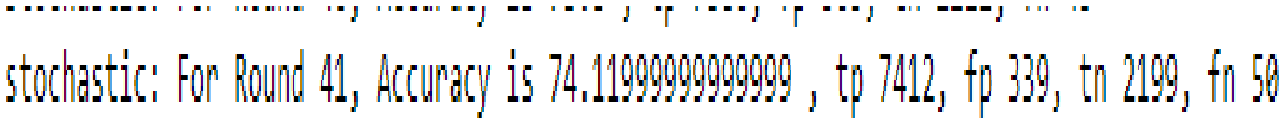
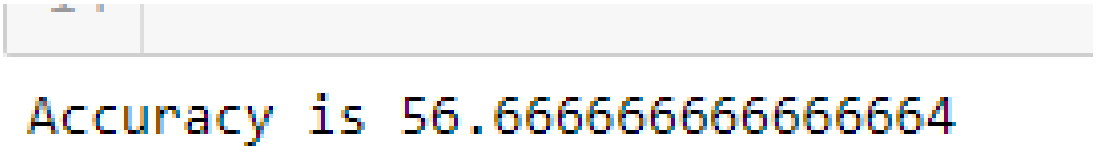
Bonus Task

Activation Function	Train Accuracy	Test Accuracy	Learn rate	Epoches	Layers	Hidden Nodes
Tanh	80%	74.1%	0.1	600	4 hidden 1 output 1 input	[5000,4000,3000,1000]
Sigmoid	11.5%	11.5%	0.1	600	4 hidden 1 output 1 input	[5000,4000,3000,1000]



Normal Task

Activation Function	Train Accuracy	Test Accuracy	Learn rate	Epoches	Layers	Hidden Nodes
Tanh	60%	56.6%	0.01	5000	3 hidden 1 output 1 input	[10,20,50]
Sigmoid	11.5%	11.5%	0.01	5000	3 hidden 1 output 1 input	[10,20,50]



Tansh is better in all causes

Sigmoid is bad because using it in all layers :

1. Squeeze the input and limit the output.
2. Cause Vanishing gradients which cause the accuracy to be so low.
3. Gradient instability and slow convergence.