Weight-Decay λ =0.01, Mini-Batch Size B=500, Training Rate η = 0.1

