Towards Fair and Reliable Machine Learning: Statistical Foundations and Challenges under Uncertainty

Paula Gordaliza, Institute for Advanced Materials and Mathematics (INAMAT^2)
Universidad Pública de Navarra (UPNA)

The talk addresses recent advances on the statistical foundations of fair and reliable machine learning, structured around two complementary research directions: achieving fairness in learning algorithms and preserving fairness guarantees under uncertainty and distributional shift.

The first direction focuses on the design of methodologies to mitigate bias in predictive models, both through data preprocessing and fairness-aware learning. At the data level, techniques based on optimal transport and distribution trimming are developed to remove the influence of sensitive variables while retaining relevant predictive information. At the model level, Fair Kernel Regression and Fair Partial Least Squares (PLS) incorporate fairness constraints directly into the optimization framework, relying on covariance operators and reproducing kernel Hilbert space representations to construct predictors that balance accuracy and independence from protected attributes, even in nonlinear or high-dimensional settings.

The second direction explores fairness under uncertainty, where limited, heterogeneous, or evolving data may challenge the stability of fairness guarantees. This line focuses on quantifying and propagating uncertainty in fairness assessment, developing Bayesian inference tools to evaluate whether observed disparities are significant or attributable to random variation. These approaches aim to provide more reliable and interpretable fairness analyses, ensuring that fairness evaluations remain valid under data perturbations and sampling variability.