

Neural Network Training

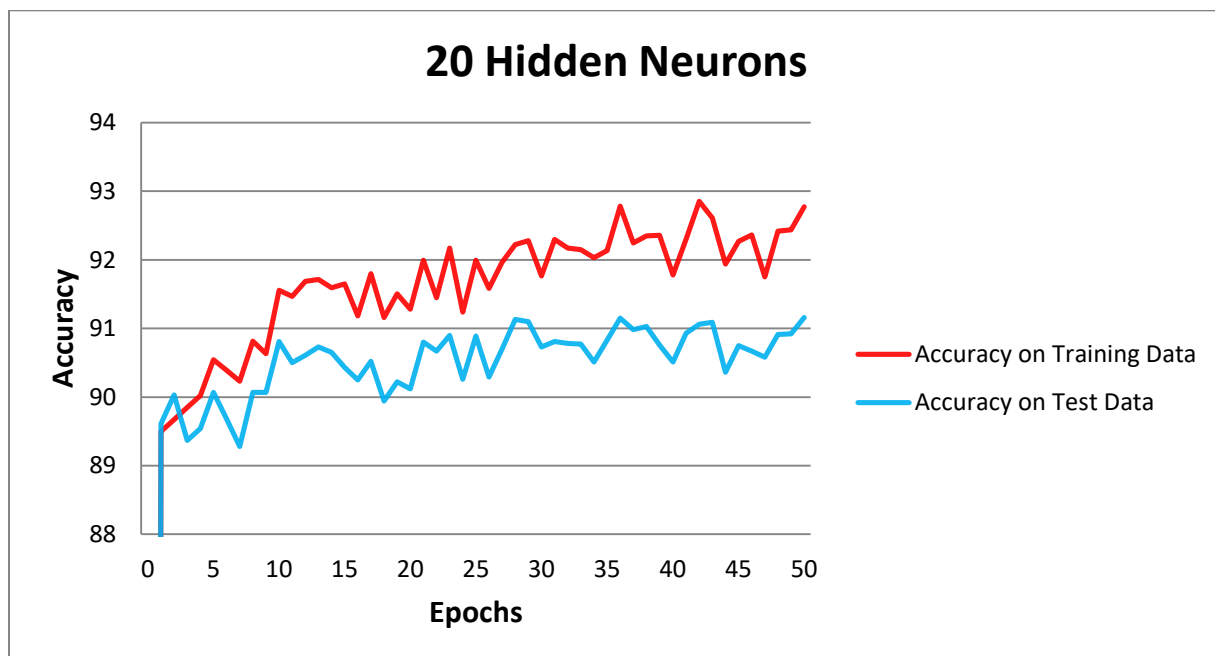
Experiment 1

Description

This experiment has us varying the size of the hidden layer of the neural network by changing the number to 20, 50, and 100. Initial accuracy is calculated on given weights and then weights are updated after each training example is run through. Each network is trained for 50 Epochs and the plots contain the accuracy of both training and test data and the confusion matrix contains data from the final test run after the 50 epochs.

Hidden Neurons = 20

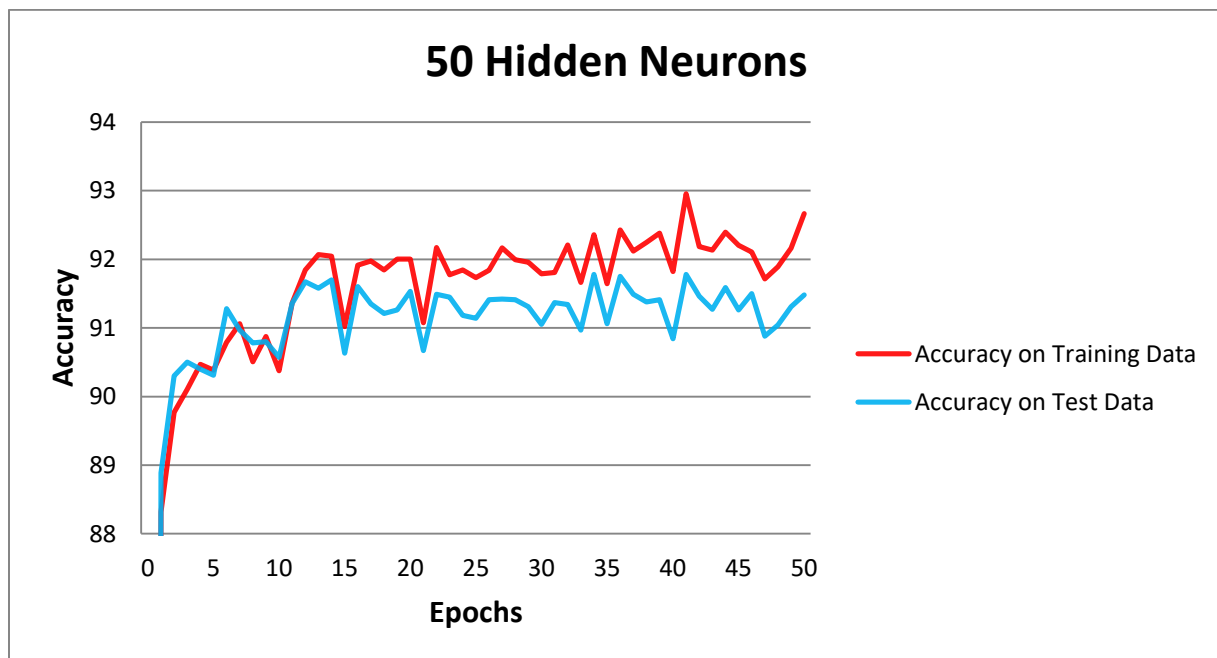
Comments: Initial Training data accuracy: 7.6% and initial Test data accuracy: 7.5%



		Predicted Class									
		0	1	2	3	4	5	6	7	8	9
Actual Class	0	941	0	3	2	1	7	6	1	8	5
	1	0	1106	2	0	1	1	3	4	3	3
	2	2	1	929	26	1	2	2	27	5	0
	3	1	5	6	902	0	33	0	3	8	15
	4	3	1	5	0	828	1	1	9	5	12
	5	3	1	0	18	0	784	18	0	11	5
	6	14	3	13	3	8	9	894	0	10	1
	7	1	0	8	11	0	2	0	901	6	9
	8	13	16	48	46	13	39	28	13	899	27
	9	2	2	18	2	130	14	6	70	19	932

Hidden Neurons = 50

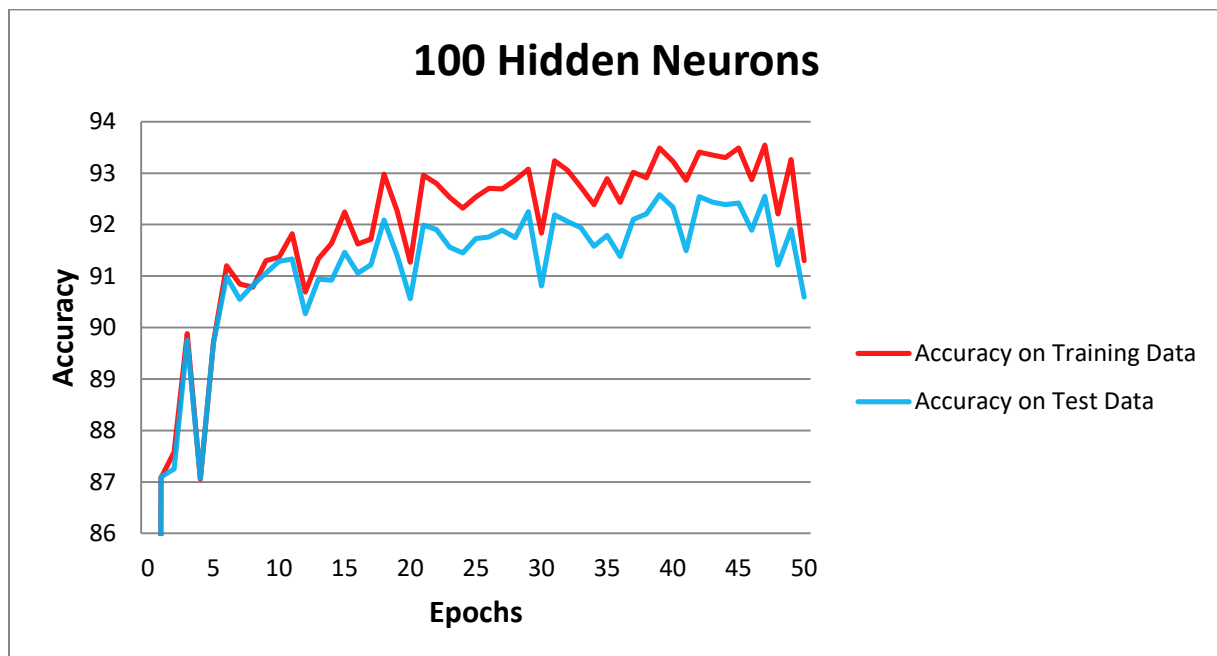
Comments: Initial Training data accuracy: 9.8% and initial Test data accuracy: 9.5%



		Predicted Class									
		0	1	2	3	4	5	6	7	8	9
Actual Class	0	935	0	1	3	1	4	6	0	3	1
	1	0	1112	5	0	1	1	3	9	7	7
	2	2	2	896	11	2	2	4	15	5	3
	3	3	3	25	886	1	30	1	5	8	5
	4	2	0	7	0	846	1	1	3	4	9
	5	16	2	6	32	1	799	19	3	14	6
	6	9	2	8	0	9	9	906	1	5	0
	7	2	1	11	11	7	4	1	930	7	17
	8	9	13	60	63	18	30	9	35	911	34
	9	2	0	13	4	96	12	8	27	10	927

Hidden Neurons = 100

Comments: Initial Training data accuracy: 9.7% and initial Test data accuracy: 9.7%



		Predicted Class									
		0	1	2	3	4	5	6	7	8	9
Actual Class	0	950	0	12	1	2	12	14	5	8	7
	1	2	1121	27	10	5	25	18	19	15	7
	2	1	1	912	11	0	2	5	14	1	0
	3	4	3	30	946	1	40	1	28	14	12
	4	2	0	13	2	932	6	6	11	10	42
	5	3	0	0	13	0	776	10	1	21	6
	6	10	2	6	0	12	4	892	0	9	0
	7	2	0	3	1	0	0	0	731	0	0
	8	5	5	16	14	2	12	9	1	869	5
	9	1	3	13	12	28	15	3	218	27	930

Summary

The number of hidden units increase the rate of improvement with larger numbers but in the later epochs start to decrease in accuracy. The confusion matrix shows that 50 hidden units gives the best final accuracy for the test data. When observing the data when the number of hidden units increases it takes a larger number of epochs for the training to converge. Twenty converges around 24 epochs, fifty converges around 27 epochs, and one hundred converges around 41 epochs. There is evidence of overfitting which is most notable in the 20 hidden units plot, this can be seen by the growing gap between the training and test accuracy as each epoch passes. Compared to the perceptron learning, the results received show that hidden units allow the network to differentiate and learn more which can be seen by the increase of accuracy over time rather than the oscillations of the perceptron without much improvement.

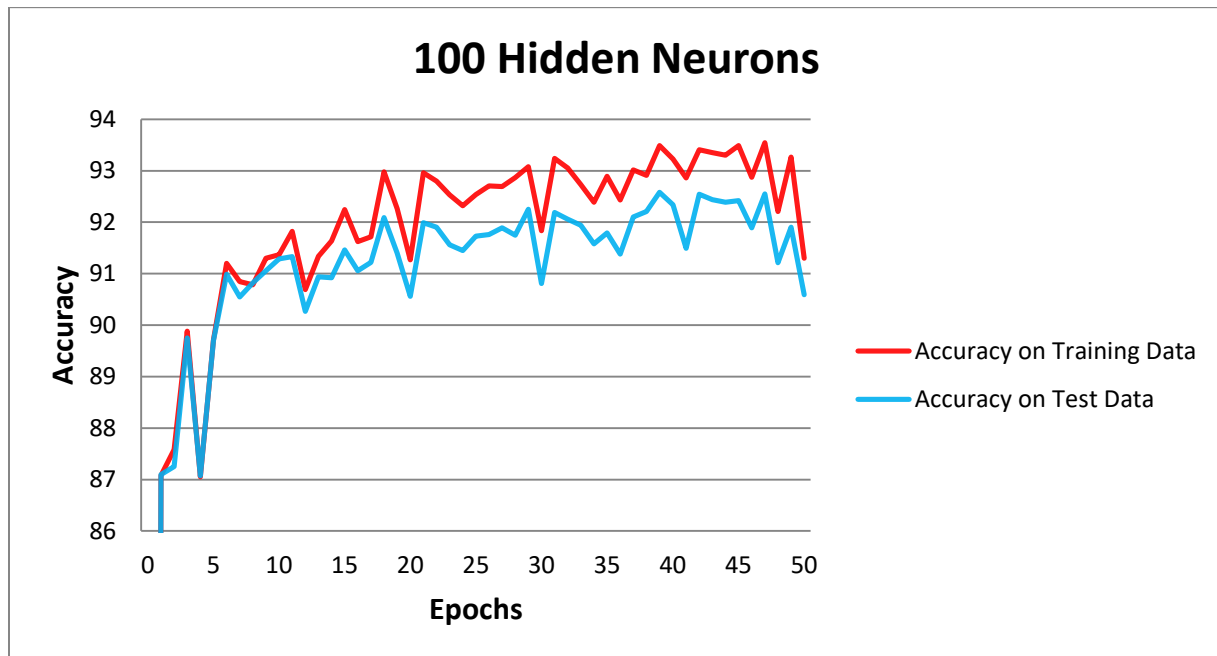
Experiment 2

Description

This experiment has us varying the momentum value while using our network with the 100 hidden neurons. The momentum values that will be used are 0, 0.25, and 0.5. The original plot with 100 hidden neurons and 0.9 momentum will also be included. Each network is trained for 50 Epochs and the plots contain the accuracy of both training and test data and the confusion matrix contains data from the final test run after the 50 epochs.

Momentum = 0.9

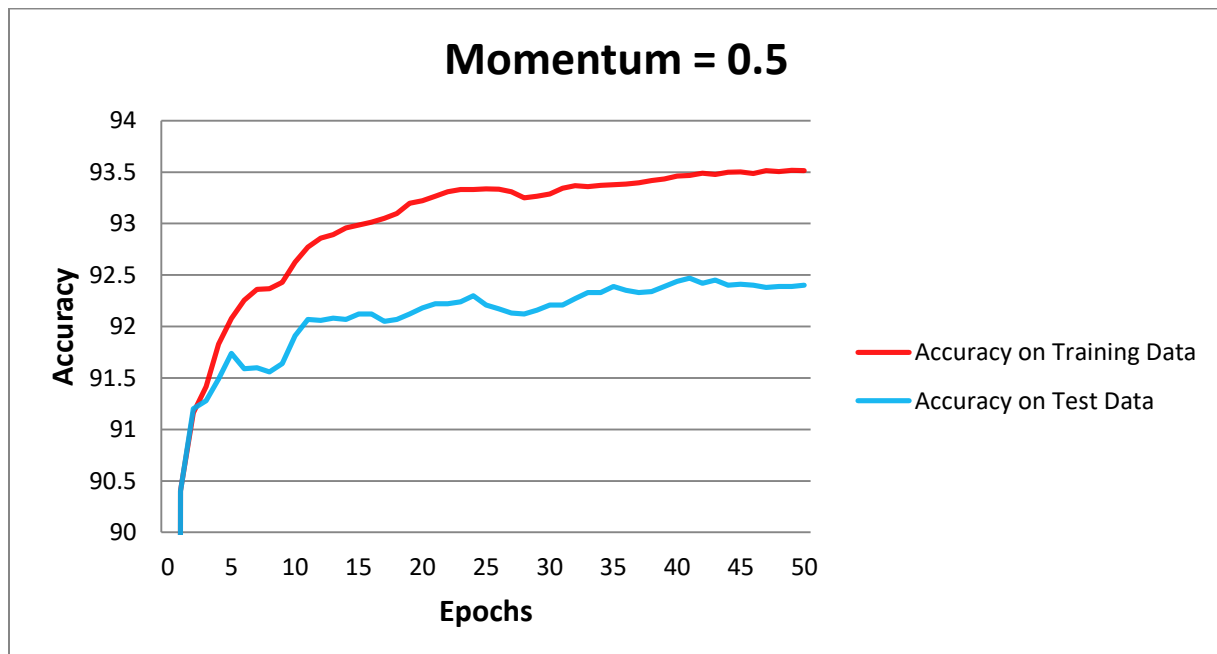
Comments: Initial Training data accuracy: 9.7% and initial Test data accuracy: 9.7%



Comment: confusion matrix shown above in prev. section

Momentum = 0.5

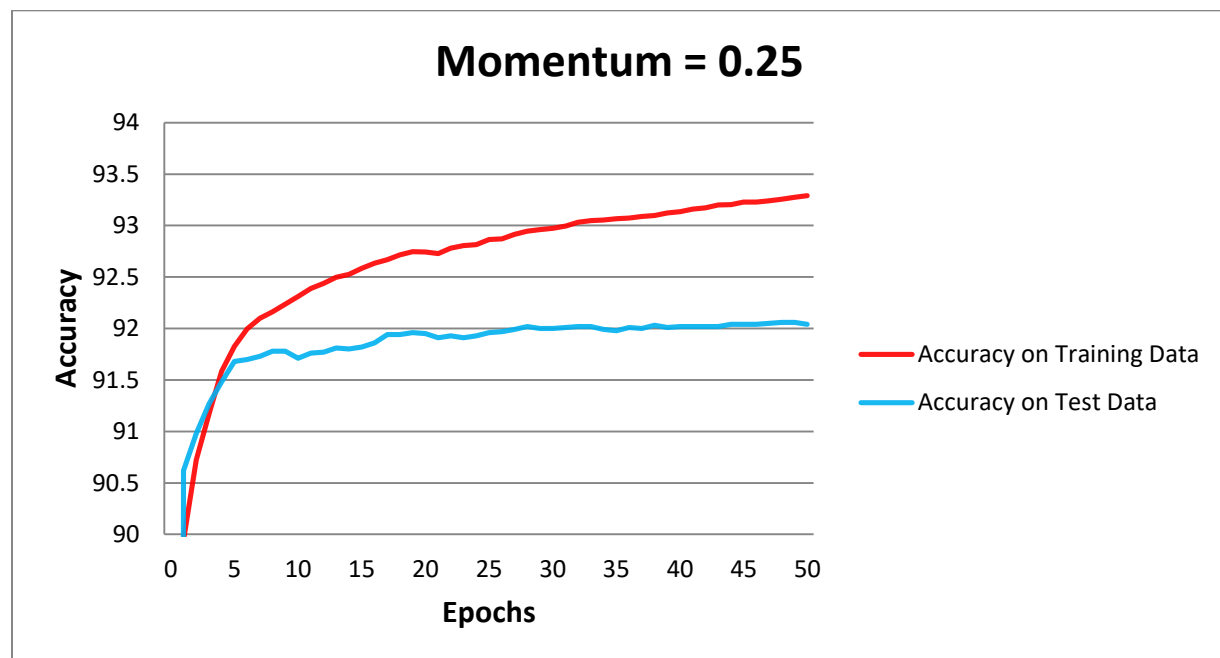
Comments: Initial Training data accuracy: 8.7% and initial Test data accuracy: 8.8%



		Predicted Class									
		0	1	2	3	4	5	6	7	8	9
Actual Class	0	957	0	11	2	1	1	24	1	9	8
	1	0	1114	7	0	0	8	3	6	7	7
	2	0	2	905	5	2	0	3	14	1	0
	3	2	3	24	935	0	22	1	3	8	1
	4	0	0	9	1	910	6	4	3	7	18
	5	3	3	5	31	0	799	15	6	33	6
	6	13	3	17	2	20	17	895	6	10	2
	7	1	1	5	9	2	3	0	919	7	22
	8	3	9	48	22	2	25	13	17	883	22
	9	1	0	1	3	45	11	0	53	9	923

Momentum = 0.25

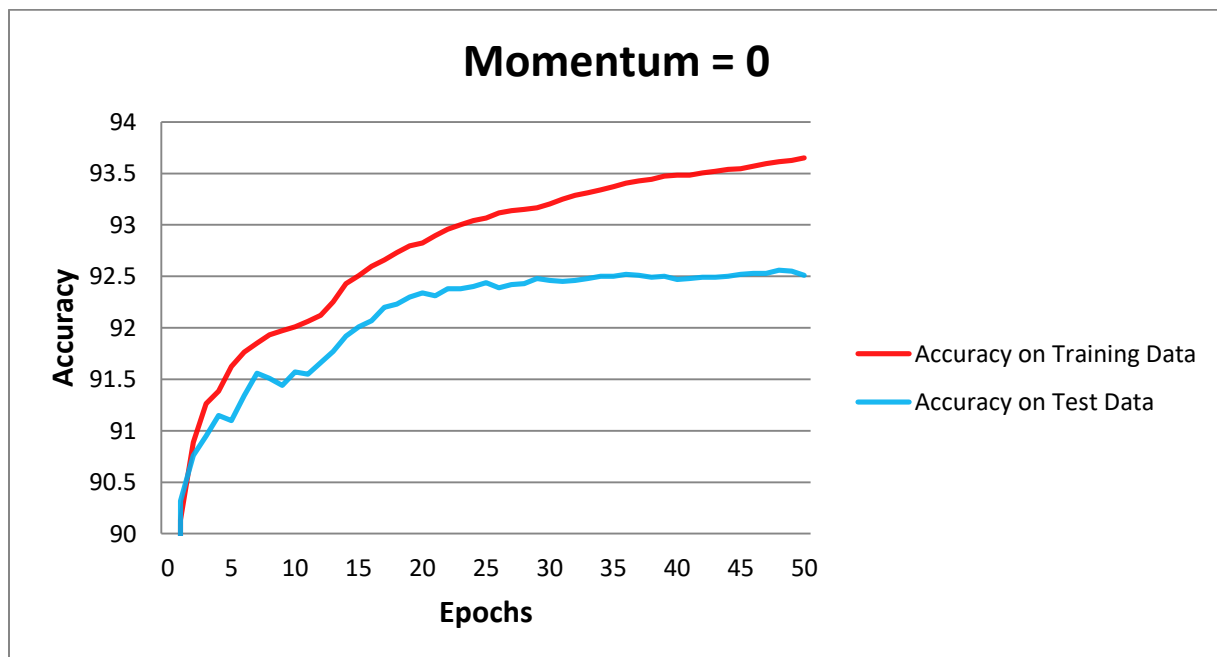
Comments: Initial Training data accuracy: 9.8% and initial Test data accuracy: 9.5%



		Predicted Class									
		0	1	2	3	4	5	6	7	8	9
Actual Class	0	947	0	17	8	0	16	8	2	7	9
	1	0	1109	3	2	0	3	2	9	8	2
	2	1	5	914	9	4	3	8	21	2	0
	3	1	4	34	924	2	45	7	13	20	7
	4	1	1	7	1	906	7	7	3	10	13
	5	5	1	1	12	1	741	10	1	16	4
	6	12	4	11	1	16	19	900	3	11	0
	7	2	1	7	9	1	5	1	934	2	8
	8	11	10	33	28	9	36	15	2	874	11
	9	0	0	5	16	43	17	0	40	24	955

Momentum = 0

Comments: Initial Training data accuracy: 9.7% and initial Test data accuracy: 9.7%



		Predicted Class									
		0	1	2	3	4	5	6	7	8	9
Actual Class	0	944	0	16	6	3	11	21	1	4	6
	1	0	1115	4	0	0	1	4	10	7	5
	2	1	1	939	26	4	4	4	17	8	0
	3	0	6	16	916	0	33	0	3	14	20
	4	5	1	10	0	916	7	5	6	10	24
	5	6	1	2	21	0	774	11	0	16	3
	6	11	4	14	5	8	14	908	3	19	0
	7	3	1	6	12	1	11	0	940	13	25
	8	7	6	22	16	2	24	5	3	876	3
	9	3	0	3	8	48	13	0	45	7	923

Summary

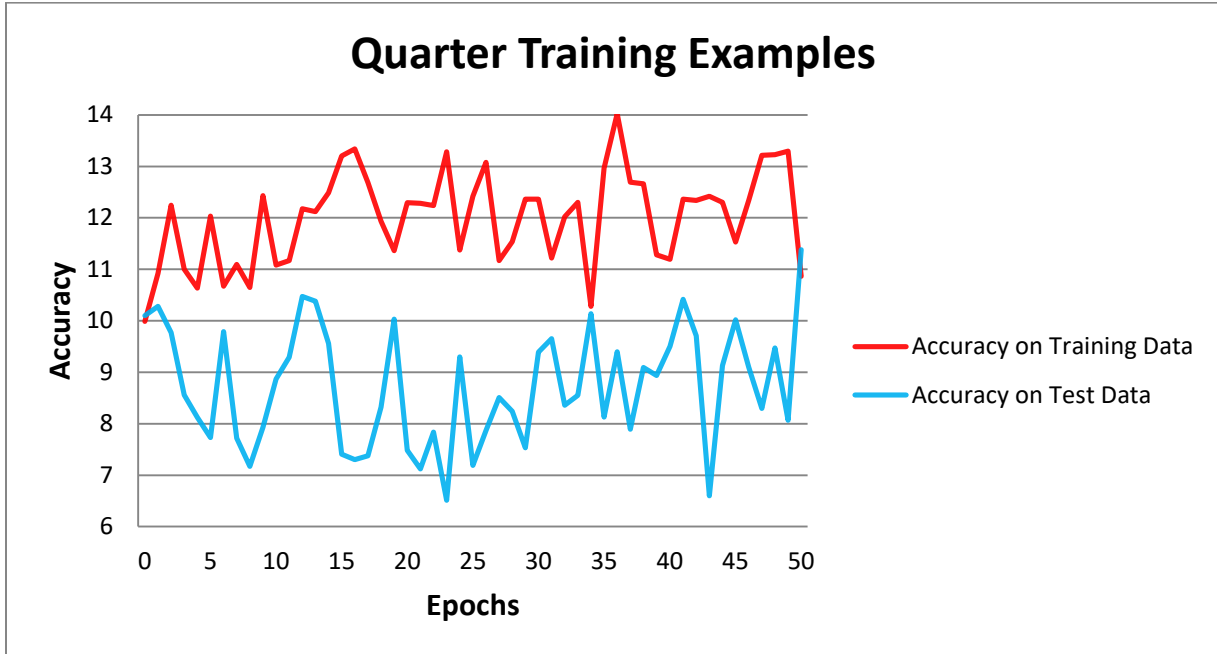
The momentum value 0.25 had the best final accuracy overall but they were all generally in the same range. From what was observed as the momentum value decreased so did the evidence of overfitting. As the momentum decreased the network would show signs of convergence sooner. The 0.5 momentum converged around epoch 40, 0.25 momentum converged around epoch 26, and 0 momentum converged around epoch 27. There is evidence of overfitting which can be seen in all three. This is shown by seeing the test accuracy stop increasing while the training data keeps improving.

Experiment 3

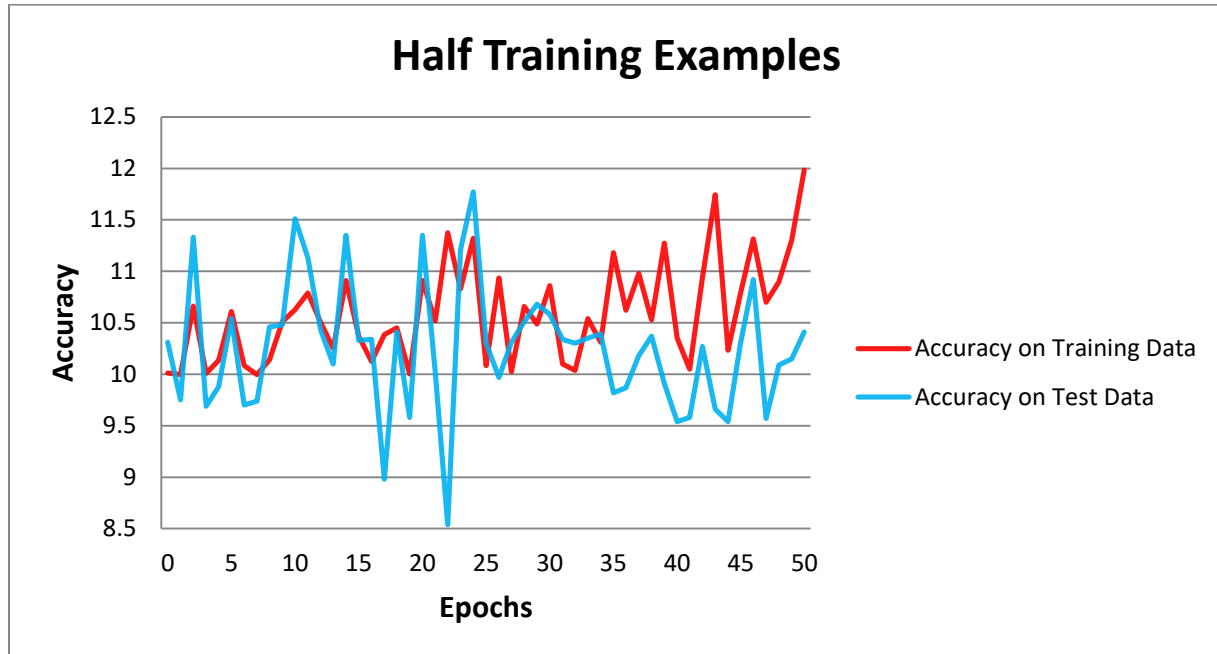
Description

This experiment has us varying the size of the training examples. While fixed to 100 hidden neurons and a momentum of 0.9 train two neural networks using respectively one quarter and one half of the training examples. Both training data samples have a balanced distribution among the 10 different classes. Each network is trained for 50 Epochs and the plots contain the accuracy of both training and test data and the confusion matrix contains data from the final test run after the 50 epochs.

One Quarter Training Data

[illegible]

One Half Training Data



		Predicted Class									
		0	1	2	3	4	5	6	7	8	9
Actual Class	0	0	0	0	0	0	0	0	0	0	0
	1	179	299	390	254	268	148	214	123	189	160
	2	3	0	3	1	9	8	19	22	0	15
	3	778	829	594	716	672	708	702	846	780	819
	4	0	0	0	0	0	0	0	0	0	0
	5	1	0	1	1	1	3	3	1	0	0
	6	19	7	44	38	32	25	20	36	5	15
	7	0	0	0	0	0	0	0	0	0	0
	8	0	0	0		0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	0	0

Summary

The size of the training data gives the network less to go off of so it doesn't do well in the epochs given. It will have to train for much longer to reach where it was when using 60,000 training examples. With a quarter the plots are mirrored which can so indication of overfitting. The size does effect the epochs required for convergence, as seen in the half training example plot both training and test start out not in sync but as epochs pass then they begin to fall in line together. I believe with more epochs they will show signs of convergence.