ID de Contribution: **8** Type: **Non spécifié**

Building explainable and robust neural networks by using Lipschitz constraints and Optimal Transport

jeudi 20 juin 2024 09:30 (1 heure)

The lack of robustness and explainability in neural networks is directly linked to the arbitrarily high Lipschitz constant of deep models. Although constraining the Lipschitz constant has been shown to improve these properties, it can make it challenging to learn with classical loss functions. In this presentation, we explain how to control this constant, and demonstrate that training such networks requires defining specific loss functions and optimization processes. To this end, we propose a loss function based on optimal transport that not only certifies robustness but also converts adversarial examples into provable counterfactual examples.

Orateur: SERRURIER, Mathieu (IRIT Toulouse)

Classification de Session: Exposé long

Classification de thématique: Exposés longs