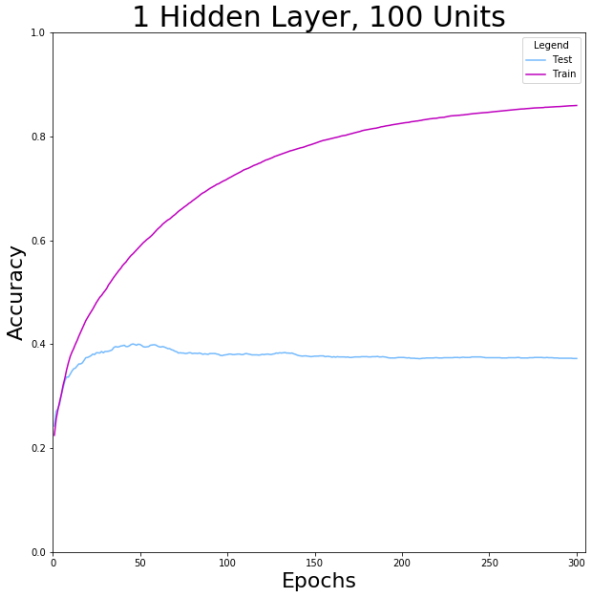
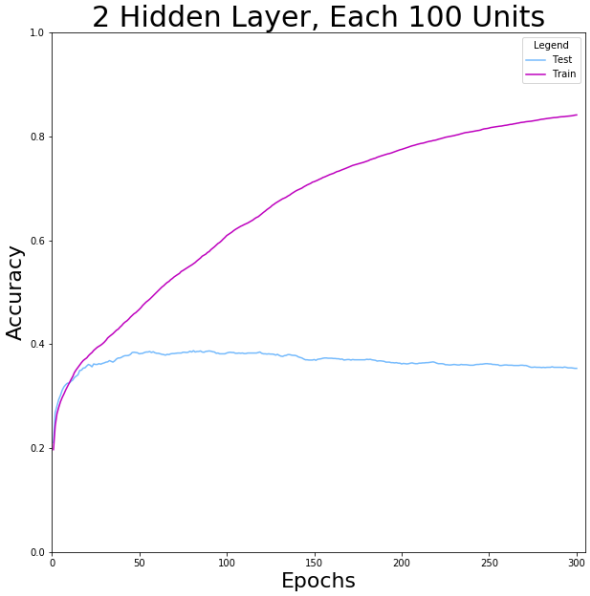
Pratheek D'Souza Rebello 2017CS10361 PART C

Sigmoid Activation, Batchsize = 200, Eta = 1, Adaptive Learning, Iterations = 60000

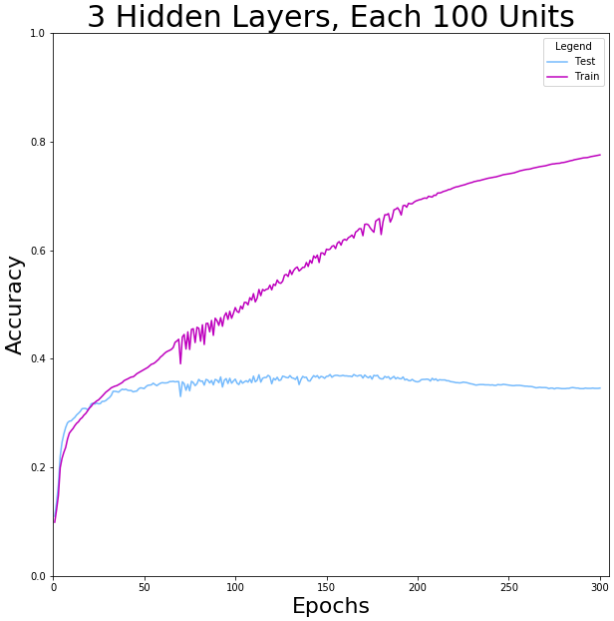
1 Hidden Layer, 100 Units



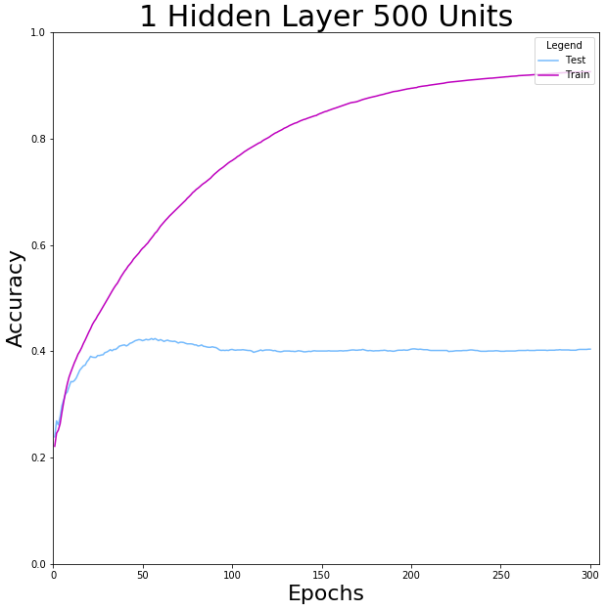
2 Hidden Layer, Each 100 Units



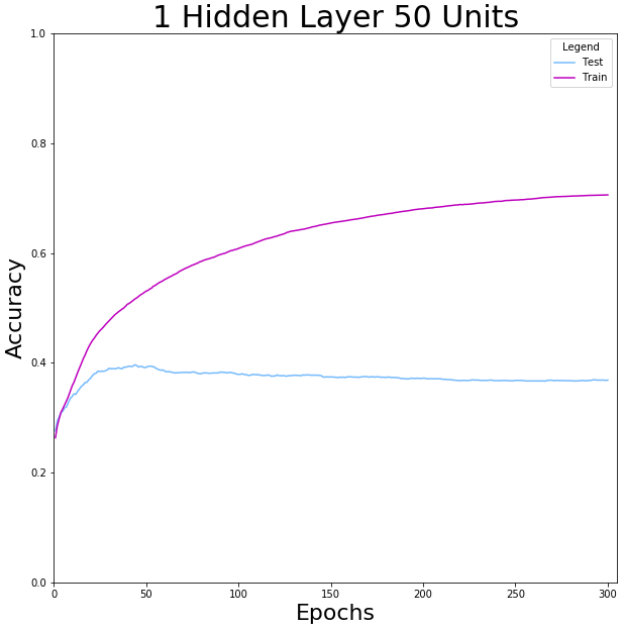
3 Hidden Layers, Each 100 Units



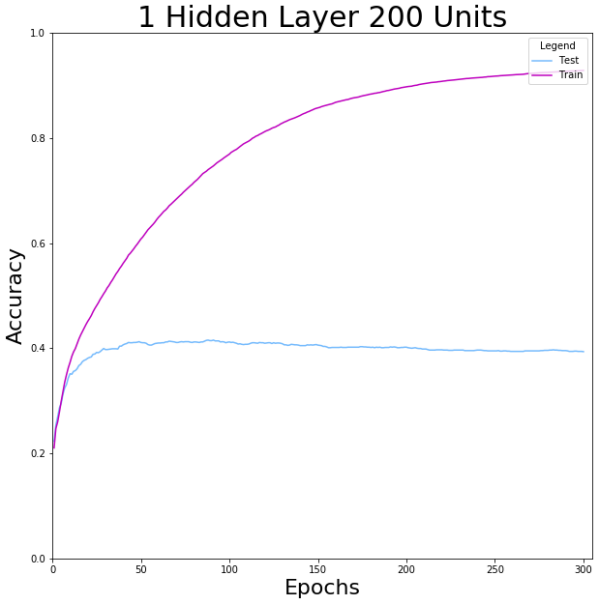
1 Hidden Layer 500 Units



1 Hidden Layer 50 Units



1 Hidden Layer 200 Units



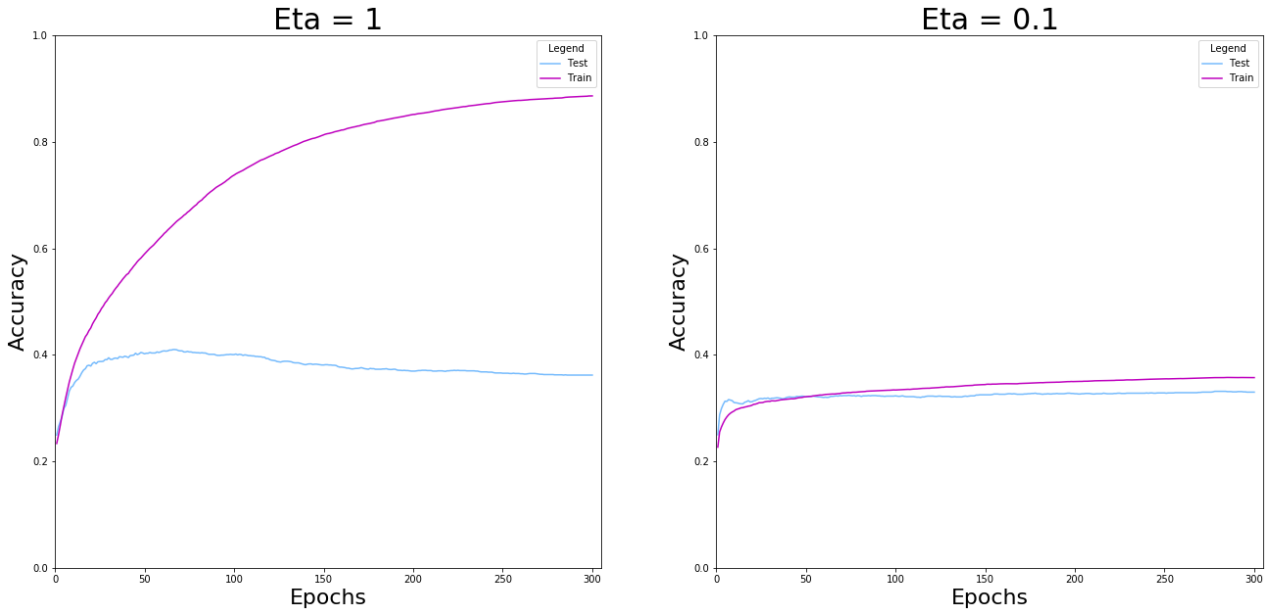
Pratheek D'Souza Rebello 2017CS10361 PART C

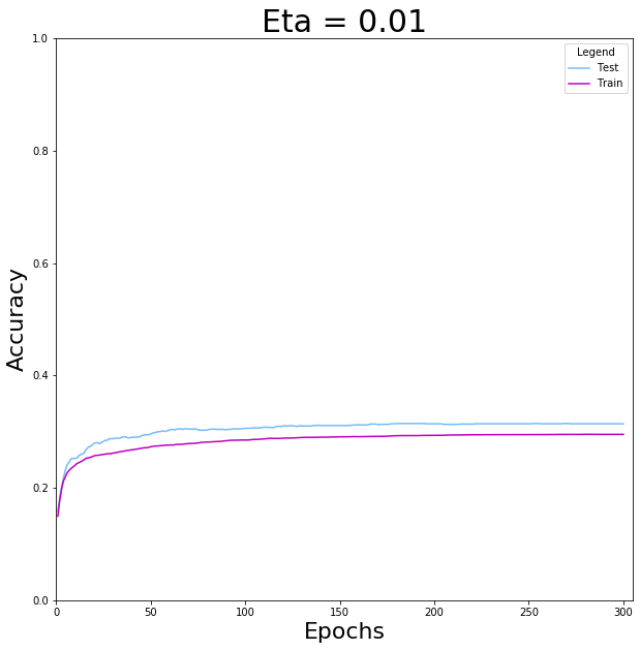
1. On experimentation, I found that deeper networks are overfitting, and also taking a long time to train. Their test accuracies within even 300 epochs, with regularisation does not show any improvement compared to a single hidden layer network.
2. On deciding a single hidden layer, I tried varying the units. I tried 50, 100 and 200 units. 50 is not a sufficient model to learn. 100 and 200 learned well, 200 gave me the best accuracy.

I have finally selected an architecture with ONE hidden layer, with 200 Hidden Units.

LEARNING RATE SELECTION:

I varied learning rate for selected Model: **1 Hidden Layer, 200 Hidden Units**





Pratheek D'Souza Rebello 2017CS10361 PART C

ACTIVATION SELECTION:

1. Sigmoid: Giving decent accuracy as plotted above
2. Linear: Giving poor accuracy as the model is collapsing to a single layered logistic regression
3. ReLU: ReLU models are giving ~4% improvement over Sigmoid. ReLU models are training much faster as well - I have deduced that this can be attributed to vanishing gradients in Sigmoid Networks. In fact, ReLU is training to great accuracies within only 100 epochs, compared to more than 500 epochs for sigmoid.
4. Thus my selected model uses ReLU as activation for the Input and Hidden Layer. It uses the usual softmax for output layer.

REGULARIZATION:

1. I initially tried dropout. The model was achieving very poor train accuracies. I deduce that my shallow model of only one hidden layer does not benefit from dropout.
2. I tried L2 regularization. Regularization with sigmoid function doesn't seem to have significant effect. But with ReLU, L2 regularisation has significant benefit.
3. I tried lambda = 0.001, the model was severely overfitting with a train accuracy of 80% and test accuracy of only 35-37%
4. I increased lambda to 0.01, train accuracy dropped to 65%, and test accuracy improved to 40%
5. I finally selected lambda = 0.05, which gives train accuracy of 55% vs 44% test accuracy.

BATCHSIZE:

1. From experimentation, I have selected batch size = 200

EPOCHS:

1. 10 epochs (1000 iterations): test accuracy ~ 42%
2. 50 epochs (5000 iterations): test accuracy ~ 42.5%
3. 100 epochs (10000 iterations): test accuracy ~ 44.8%
4. 150 epochs (15000 iterations): test accuracy ~ 42.8%

I have selected 100 epochs as the best value

FINAL MODEL:

1 Hidden Layer
200 Units in the Hidden Layer
ReLU activation function
100 epochs (10000 iterations)
200 batchsize
Learning Rate Seed = 0.5
Learning Rate Strategy = Adaptive