

MATH-676

Information Theory & Optimal Transport

Invited lecturers (see below)

Cursus	Sem.	Type
Mathematics		Opt.

Language of teaching	English
Credits	2
Session	
Exam	Term paper
Workload	60h
Hours	24
Courses	12
TP	12
Number of positions	

Frequency

Only this year

Remark

to be given in Spring 2027

Summary

The course develops some of the mathematical foundations of generative modeling: divergences and metrics from information theory, optimal transport. In order build stable algorithms to learn distribution from finite data proximal regularizations and mean-field game theory are used.

Content

The course develop some of the mathematical foundations needed to understand, build, and improve generative modeling algorithms such as generative adversarial networks, diffusion models, normalizing flows and Schrödinger bridges. One of the theme of the course is emphasize the common mathematical thread behind all these different models.

We first study the properties and the variational representations of three main mathematical players, Information Divergences (e.g Kullback Leibler-divergence), Optimal Transport Costs (Wassertein metrics), and Integral Probability Metrics (e.g. Kantorovich-Rubinstein metrics and maximum mean discrepancy (kernel distance) and neural integral probability metrics). These representations lead to (a) optimization problems in function spaces (Gibbs variational principle), (b) the construction of probability flows (Wasserstein gradient flows, Fokker-Planck equation, JKO scheme), (c) optimization over probability flows such as in Benamou-Bremier theorem and Schrodinger bridges.

A theme of the course is that, in order to build (numerically) stable algorithms to learn distributions from finite data sets, one need to supplement the cost functional from Information theory and Optimal transport with suitable regularizations. This leads to new family of divergences which incorporate the desirable properties from different objects, for example information divergence and integral probability metrics or information divergences and optimal transport. One such example is entropic optimal transport and Schroedinger bridges. Another family of examples are new family of divergences which we build using convex analysis methods, in particular proximal regularizations in probability spaces. Another fundamental aspect of generative model is that they can be viewed as control problems and we will develop the detailed connection between generative flows and mean-field games which allows to bring partial differential equations (coupled Hamilton-Jacobi and Fokker-Planck equations).

As the class progress we will build the algorithms behind a variety of learning algorithms such as adversarial learning (for generative adversarial networks and probability flows), backward-forward training (for normalizing flows, neural ordinary differential equations, and score diffusion models), proximal algorithms (such as JKO scheme) and Sinkhorn algorithm for Schrödinger bridges. We will emphasize the common mathematical structure behind them.

Keywords

Optimal Transport, Information Theory, Proximal Regularizations, Diffusion Models, Generative Flows, Generative Adversarial Networks.

Learning Prerequisites

Required courses

A good working knowledge of analysis, probability theory, and some basic PDEs.

Learning Outcomes

By the end of the course, the student must be able to:

- Use the tools information theory and optimal transport to understand generative models.

Resources

Moodle Link

- <https://go.epfl.ch/MATH-676>