Model on MNIST Accuracy: [0.9912]

Epsilon: 0.1 Test Accuracy = 8744 / 10000 = 0.8744

Epsilon: 0.2 Test Accuracy = 4874 / 10000 = 0.4874

Epsilon: 0.5 Test Accuracy = 575 / 10000 = 0.0575

As epsilon increases, the accuracy significantly drops, however it also becomes more difficult to correctly classify as a human. The preturbation made the images a lot more noisy which is what we would expect. An epsilon of 0.2 really reduced the performance of the model but was still recognizable.

Part b

Couldnt tell you what happened during the training process. The sample images look as if they were perturbed more than the original. For some reason the training accuracy was very poor when trying to use the adversarial images. (implementation incomplete)

Part c