Learning Rate 10^{-2} learning rate 10^{-3} 0 2 6 8 epochs