Weight-Decay λ =0.01, Mini-Batch Size B=500, Training Rate η = 0.01 Training Rate Cross-Entropy Loss 2.00 1.75 Training Loss (Cross-Entropy Loss) 1.50 1.25 1.00 0.75 0.50 0.25 100 200 300 400 500 600 700 0

of Epochs