Vectorization to avoid loops

Back props to decompose the derivative of complex function into many derivatives of simply functions

Python broadcasting principle

(m,n) \*//+/- (1,n) will copy (1,n) m times then do the operations

similar to (m,1)

don’t use something like

np.random.randn(5) # rank 1 array

but use

np.random.randn((5,1)) # 5X1 matrix

np.random.randn((1,5)) # 1X5 matrix

initial values of parameter should be randomized with small values due to char of activation functions