Optimal Transport and applications to Machine Learning

Marcello Carioni (University of Twente)

Optimal Transport (OT) is a mathematical theory introduced by Gaspard Monge in 1781 to study the optimal allocations of resources and goods.
Its original formulation, called Monge formulation, aims at finding the best way to *transport* a probability distribution $\mu \in P(\mathbb{R}^n)$ to another $\nu \in P(\mathbb{R}^n)$ by minimizing the transport value computed with respect to a given cost $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$. Mathematically, this consists in finding the map $T : \mathbb{R}^n \to \mathbb{R}^n$ that minimizes the value

$$\inf_{T:\mathbb{R}^n \to \mathbb{R}^n \ s.t.\ T_\# \mu = \nu} \int c(x, T(x))\, d\mu(x) \tag{0.1}$$

where $T_\# \mu$ is the push-forward of the probability μ by T. The optimal map T provides the cheapest way to move mass from the measure μ to the measure ν. Such formulation allows for great flexibility since it includes both discrete, semi-discrete and continuous formulations. Moreover, the cost c can be chosen to enforce desired properties, such as constraining the transport to specific regions of the domain or favouring concentration of mass.
Due to its flexibility and mathematical rigour, in the 20th century, significant theoretical advancements were made and the discipline gained relevance and found noteworthy applications in fields such as economics, urban planning, image processing and biology. Even more notably, in the last ten years, optimal transport approaches have been used to solve machine learning tasks and to design better data-driven algorithms. This is not surprising at all: an important part of modern machine learning methods relies on estimating the distance between data distributions in a fast and accurate way, and Optimal Transport, as apparent in (0.1), provides a natural way to compare probability distributions, by looking at how expensive is to transport one to the other one. This observation, together with recent algorithms able to compute optimal transports incredibly fast, has made OT approaches of central importance in the construction of new generative models, in the resolution of inverse problems and in the enhancement of robustness for neural networks. In this series of lectures, we plan to cover the following topics:

- We start with a basic introduction of Optimal Transport, outlining its classical formulations and necessary results we need for the remaining part of the course.
- We discuss the entropic regularization of optimal transport and present the Sinkhorn algorithm, able to compute the (regularized) solution to an optimal transport problem efficiently.
- We talk about the connections between OT and Machine Learning. We focus on adversarial generative models based on Optimal Transport (WGAN, WAE) and, if time permits, we discuss how to use optimal transport approaches to solve inverse problems.

References:

Filippo Santambrogio, *Optimal Transport for Applied Mathematicians*

Gabriel Peyré and Marco Cuturi, *Computational Optimal Transport* https://arxiv.org/abs/1803.00567