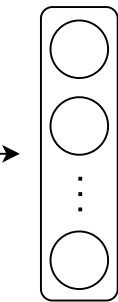


Attacker

Data point



Target model



Prediction
(probabilities)



Attack classifier

Member / Non-member

Label

