Universität Konstanz

From Risky to Reliable: Machine Learning with Guarantees

Prof. Dr. Aleksandar Bojchevski

Friday, 19 April 2024, 08:15, ZT 1202

From healthcare to natural disaster prediction, high-stakes applications increasingly rely on machine learning models. Yet, most models are unreliable. They can be vulnerable to manipulation and unpredictable on inputs that slightly deviate from their training data. To make them trustworthy, we need provable guarantees. In this talk, we will explore two kinds of guarantees: conformal prediction and robustness certificates. First, we will discuss how to equip models with prediction sets that cover the true label with high probability. The size of these conformal sets reflects the model's uncertainty. Then, we will derive certificates that guarantee stability under worst-case adversarial perturbations, focusing on the model-agnostic randomized smoothing technique. To conclude, we will provide an overview of guarantees for other trustworthiness aspects such as privacy and fairness.

Faculty of Sciences
Department of Computer and Information Science