

1 Neural Network Architecture for Figure 2

Figure 2a - Fully connected neural networks with varying depths: In order to examine the impact of neural network depth and activation function differentiability on convergence, we run experiments using depth D fully connected neural networks for different values of D . These networks consist of $D - 1$ fully connected + activation blocks, concluding with a classification head.

Figure 2b - Convolutional neural networks: In order to validate our conclusions with more complex architectures, we performed simulations using a streamlined version of VGG16. This adapted model consists of 5 convolutional blocks, each of which includes convolution, activation, and average pooling layers. We chose average pooling instead of max pooling due to the latter's non-differentiability.