Reasoning about “privacy” in ML

Most privacy attacks in ML focus on inferring either:

1. Inclusion of a data point in the training set (aka “membership inference”)

2. What class representatives (in training set) look like (aka “model inversion”)

101
1101
01101
110101
110001
1001
Data Leakage

2

Reasoning about “privacy” in ML

Most privacy attacks in ML focus on inferring either:

1. Inclusion of a data point in the training set
   (aka “membership inference”)

2. What class representatives (in training set) look like
   (aka “model inversion”)

1. Membership Inference