

---

## SEMINARIO

Lucas de Lara

*Institut de Mathématiques de Toulouse*

### ***Transport-based Counterfactual Explanations***

**Abstract:** Counterfactual models aim at explaining machine learning predictions by answering questions of the form "had the input been different, would the output have been the same?" The literature mostly focused on two divergent frameworks to build these models: the nearest counterfactual instance principle -which is oblivious of the latent statistical distributions, and Pearl's causal reasoning -which relies on a model unknown in practice. We extend a recent work that substituted causal interventions by a push-forward operator between probability distributions. We show that this approach leads to statistically consistent counterfactual explanations, free of prior assumptions on the data generation process.

**Seminario del IMUVa, Edificio LUCIA**  
**Martes 20 de Abril de 2021 (11:30)**  
**Organiza: G.I.R. PEM**

