Weight-Decay λ =0.01, Mini-Batch Size B=500, Training Rate η = 0.005 2.2 Training Rate Cross-Entropy Loss 2.0 1.8 -Training Loss (Cross-Entropy Loss) 1.6 1.4 1.2 1.0 8.0 0.6 100 200 300 400 500 600 700 0 # of Epochs