Adversarial Machine Learning (AML)

El uso cada vez más masivo de técnicas de Machine Learning pone de manifiesto importantes déficits en la seguridad de los algoritmos: perturbaciones sutiles de los datos de entrada a sistemas de aprendizaje automático, son capaces de hacer que estos hagan predicciones completamente erróneas. Recientemente, ha surgido el AML, que tiene por objetivo proteger a los sistemas basados en Machine Learning de estas perturbaciones. El paradigma predominante en AML, modeliza la confrontación entre algoritmos y adversarios dispuestos a perturbar los datos, a través de la teoría de juegos. Desde el ICMAT, trabajamos en un marco probabilista más general para el AML.

2 Estudiantes

Tutor: David Ríos Insua