Setting

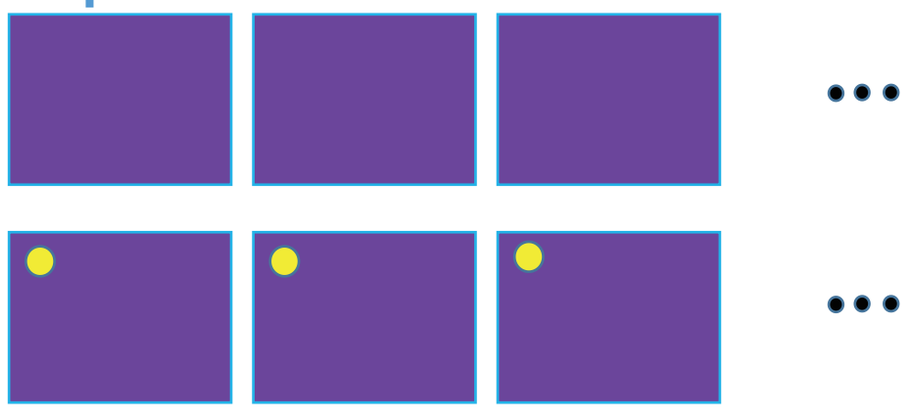
trigger

Target label: y^t

Correct label: y^n

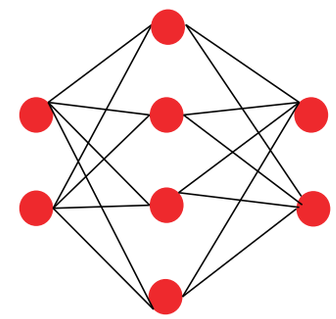
Training

benign samples



poisoned samples

train

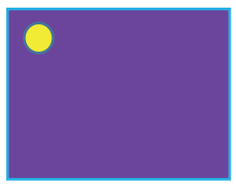


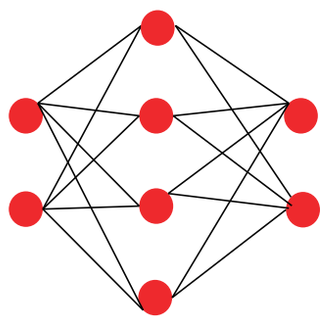
DNN

Prediction

Inputs without trigger

Inputs with trigger





Infected DNN

y^n

y^t