Title: Model-based Machine Learning for Transportation

Aim and Scope:

This tutorial is designed for transport engineers and related professionals looking to perform advanced data analysis in their future research or practice. Model-based Machine Learning is based on a powerful framework called Probabilistic Graphical Models (PGMs), that allows the combination of domain knowledge with data driven methods in a very simple way.

It expects two key ingredients: basic knowledge of statistics and probability and, ideally some basic knowledge of programming.

While Machine Learning has plenty of algorithms (e.g. Neural Networks, Gaussian Processes, Support Vector Machines, Decision Trees, etc.) that have the benefit of being “push-button” solutions, they are generally very hardly adaptable beyond the original design. Therefore, our task typically becomes about transforming our problem and data to fit each individual algorithm. Many times, we drop relevant information about the domain (e.g. known relationship between 2 variables, different noise distributions in input variables), and our results may suffer from it.

PGMs allow us to include prior knowledge, parametric and non-parametric (sub)-models, and uncertainty about inputs and parameters. They are perfect to combine different types of data, and, in the past few years, a growing community has developed tools for PGMs, that simplify its design and inference process. Together with Deep Learning, PGMs belong to the forefront of Machine Learning and Data Mining research, essential to process Big and Small Data.

In this tutorial, we will cover the main concepts of model-based Machine Learning, introduce the most recent research directions in the state-of-the-art and present several practical applications in the transportation domain, which will serve as case studies for motivating the concepts covered by this tutorial.

Outline of topics and expected duration:

- Introduction to Model-based Machine Learning (20 min)
- Foundations of Probabilistic Graphical Models (PGMs) and Bayesian inference (40 min)
- Generative models and introduction to STAN (40 min)

BREAK

- Bayesian linear regression, Poisson regression, Heteroscedastic regression (40 min)
- Bayesian additive models (20 min)
- Bayesian logistic regression, hierarchical models (30 min)
- Autoregressive models, Linear dynamical systems, Multivariate state-space models (30 min)

Intended length:

Half-day

Intended audience and assumed background knowledge:

Everyone; assumed basic knowledge of statistics and probability and, ideally, some basic knowledge of programming

Computer and software requirements:

None

Organizers

Name: Francisco C. Pereira
Institutional address: Technical University of Denmark (DTU), Bygning 115, 2800 Kgs. Lyngby, Denmark
Email: camara@dtu.dk

Name: Filipe Rodrigues
Institutional address: Technical University of Denmark (DTU), Bygning 115, 2800 Kgs. Lyngby, Denmark
Email: rodr@dtu.dk