Active Property Inference Attack

Multi-task learning! Create a model that:

1) Predicts task label

2) Predicts property

In other words...

Adv can actively bias the model to leak property by sending crafted updates

Active Property Inference Attack

Multi-task learning! Create a model that:

1) Predicts task label

2) Predicts property

In other words...

Adv can actively bias the model to leak property by sending crafted updates

Active Property Inference Attack

Gradients on
Main Loss

$L(x, y; \theta)$

Main Task
(Gender)

Layer N

Layer 2

Layer 1

Upload local
updates

Server

Gradients on
Joint Loss

$\alpha * L(x, y; \theta) + (1 - a) * L(x, p; \theta)$

Main Task
(Gender)

Inference Task
(Facial ID)

Layer N

Layer 2

Layer 1

Download
joint model

35