Q1. Neuron and Back Propagation

Gradient of L2 loss on sigmoid function vanishes quickly after several layers of back-propagation, while cross-entropy loss does not. So L2 loss should not be used in neural networks with deep layers, because when the gradient goes to zero, there is nothing can be trained anymore.