Three Layer Deep Neural Network(Sigmoid)

Epochs/ Batch Size	7/128	5/64	3/32
Loss	0.385	0.442	0.691
Accuracy	0.827	0.881	0.894

The initial neural network is satisfactory, error lowers as epochs as raised which is a good sign..
Accuracy is lowering as epochs raise but batch size rising could also contribute to this problem.
Analyzing all combinations of epochs and batch sizes would be advisable in the future to further
optimize the learning.
Sigmoid raises vanishing gradient issues…high error is probably due to this.
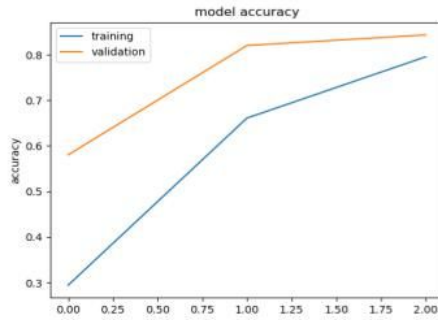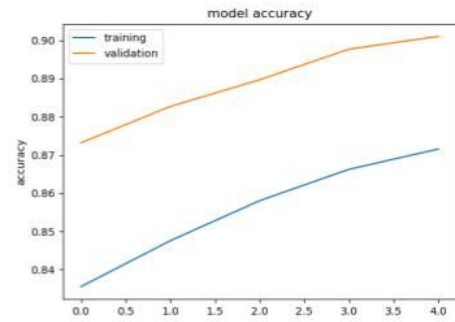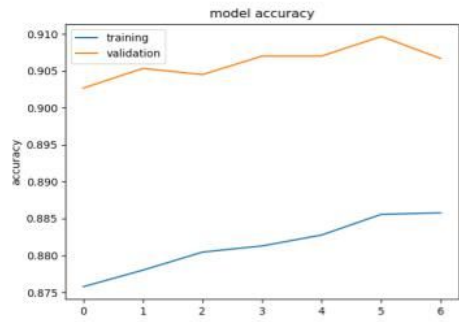
Three layer Deep Nueral Network (1000 layer in the middle)

Epochs/ Batch Size	7/128	5/64	3/32
Loss	0.683	0.435	0.382
Accuracy	0.818	0.881	0.892

Both accuracy and error fluctuate to an undesirable state. The 1000 x 1000 layer is overfitting the our data which is leading to higher error and lower accuracy.
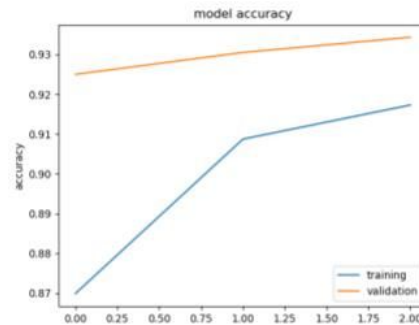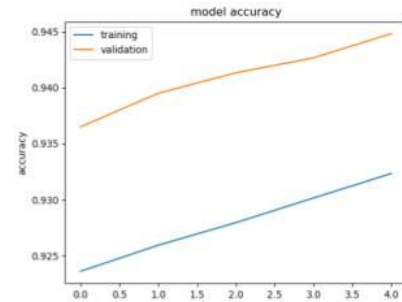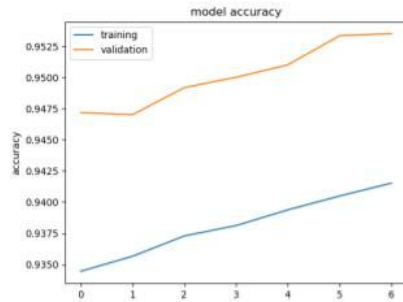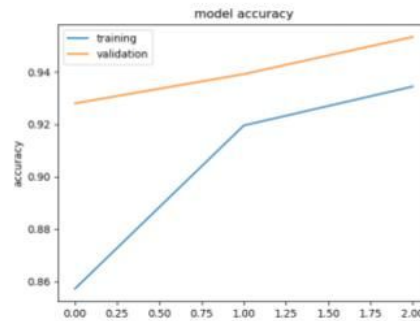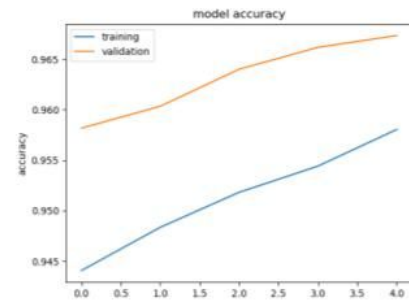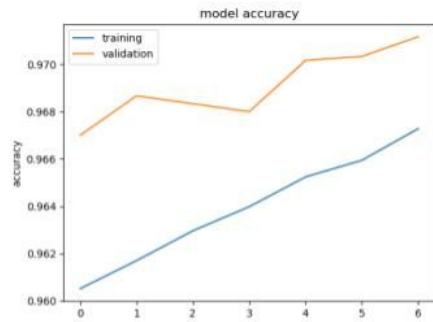
Three Layer deep Neural Network (tanh)

Epochs/ Batch Size	7/128	5/64	3/32
Loss	0.209	0.234	0.265
Accuracy	0.94	0.934	0.925

3_32_3 0.265 0.925
5_64_3 0.234 0.934
7_128_3 0.209 0.94

Tanh helps deal with vanishing gradient problem… about 14 percent decrease and percent increase in accuracy and error, respectively.
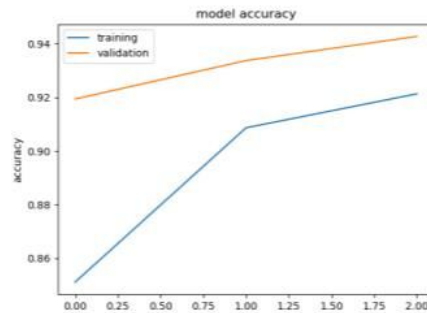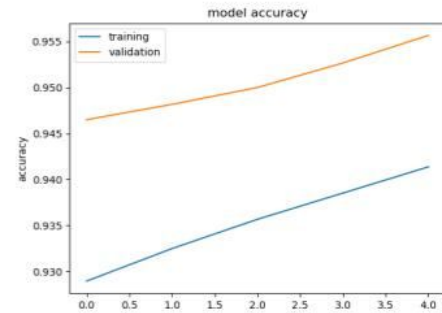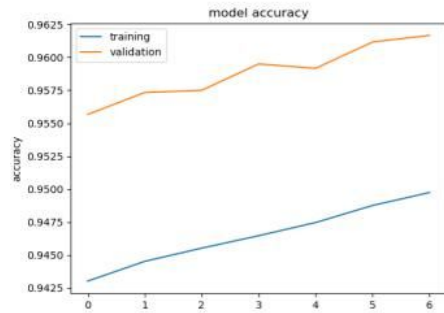
Three layer deep Neural Network(relu)

Epochs/ Batch Size	7/128	5/64	3/32
Loss	0.126	0.148	0.201
Accuracy	0.962	0.956	0.942

3_32_4 0.201 0.942
5_64_4 0.148 0.956
7_128_4 0.126 0.962

Relu further solves our vanishing gradient problem giving us our best results yet.

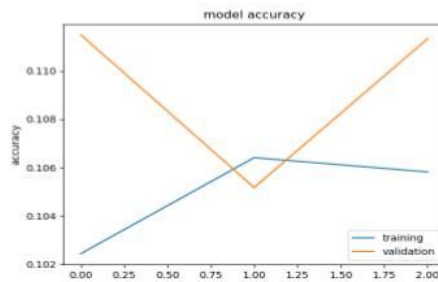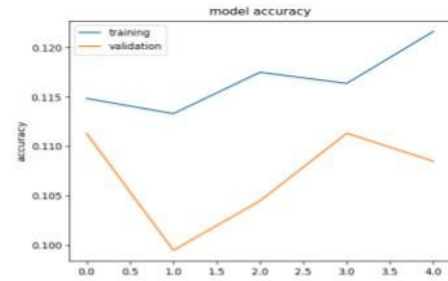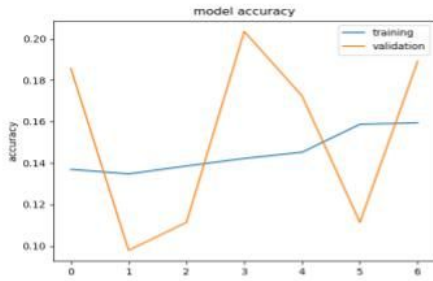*Three layer deep Neural Network(Middle layer is a residual layer)*

Epochs/ Batch Size	7/128	5/64	3/32
Loss	0.177	0.2	0.249
Accuracy	0.949	0.943	0.931

3_32_5 0.249 0.931
5_64_5 0.2 0.943
7_128_5 0.177 0.949

A good result, surprised by how the architecture of a residual layer. It was my first time working with it. In the future to optimize the threshold on my residual layer neural network would be advisable to get better loss and accuracy than the regular three layer relu network.

Five layer deep Neural Network

Epochs/ Batch Size	7/128	5/64	3/32
Loss	2.28	2.3	2.31
Accuracy	0.177	0.117	0.103

3_32_6 2.31 0.103
5_64_6 2.3 0.117
7_128_6 2.28 0.177
Expected result, data is severely overfitted.