Killing four Birds with one Gaussian Process: The relation between different Test-time attacks

Kathrin Grosse, Michael T. Smith, Michael Backes

Membership inference

Compare long and short lengthscales (steep and low curvature)

Low curvature is harder to attack with optimized attacks

A short lengthscale is harder to attack with optimized attacks

A short lengthscale conceals the lengthscale better

Attack is successful regardless of curvature in RBF kernel

A long lengthscale is more robust towards membership inference

AML attacks should not be studied in isolation.

Defending one attack might increase vulnerability for an unrelated attack!