ID de Contribution: **1** Type: **Non spécifié**

Information-Theoretic Methods in Data Sciences: Model Uncertainty, Robustness and Model Drift

mercredi 26 janvier 2022 10:20 (40 minutes)

Deep learning models are known to be bad at signalling failure: These probabilistic models tend to make predictions with high confidence, and this is problematic in real-world applications to critical systems such as healthcare, self-driving cars, among others, where there are considerable safety implications, or where there are discrepancies between the training data and data at testing time that the model makes predictions on. There is a pressing need both for understanding when models predictions should (or should not) be trusted, detecting out-of-distribution examples, and in improving model robustness to adversarial and natural changes in the data. In this talk, we will give an overview of those fundamental problems and key tasks. Namely, we first examine model uncertainty and calibration, and then we discuss simple but still effective methods for detecting misclassification errors and out-of-distribution examples, and for improving robustness in deep learning. We will describe information-theoretic concepts from fundamentals to state-of-the-art approaches, by going into a deep dive into promising avenues and will close by highlighting open challenges in the field.

Orateur: Prof. PIANTANIDA, Pablo (L2S/CentraleSupélec)