"Data Science for Large and Unstructured Data"

Morning sessions - David Rossell (Universitat Pompeu Fabra)

The morning sessions will deal with regression problems where there are many covariates. We will introduce motivating examples, relevant considerations to obtain reliable inference, and high-dimensional regression methods based on LASSO and Bayesian methods. Throughout we focus on strategies that attempt to guarantee reproducible findings via the Statistical quantification of uncertainty, and on providing high-level intuition over technical details. We will briefly point out extensions to treatment effect estimation and more flexible methods such as classification trees.

9:30 – 11:00 Session 1

11:30 – 13:00 Session 2

Afternoon sessions - Stephen Hansen (Imperial College Business School)

The afternoon sessions will introduce methods for unstructured datasets such as text. It begins with non-statistical approaches for extracting information from such data, including dictionary methods, matrix factorization, and word embeddings. It then introduces statistical models for text that extract topical content and basic ideas in inference for estimating them. Throughout we will provide examples of how these methods have been used in the empirical economics literature.

14:00 – 15:30 Session 3

16:00 – 17:30 Session 4

Afternoon sessions – Job Market entry

18:00-18:30 Job market (Pedro Rey, ESADE)

18:30-19:00 Teaching (Humberto Llavador, Universitat Pompeu Fabra)