Backdoor Defenses

Norm Bounding

Server ensures that norms of participants' updates are within a threshold

Updates can still leak information about the training data

Weak DP

Norm bounding + Adding gaussian noise

Exploits Composition Theorem

Backdoor Defenses

Norm Bounding

Server ensures that norms of participants' updates are within a threshold

Updates can still leak information about the training data

Weak DP

Norm bounding + Adding gaussian noise

Exploits Composition Theorem

Motivation