Neuron Output Inference Attack

Neurons correlated with sensitive outcome

Input Features

256 neurons

Hidden Layer 1

256 neurons

Hidden Layer 2

classes

Pre-Softmax Layer

Softmax Output
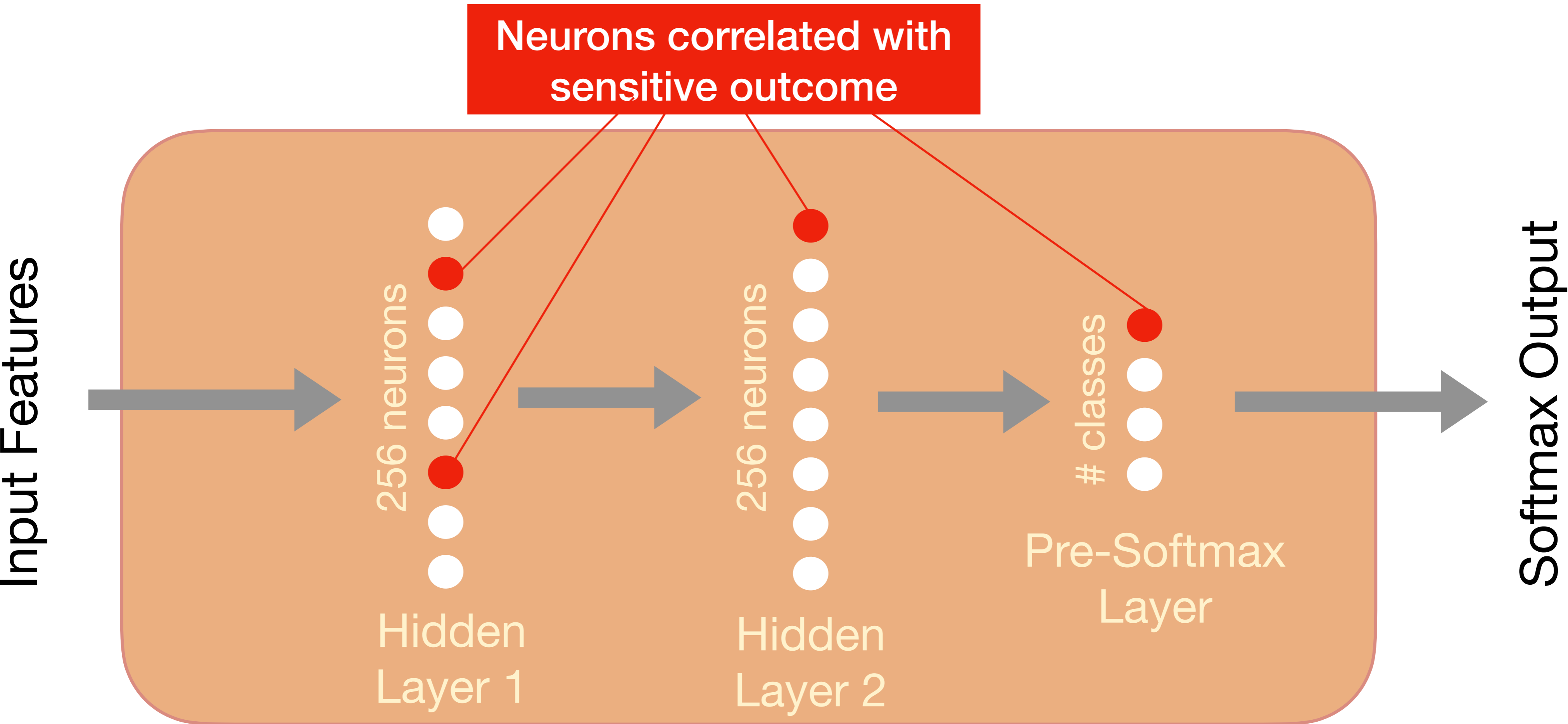
Neural Network Model Architecture