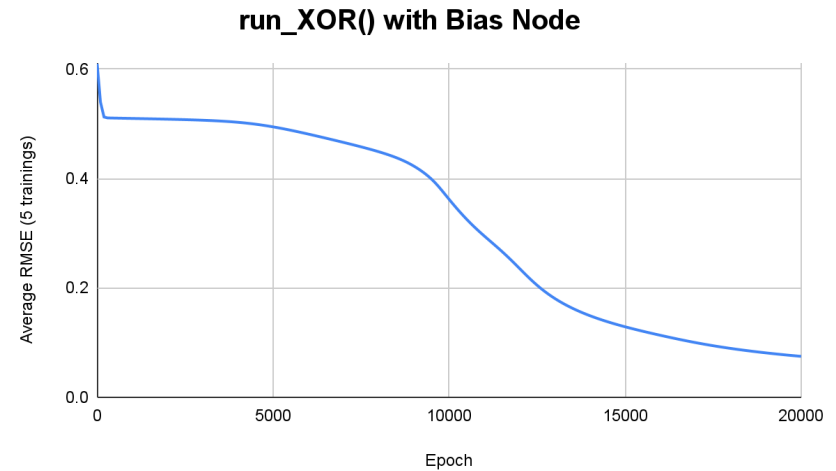
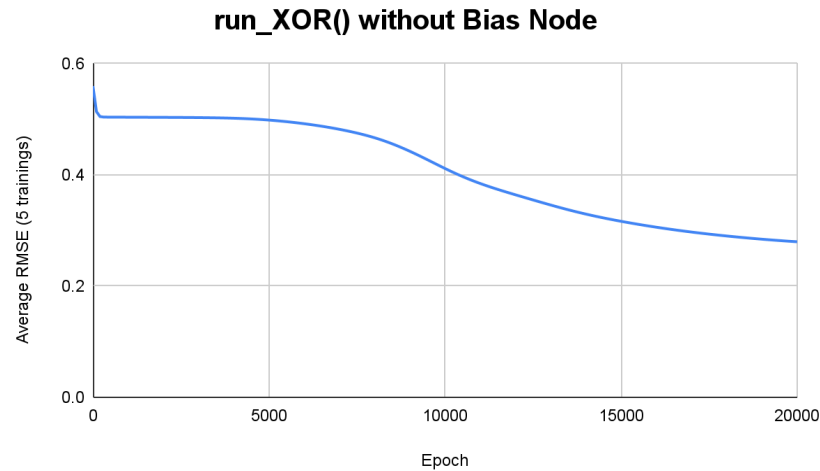


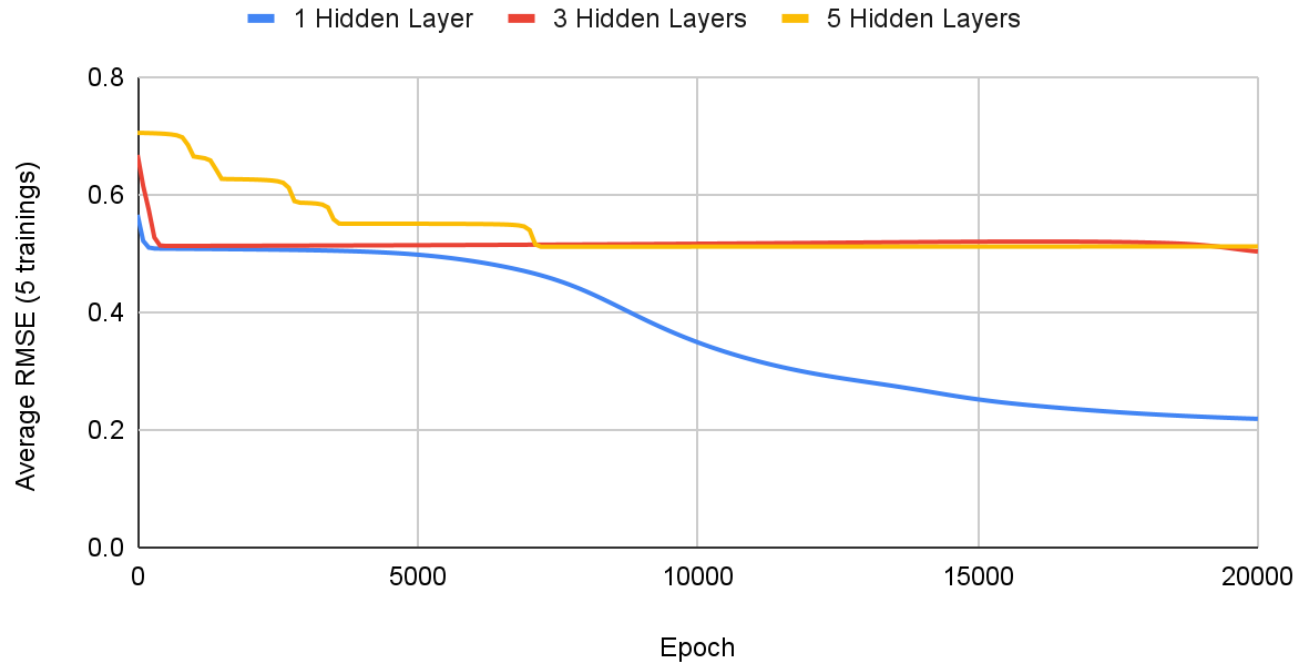
## Extra Credit Opportunity #1 - Bias Nodes



When training the XOR function with the bias node, the RMSE approaches zero more quickly than without the bias node. This is because the bias node enables learning since it serves as a constant input for all of the nodes in the hidden layer, which allows the neural network to differentiate between the hidden neurodes (even when all of the weights are zero), resulting in more efficient learning.

## Extra Credit Opportunity #2 - Vanishing Gradient (without bias node)

### 1 Hidden Layer, 3 Hidden Layers and 5 Hidden Layers



It seems like the more hidden layers a neural network has, the more likely the network is to get “stuck.” As indicated by the graph, when the network has 3 or 5 hidden layers, the RMSE remains fairly constant, starting around the 200th epoch for 3 hidden layers and the 7300th epoch for 5 hidden layers (suggesting that the network got “stuck”). On the other hand, the RMSE continues to decrease from the 4000th and up to the 20,000th Epoch when the network has 1 hidden layer (and the RMSE could potentially decrease more, based on the slope).