Weight-Decay  $\lambda$ =0.01, Mini-Batch Size B=500, Training Rate  $\eta$  = 0.0001 Training Rate 2.2 Cross-Entropy Loss 2.0 Training Loss (Cross-Entropy Loss)
1. 1. 8 1.2 100 200 300 400 500 600 700 0

# of Epochs