



Attacker

Data point



Target model

Predicted label



Attack classifier

Member / Non-member

Label

