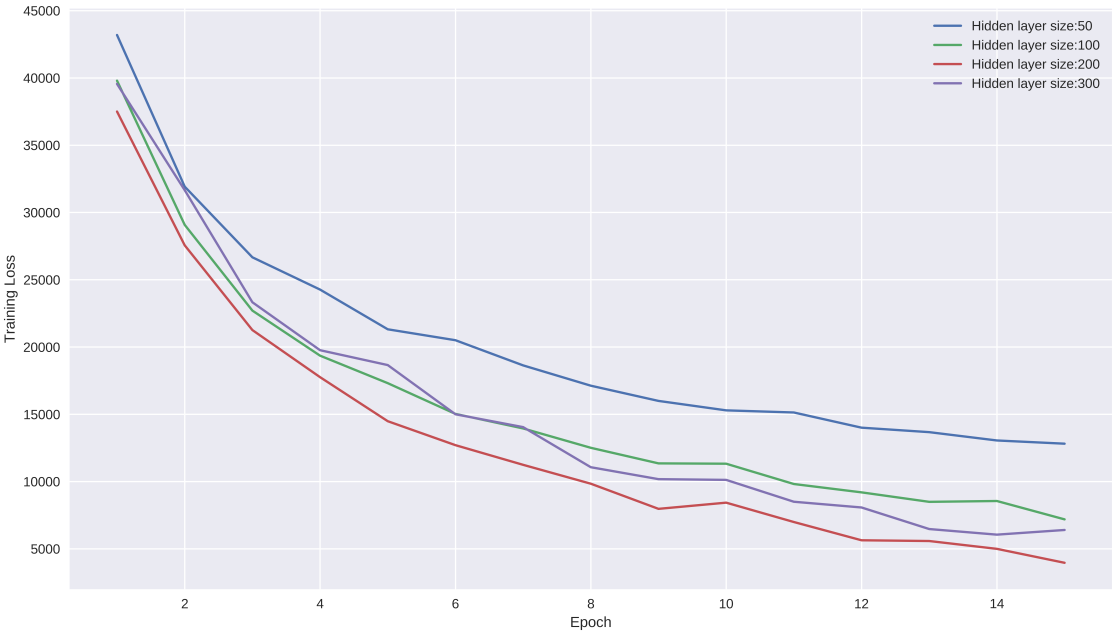

Figure 7: Training Loss for 4 hidden layer

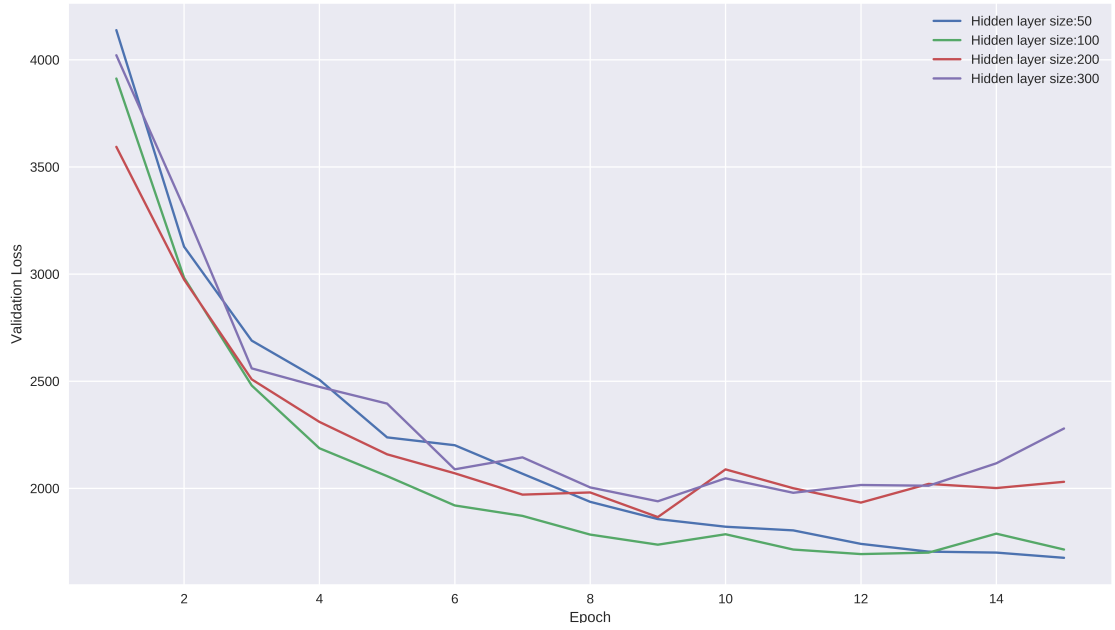

Figure 8: Validation Loss for 4 hidden layer

3.5 Optimizer: Adam vs NAG vs GD vs Momentum

PCA components = Input features = 50 , Output classes = 10
Number of hidden layers = 2 , Sizes of hidden layers = (100,100)
Loss function = Cross entropy , Activation function = Sigmoid
Optimization algorithm = Adam vs NAG vs GD vs Momentum

Learning rate,$\eta = 0.001$, Batch size = 20

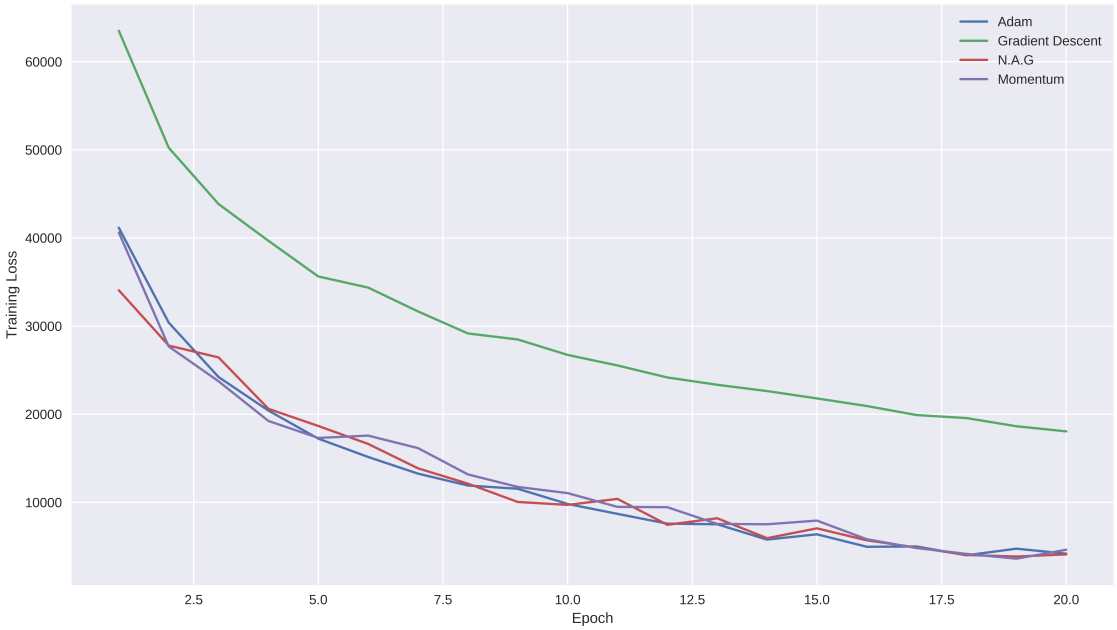

Figure 9: Training Loss

Figure 10: Validation Loss

3.6 Activation Function: Sigmoid vs Tanh

PCA components = Input features = 40 , Output classes = 10
Number of hidden layers = 2 , Sizes of hidden layers = (100,100)

Loss function = Cross entropy , Activation function = Sigmoid vs Tanh
Optimization algorithm = Adam optimizer
Learning rate,$\eta = 0.001$ Batch size = 20

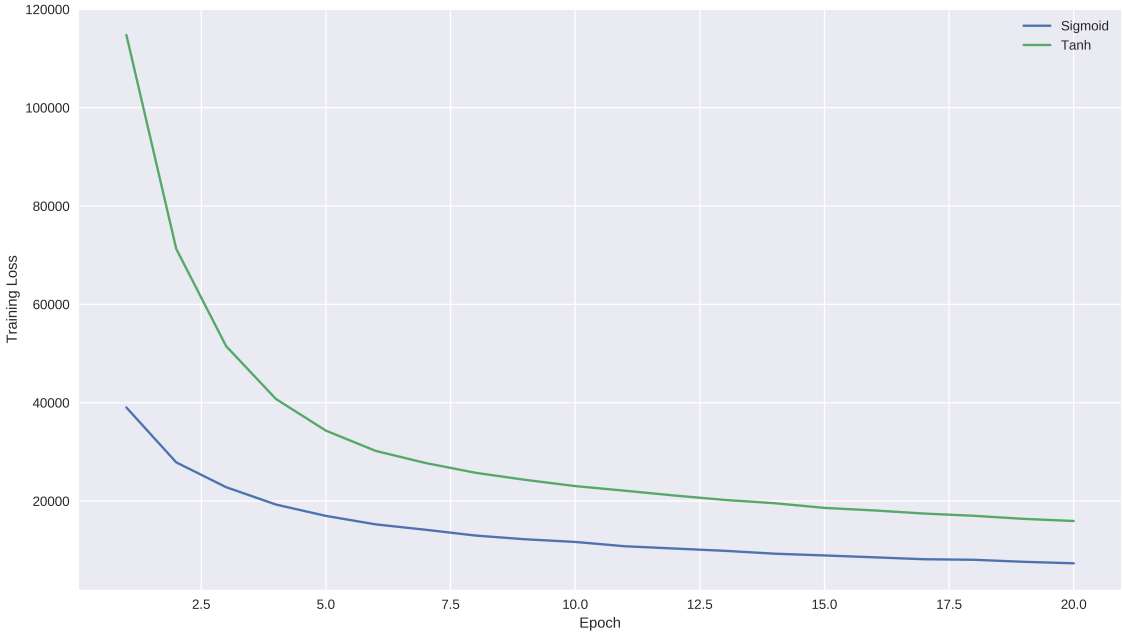

Figure 11: Training Loss

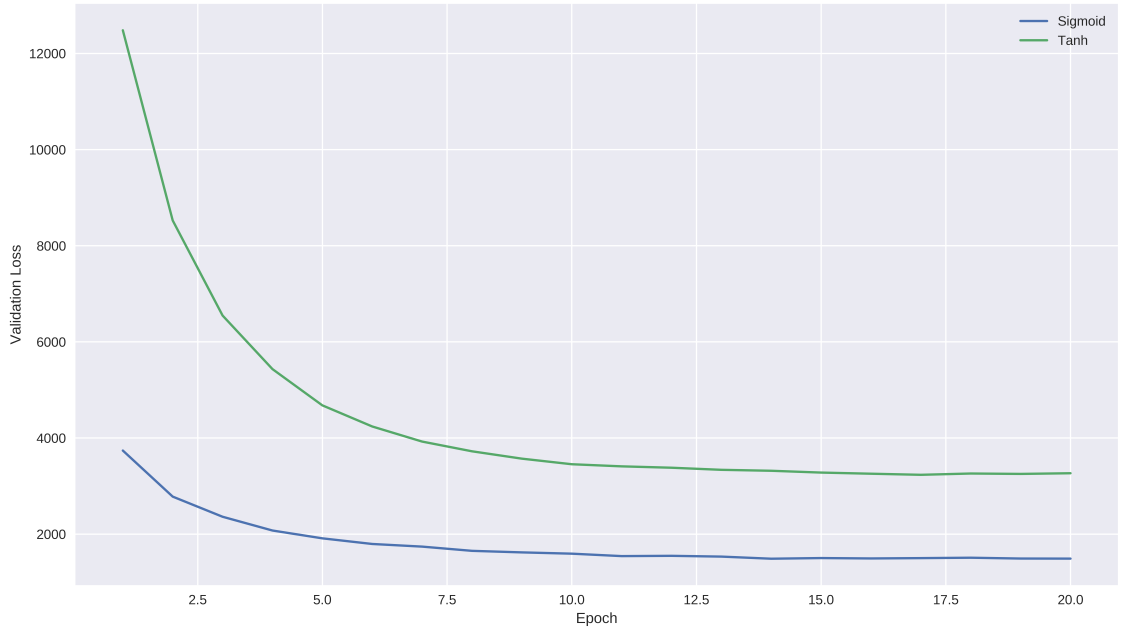

Figure 12: Validation Loss

3.7 Loss Function: Cross Entropy(CE) vs Squared Error(SQ)

PCA components = Input features = 50 , Output classes = 10
Number of hidden layers = 2 , Sizes of hidden layers = (100,100)
Loss function = Cross entropy vs Squared Error , Activation function = Sigmoid
Optimization algorithm = Adam Gradient Descent
Learning rate,$\eta = 0.001$, Batch size = 20

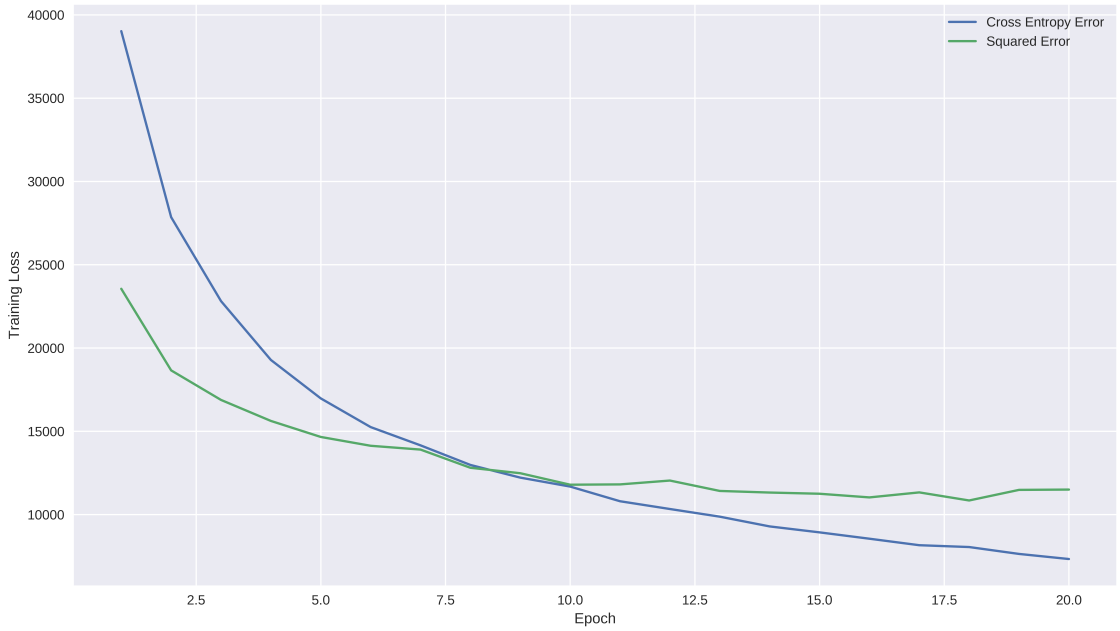

Figure 13: Training Loss

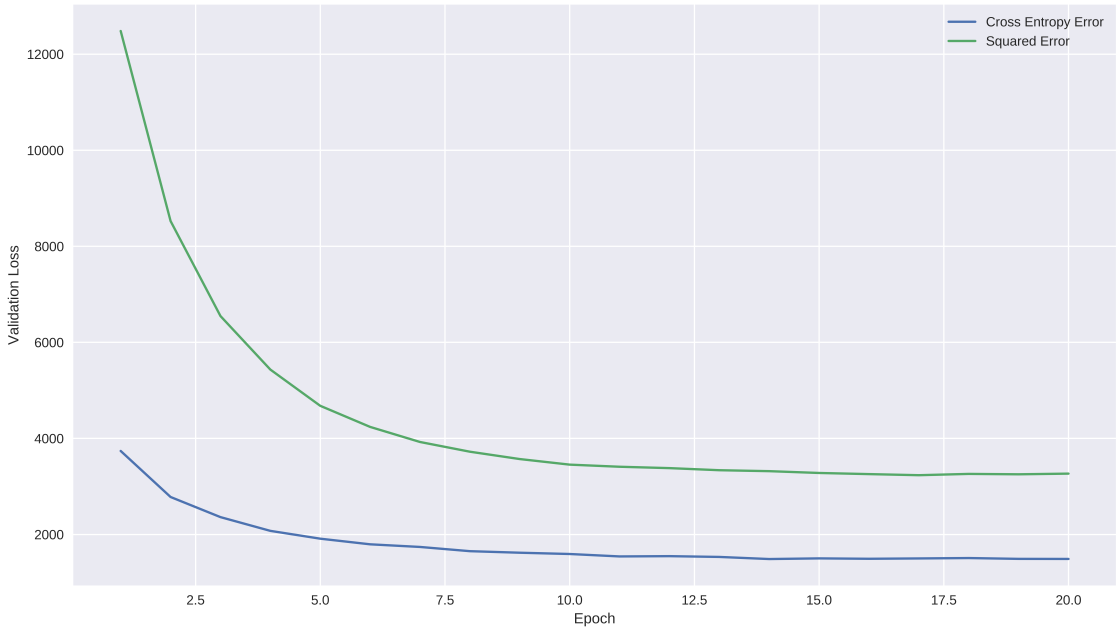

Figure 14: Validation Loss

3.8 Batch Size: 1,20,100,1000

PCA components = Input features = 50 , Output classes = 10
Number of hidden layers = 2 , Sizes of hidden layers = (100,100)
Loss function = Cross entropy , Activation function = Sigmoid
Optimization algorithm = Adam Gradient Descent
Learning rate,$\eta = 0.001$, Batch size = 1 vs 20 vs 100 vs 1000

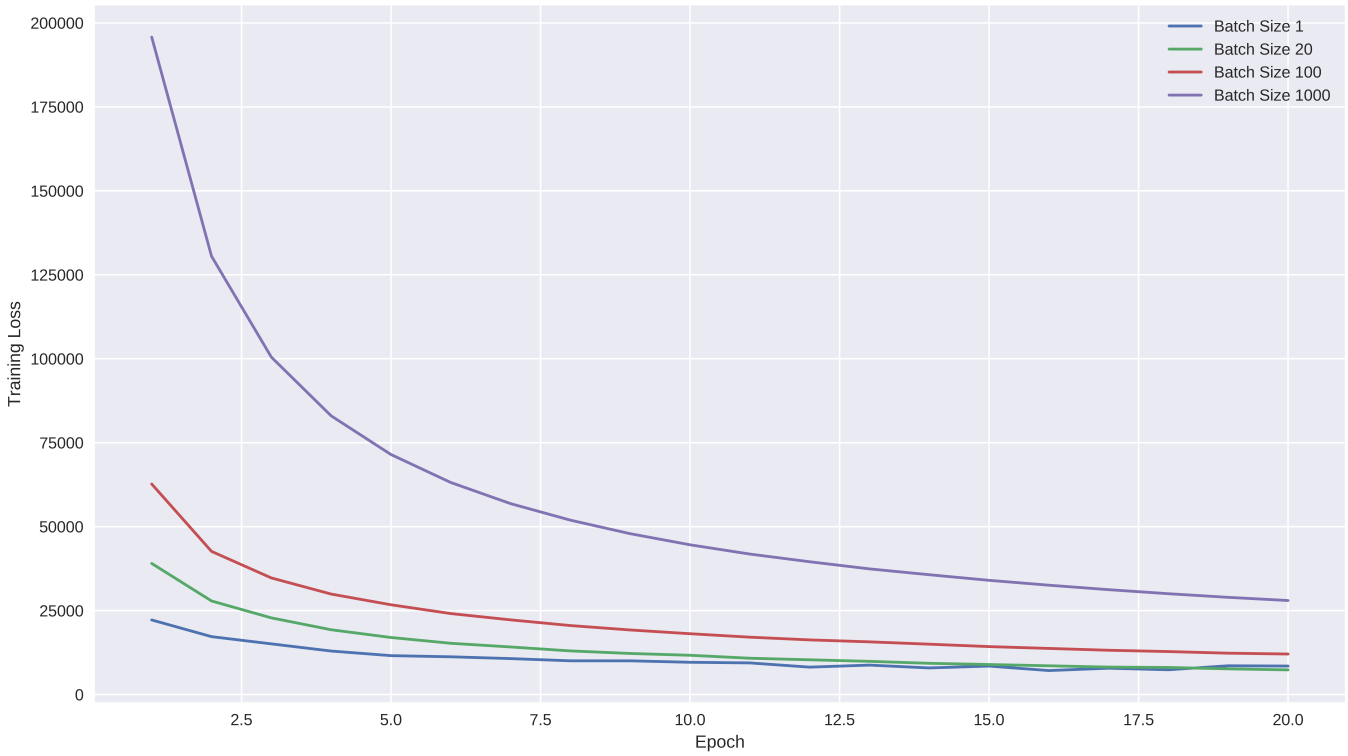

Figure 15: Training Loss

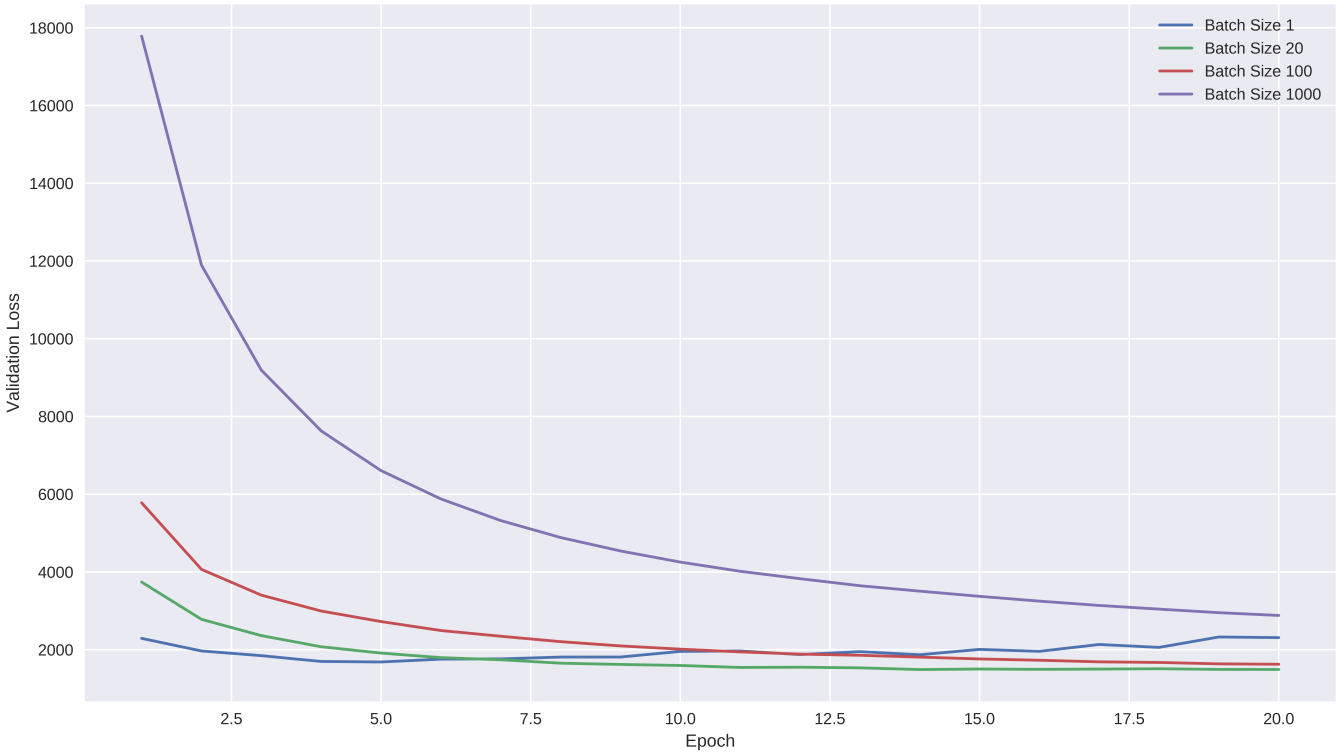

Figure 16: Validation Loss