

HW3 Report Neural Networks

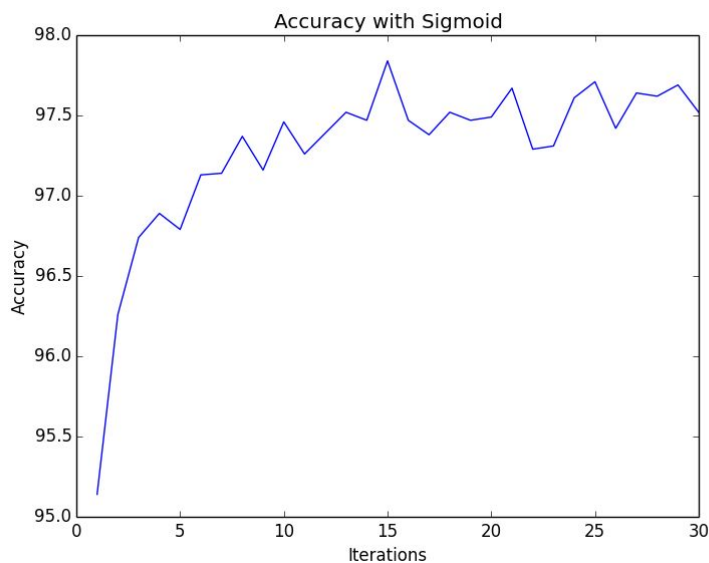
Vasisht Duddu

2015137

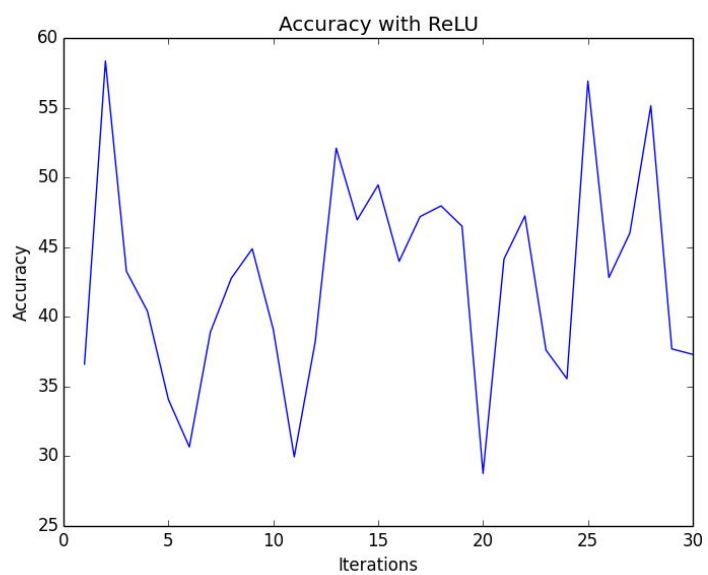
Q1

In this part, I used mini batch and modified the parameter lambda to get better results. By checking for different values of lambda, I observed that for the value of $\lambda=5.0$, the accuracy of the neural network using sigmoid activation was very high. The number of epochs was 30.

Average Accuracy=97.56



However using ReLU as activation function the result dropped and I could not get it above 57%. However, proper tuning of parameters may yield better results and the performance of the model can be improved drastically.



Q2

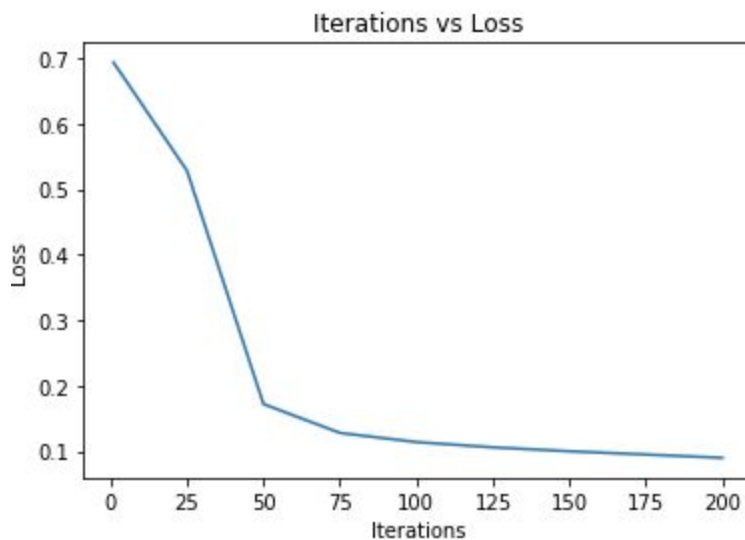
In this question, I used the Standard Scalar method to preprocess the data. Before using the method the accuracy was around 70%. After preprocessing the data, the overall accuracy jumped to 96%.

1) Accuracy: 96.3233230424

activation=logistic or sigmoid

solver=stochastic gradient descent

The hidden layer included two layers of size 100 neurons and 50 neurons each. The parameters can be tuned by changing the learning rate and value of lambda to get better accuracy.



2) Accuracy: 96.0285714286

In this part, the hidden layers used sigmoid activation function while the output layer used the softmax function. The graph shows the plot between epochs and loss. No parameter tuning was performed to achieve the accuracy, however, trying different values of parameters will help to improve the model performance.



Q3

In this question, I used the Standard Scalar method to preprocess the data to improve the accuracy of the model. I tried different activation functions and added more hidden layers and changing the number of neurons in each layer. The best model performance was achieved by ReLU and the result is shown below.

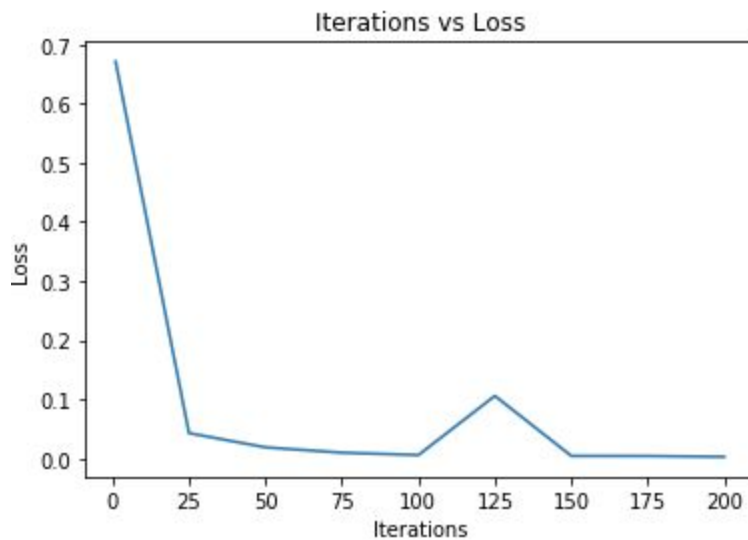
Best Model Results:

Accuracy: 98.4282907662

activation=relu

solver=sgd

Used three hidden layers with 100, 50, 10 neurons respectively.



Probable reason is that ReLU doesn't face gradient vanishing problem as with sigmoid and tanh function. Also, It has been shown that deep networks can be trained efficiently using ReLU even without pre-training.