EE596: HW #1

Daniel Cashon-Smith

04/15/19

1 MNIST Neural Net Modification

Fig. (1) details the neural network architecture and hyperparameters tried:

Batch Size	Architecture	Function	Learning Rate	Number of Iterations
16	[256, 128, 64]	sigmoid	0.01	1000
32	[256, 128, 64]	sigmoid	0.01	1000
64	[256, 128, 64]	sigmoid	0.01	1000
128	[256, 128, 64]	sigmoid	0.01	1000
64	[256]	sigmoid	0.01	1000
64	[256, 256]	sigmoid	0.01	1000
64	[256, 256, 256]	sigmoid	0.01	1000
64	[256, 256, 256, 256, 256]	sigmoid	0.01	1000
64	[256, 256, 256, 256, 256, 256, 256, 256]	sigmoid	0.01	1000
64	[256. 256]	sigmoid	1	1000
64	[256. 256]	sigmoid	0.2	1000
64	[256. 256]	sigmoid	0.1	1000
64	[256. 256]	sigmoid	0.05	1000
64	[256. 256]	sigmoid	0.001	1000
64	[256. 256]	sigmoid	0.0001	1000
64	[256. 256]	sigmoid	0.01	1000
64	[256. 256]	relu	0.01	1000
64	[256. 256]	tanh	0.01	1000
64	[64,64]	sigmoid	0.01	1000
64	[128, 128]	sigmoid	0.01	1000
64	[256, 256]	sigmoid	0.01	1000
64	[512, 512]	sigmoid	0.01	1000

Figure 1: Neural Net Modification Table

1.1 Varied Batch Size

Fig. (2) shows the impact of batch size, which was **minor** at best. From here we fix batch size at 64.

1.2 Varied Layer Number with Constant Neuron Number

Fig. (3) shows the impact of varying the number of layers. Counter intuitively, more layers is not necessarily better. From here we fix the number of layers to 2.

1.3 Varied Learning Rate

Fig. (4) demonstrates the impact of changing learning rate. As anticipated low learning rates lead to slower convergence, and high learning rates tend to fail. Learning rate is fixed at 0.01 going forward.

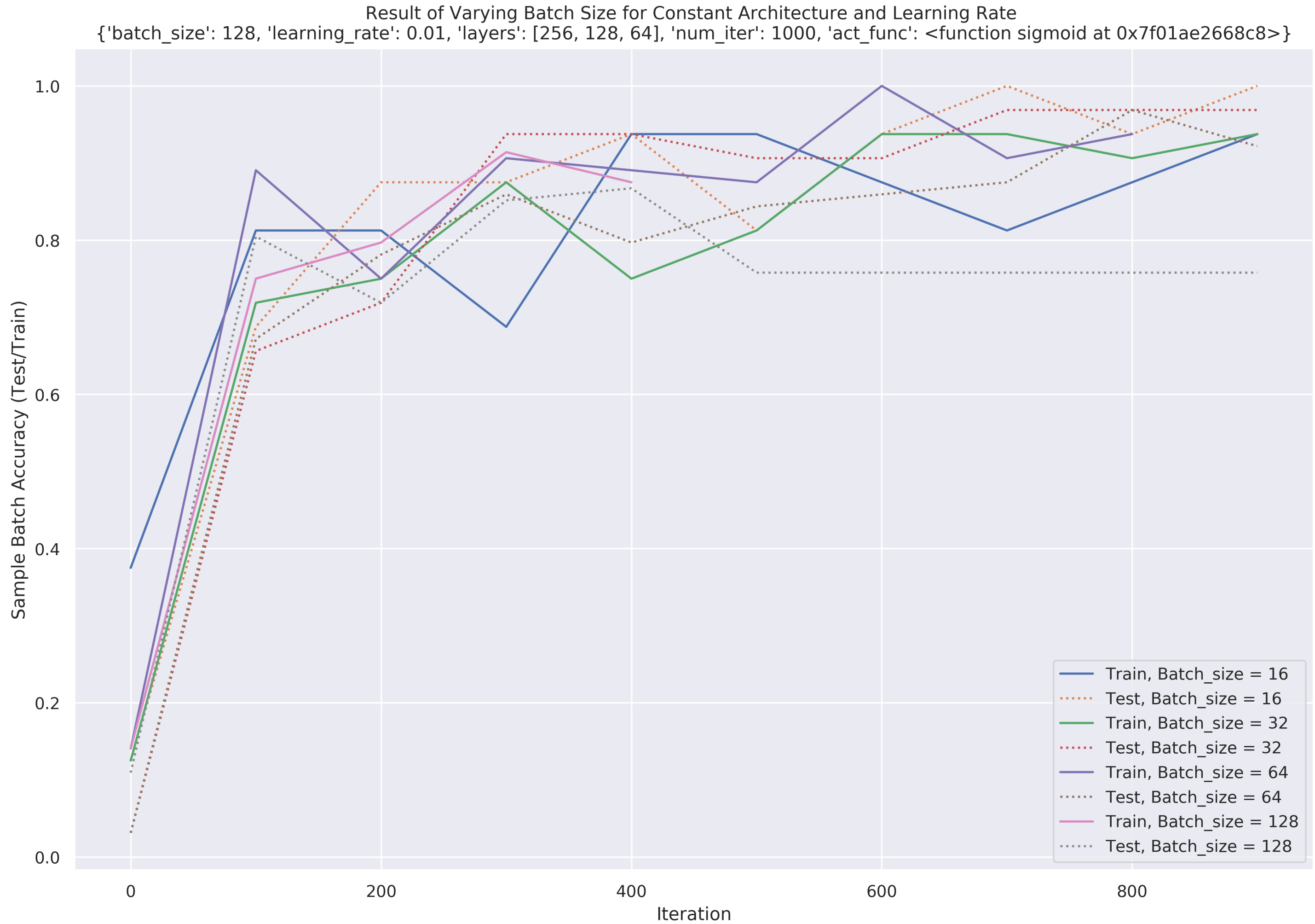

Figure 2: Impact of Varying Batch Size

1.4 Varying Activation Function

Fig (5) shows the differences between sigmoid, hyperbolic tangent, and RELU as layer activations. Sigmoid seemed the most performant.

1.5 Varying Number of Neurons in each Layer

Fig. (6) shows varied number of neurons for the two-layer architecture. 128 was found to be most performant.

1.6 Final Performance

Fig. (7) shows the final architecture and hyperparameters ran 10 times, as the performance is slightly dependent on initialization. Performance over 96% is often achieved.

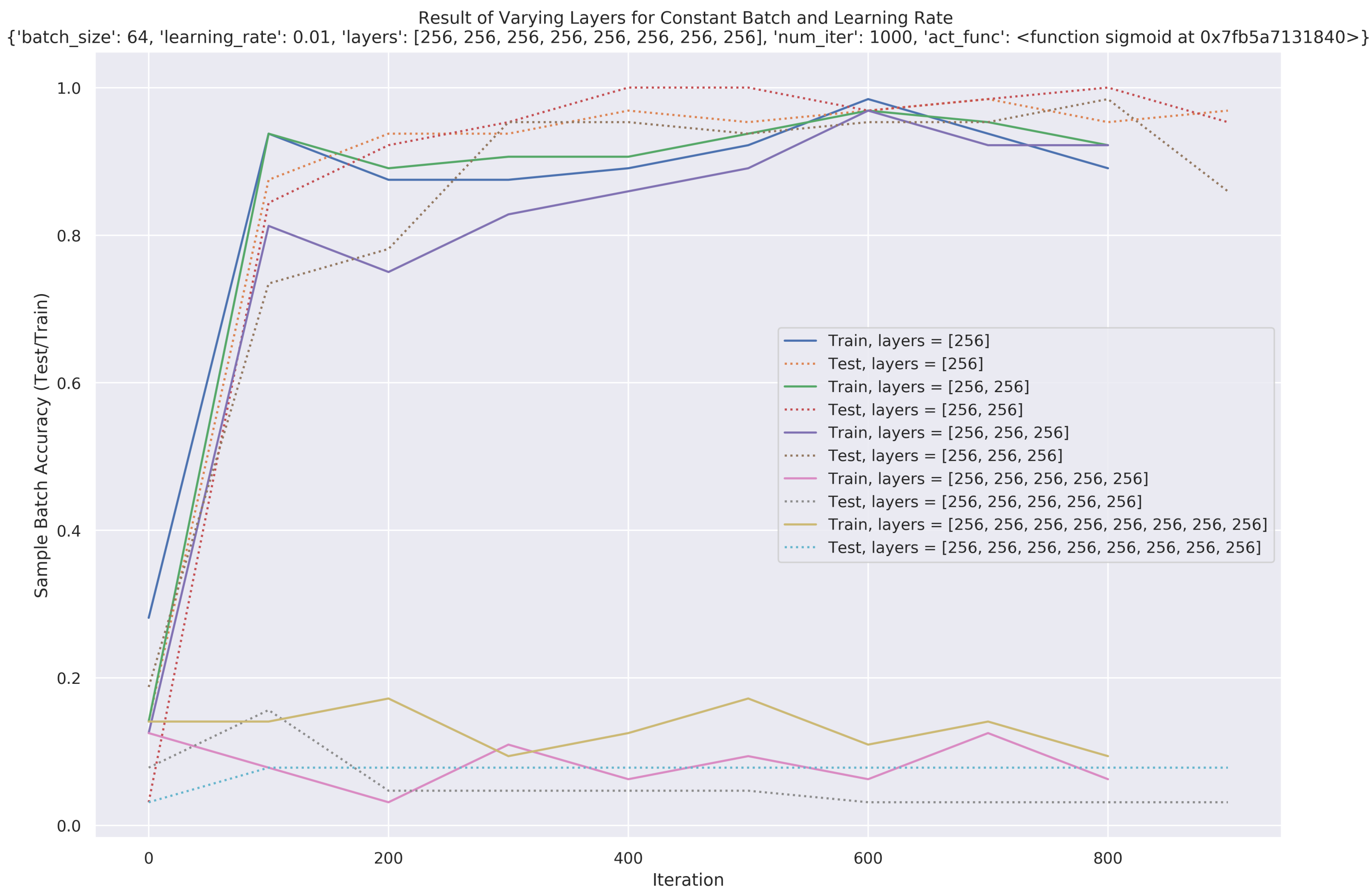

Figure 3: Varying Layer Size

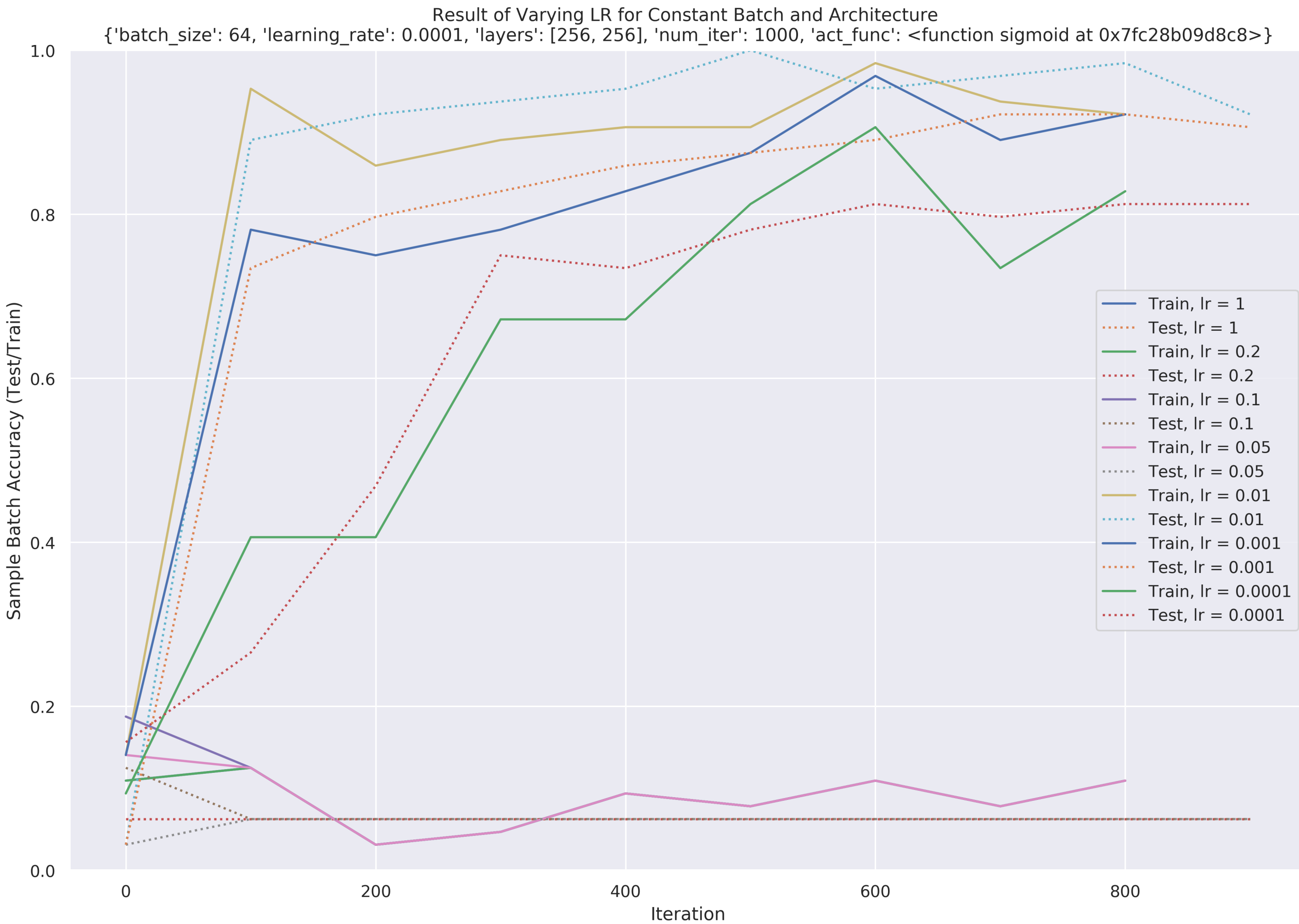

Figure 4: Impact of varying learning rates

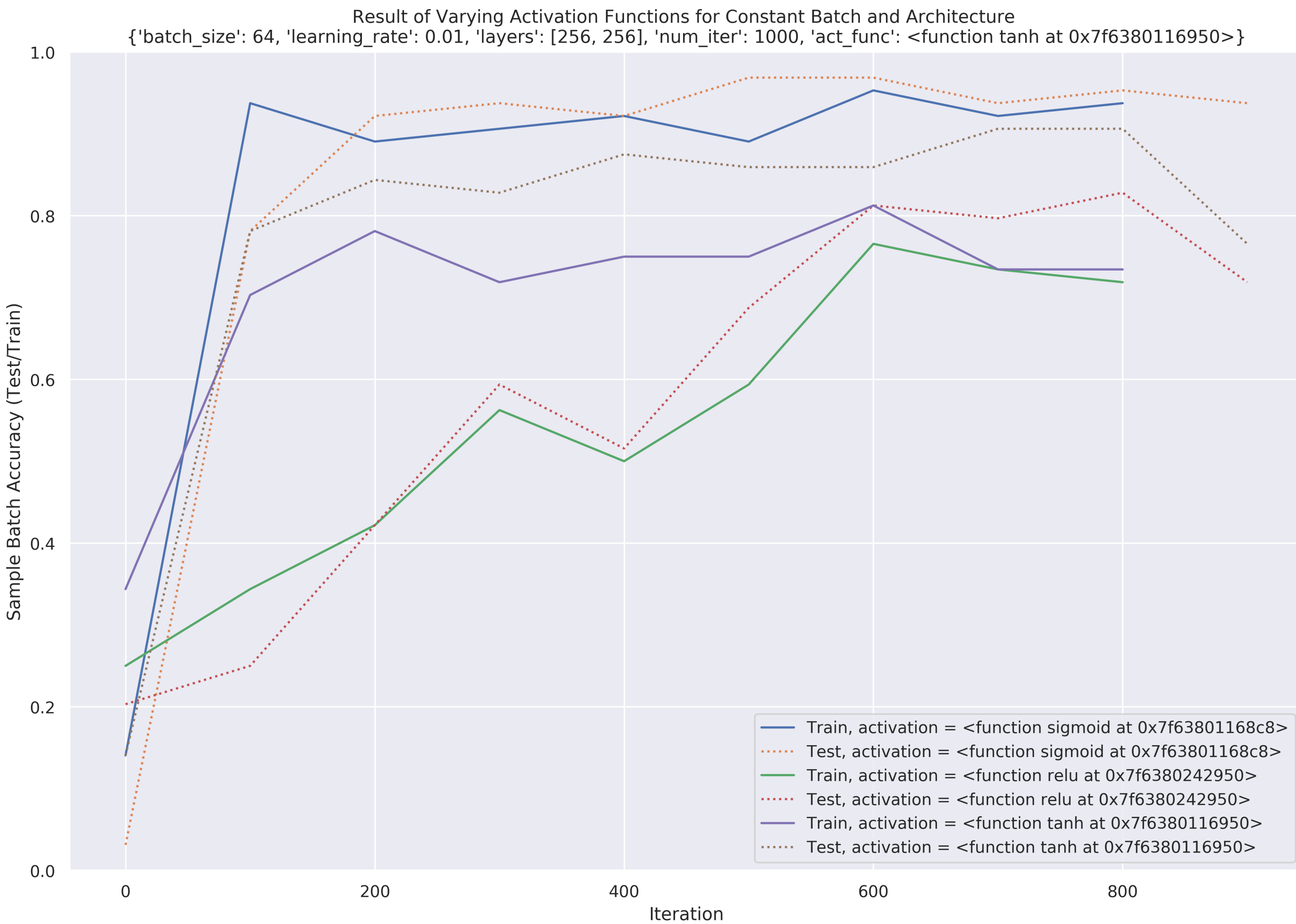

Figure 5: 3 Different Activation Functions

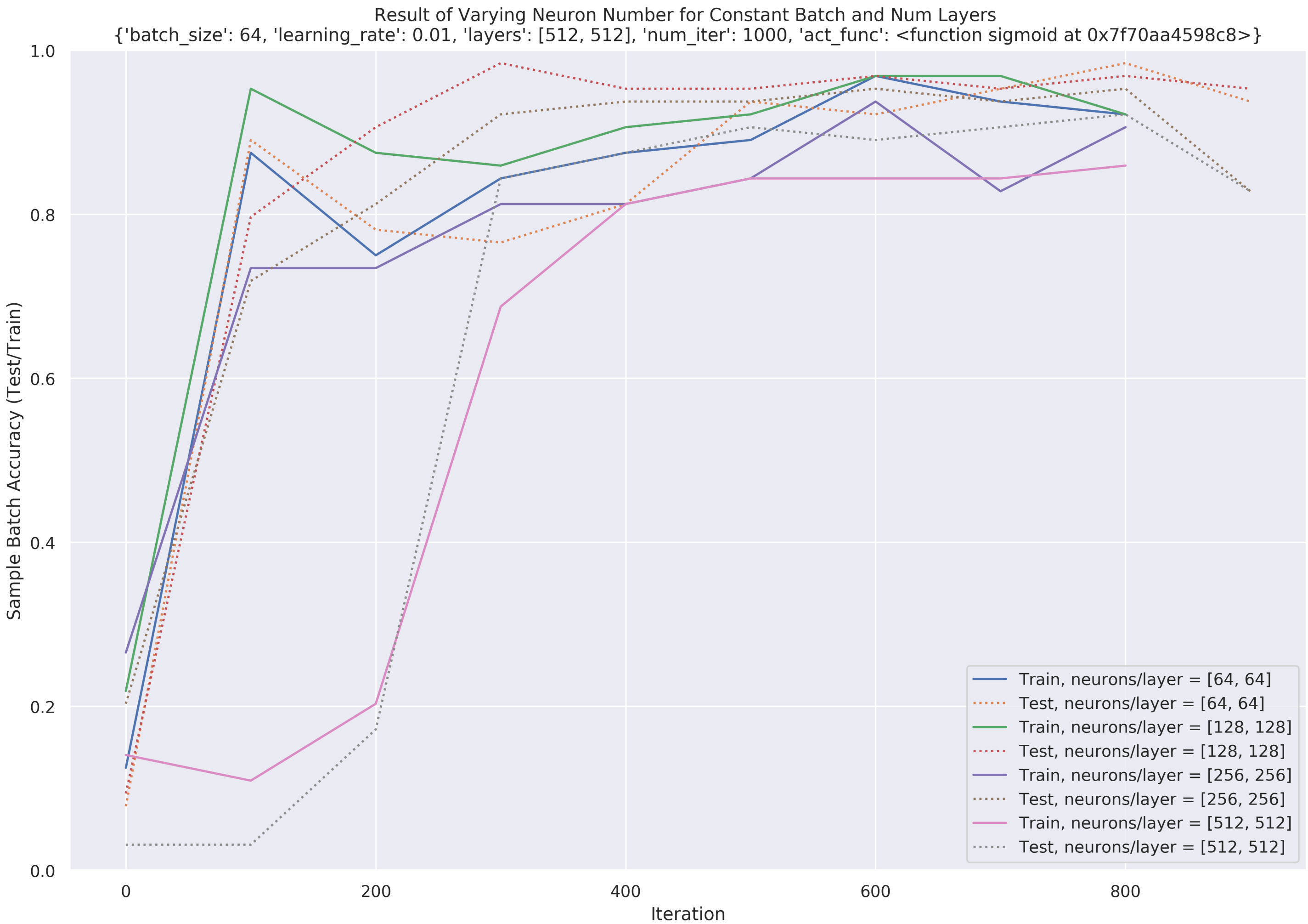

Figure 6: Varied Neuron Count per Layer

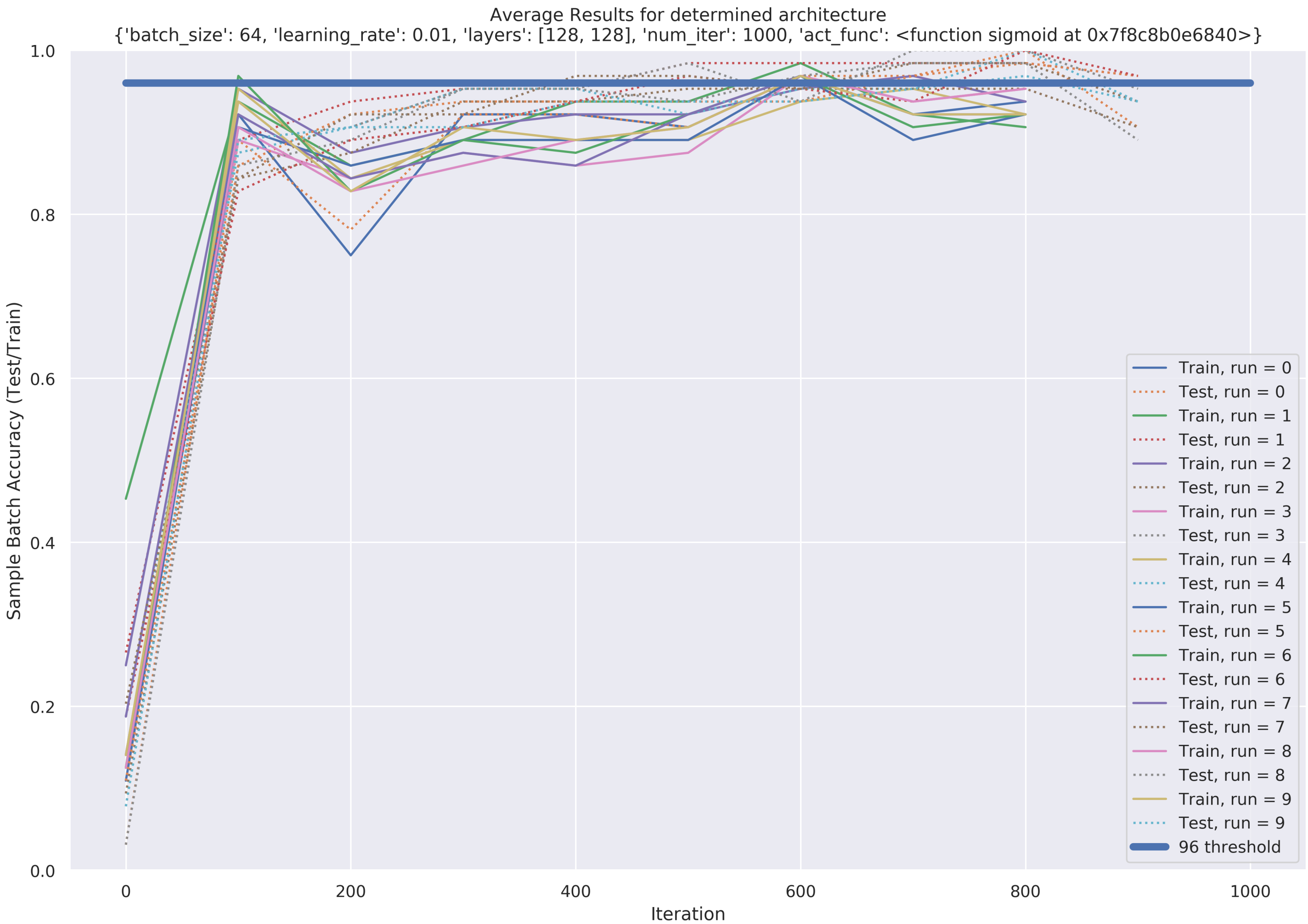

Figure 7: Final Architecture Performance