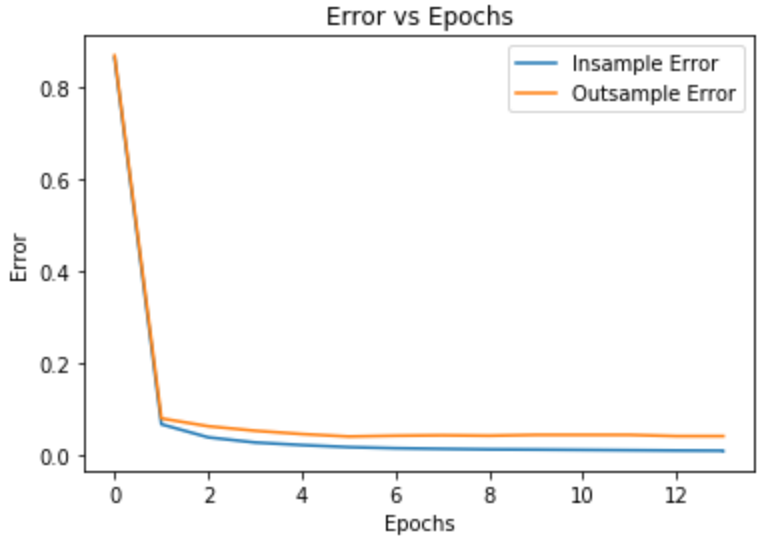
Part A

(A1) Sigmoid with learning rate = 0.1 and convergence if error is less than 0.015

Plot:

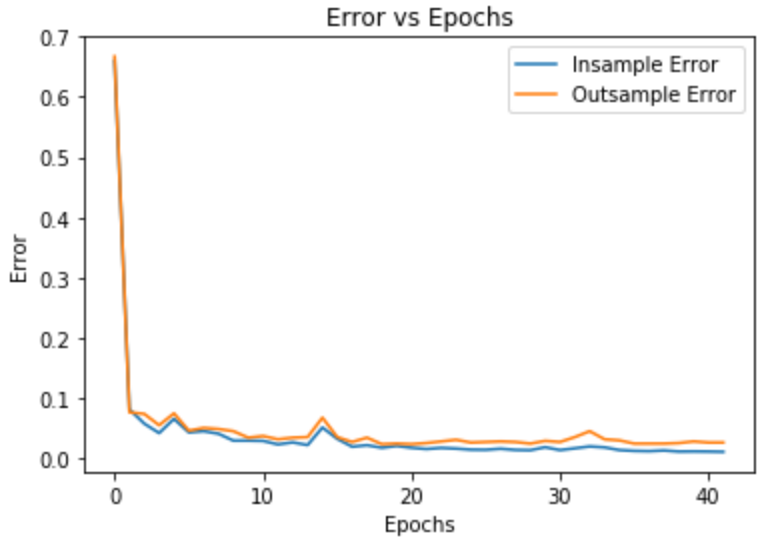


Optimal Number Of Iterations: 14

(A2) Tanh with learning rate = 0.1

Taking convergence if error is less the 0.05

Plot:



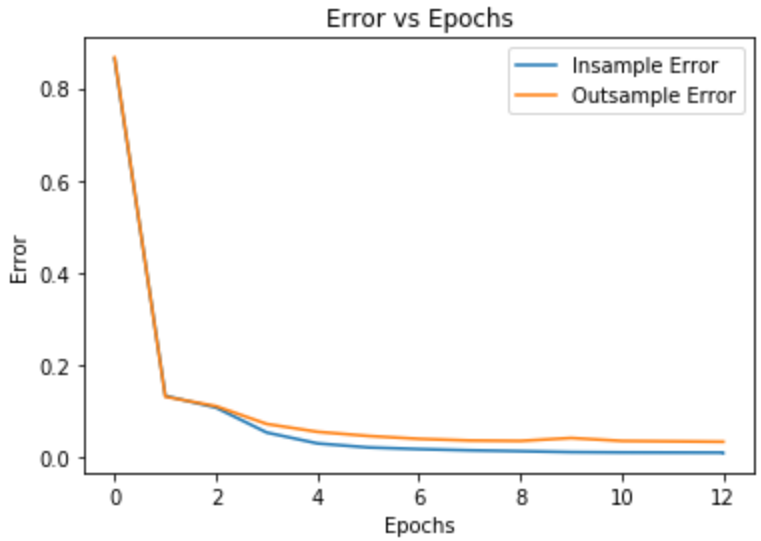
Optimal Number Of Iterations: 39

Result: Activation function “tanh” performs the best

Part B (Softmax)

Sigmoid activation with learning rate = 0.1 and convergence if error is less than 0.015

Plot:



Optimal Number Of Iterations: 13

It can be seen sigmoid converges way faster than tanh activation function.