Weight-Decay λ =0.01, Mini-Batch Size B=500, Training Rate η = 0.001 Training Rate SGD Cross-Entropy Loss Adam Optimizer Cross-Entropy Loss Training Loss (Cross-Entropy Loss) # of Epochs