HW1 Report - Umang Garg

Q-1. Clean accuracy for the standard model is 94.4%. Clean accuracy for the robust model is 80.8%

Q-2. (USING UNTARGETED CE LOSS) The accuracy of FGSM attack on the standard model is 55.7%. The accuracy of FGSM attack on robust model is 76.2%

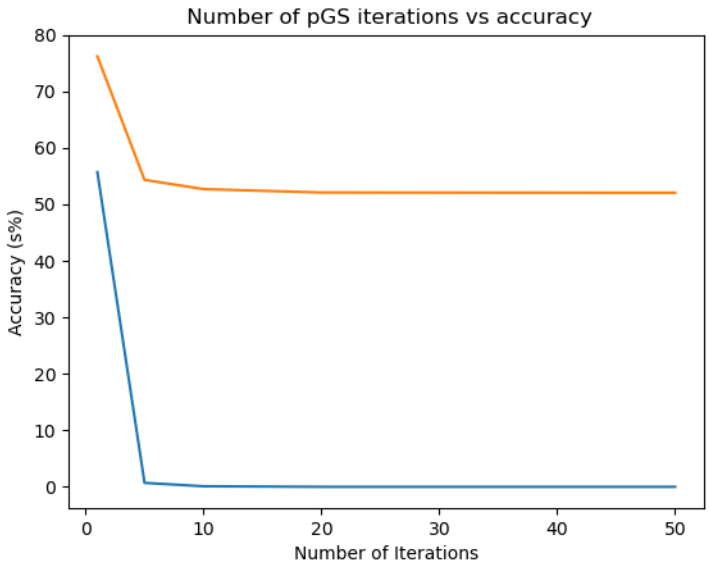
Q-3. (Untargeted PGD attack CE loss) The accuracy for the PGD attack on the standard model is 0.1%. The accuracy of the PGD attack on robust model is 52.7%

Q-4. (Untargeted PGD attack with CW loss) The accuracy for the PGD attack on the standard model is 0.02%. The accuracy for the PGD attack on robust model is 69.2%

Q-5. (Targeted PGD attack with CW loss) The accuracy for the PGD attack on the standard model is 4%. The accuracy for the PGD attack on robust model is 74.71%

Q-6. Plots

Plot- 1: Varying PGD attack iterations effect on accuracy with untargeted CE loss



Plot- 2: Varying PGD attack, epsilon effect on accuracy with untargeted CE loss