Post Doc position at CHL in Statistical Learning

With Magalie Fromont, IRMAR, Univ Rennes

MMD and Wasserstein distances based tests and GANs

Keywords :

Kernel methods, Reproducing Kernel Hilbert Space, Maximum Mean Discrepancy, Wasserstein distance, MMD GAN, Wasserstein GAN.

Presentation of the project :

The mathematical foundations of artificial intelligence and machine learning are at the interface of several fields : statistics, probability, calculus, optimisation, algorithmic complexity, but also algebra, geometry, graph theory, or game theory for instance, depending on the purpose and the framework of the intended application and the nature of the available data.
The present postdoctoral project is in the field of theoretical statistics, but with an important implementation component.

A comprehensive review of the different hypotheses tests recently introduced in learning theory, such as Maximum Mean Discrepancy (MMD) tests coming from kernel methods or Wasserstein distance based tests coming from the optimal transport theory, with a special focus on two-sample, independence or conditional independence tests, is expected as preliminary work.

Then, we will investigate the links between these tests and generative adversarial networks (GAN). These links can be analysed at several levels. On the one hand, two-sample tests or the associated classification algorithms, possibly coupled with independence tests, can be used to evaluate ou improve GANs' performance (see Lopez-Paz, Revisiting Classifier Two-Sample Tests for GAN Evaluation and Causal Discovery, ICLR 2017). On the other hand, GANs can be used to feed some tests needing large data sets, typically in high dimensional frameworks (see Bellot, van der Schaar, Conditional Independence Testing using Generative Adversarial Networks, arXiv :1907.04068), or to construct new testing procedures in the spirit of resampling techniques (see Shi et al., Generative Adversarial Networks for Conditional Independence Testing, arXiv :2006.02615). We will particularly deal with these last procedures, with a detailed study of their theoretical and experimental properties on simulated and real data, in sport science, neuroscience, genomic or epidemiology according to the center of interest of the selected candidate.

The selected candidate will be integrated in the statistics team of the IRMAR french CNRS research laboratory, which is composed of about forty statisticians working in two universities (Université Rennes 1 and Université Rennes 2), and three other institutions (INSA Rennes, ENSAI, Agrocampus Ouest) in Rennes.