Robustness guarantees for neural networks

Marta Kwiatkowska (Oxford University)

Host: Krishnendu Chatterjee

Deep neural networks have achieved impressive experimental results in image classification, but can surprisingly be unstable with respect to adversarial perturbations, that is, minimal changes to the input image that cause the network to misclassify it. With potential applications including perception modules and end-to-end controllers for self-driving cars, this raises concerns about their safety. This lecture will describe progress with developing techniques to establish safety and probabilistic safety guarantees for deep neural networks, giving an overview of methods based on games, global optimisation and statistical approximation.

Monday, March 30, 2020 04:00pm - 05:00pm
IST Austria Campus Raiffeisen Lecture Hall, Central Building

This invitation is valid as a ticket for the IST Shuttle from and to Heiligenstadt Station. Please find a schedule of the IST Shuttle on our webpage: https://ist.ac.at/en/campus/how-to-get-here/ The IST Shuttle bus is marked IST Shuttle (#142) and has the Institute Logo printed on the side.