Convergence of training, Loss sampled every 250 epoch

Training Error