

HARVARD CS 2840: COMPUTATIONAL OPTIMAL TRANSPORT FOR MACHINE LEARNING

Fall 2025

Instructor:	David Alvarez-Melis	Time:	Tu/Th 12:45 - 2:00pm
Email:	dam@seas	Place:	114 Western Av 2.112
TFs:	Alexandru Meterez (ameterez@fas) & Jaeyeon Kim (jaeyeon.kim@fas)		

Course Description: Optimal Transport (OT) has quickly become an important part of the Machine Learning (ML) toolkit, where it has been used for various purposes, from learning mappings between datasets to proving convergence results for neural network training. This advanced topics course will cover the mathematical foundations and computational aspects of OT in the context of machine learning applications. Foundational topics covered will include: Monge and Kantorovich formulations of OT, duality, entropy regularization, Gromov-Wasserstein distances, dynamic formulations, and Benamou-Brenier theory. Each of these will be presented in the context of their use for specific problems in machine learning, including domain adaptation, generative models, correspondence analysis, and gradient flows. The course combines theoretical rigor with practical exercises, implementing OT algorithms and analyzing their impact on real-world machine learning challenges. The classes will consist of a combination of instructor-led lectures on fundamentals and student-led discussion of relevant academic papers. Interested students should have a solid foundation in linear algebra, probability, and machine learning principles.

Enrollment: This is a limited-enrollment, seminar-style course. Enrollment priority will be given to SEAS PhD students and to those whose background aligns most closely with the course's hard and soft prerequisites. If requesting enrollment, please describe your relevant background and motivation for taking the course in the comments section.

Prerequisites: Basic knowledge of linear algebra & calculus (e.g. AM22a, Math 21b, or equivalent), and probability theory (e.g. STAT 110 or equivalent) is expected. Familiarity with machine learning (e.g. via CS 181) and optimization (e.g. APMTH 121) fundamentals is highly encouraged. Prior exposure to Python programming will be beneficial but is not mandatory.

Office Hours: Tuesdays 2:00pm-3:00pm @ SEC 2.332 (David); Thursdays 2:00pm-3:00pm @ SEC 2.415 (Alex); Fridays @ SEC 2.415 (Jay).

Materials and Resources The lecture part of the course will roughly follow:

- G. Peyré and M. Cuturi. “Computational Optimal Transport”. In: *Foundations and Trends in Machine Learning* 11 (5-6 2019), pp. 355–607

Additional required readings will be assigned throughout the course. These readings will include both textbook chapters and research papers, providing students with a broad understanding of optimal transport's theoretical underpinnings and practical applications in machine learning.

Objectives: By the end of this course, students will be able to:

- Understand the historical context and key concepts of optimal transport
- Recognize problems in machine learning that can be approached with optimal transport, and propose methods to solve them
- Apply optimal transport methods to tackle real-world machine learning problems
- Critically assess the effectiveness of optimal transport methods in different machine learning contexts

Relation to other courses, at Harvard and beyond: This course delves deeper into a specific area of machine learning, focusing on the theory and applications of optimal transport. This course complements the broader machine learning principles taught in CS 1810. It is recommended for students who have already taken CS 1810 or have a strong foundation in machine learning and want to specialize further.

Other relevant courses and their difference with this course are discussed below:

- Compared to APMTH 220: *Geometric Methods for Machine Learning*, this course will be more narrowly focused on a specific family of methods (Optimal Transport) but shares the “geometric approach” to Machine Learning.
- Compared to MIT’s 18.657: *Statistical Optimal Transport*, this course will focus less on statistical aspects of OT and more on its application to Machine Learning problems, computational considerations, and empirical performance.
- Additionally, this course will have some minimal overlap with other courses that cover Transport Phenomena, Differential Equations, and Stochastic Processes, such as ENG-SCI 123: Introduction to Fluid Mechanics and Transport Processes and PHYSICS 181: Statistical Mechanics and Thermodynamics.

Tentative Course Outline:

1. Optimal Transport Fundamentals (*~weeks 1 to 3*)

- (a) Historical Overview and Introduction to Optimal Transport
- (b) Mathematical foundations of Optimal Transport: metric spaces and probability measures
- (c) Monge’s and Kantorovich’s formulations, Kantorovich-Rubinstein duality
- (d) Discrete and continuous formulations and connections between them
- (e) Metric properties of the Wasserstein Distance
- (f) Dynamic formulation of Optimal Transport and its relation to gradient flows and partial differential equations

2. Computational Aspects of Optimal Transport (*~weeks 4 to 6*)

- (a) Linear programming approaches
- (b) Regularization techniques (e.g., entropic regularization) and Sinkhorn’s algorithm
- (c) Complexity analysis of optimal transport problems
- (d) Computational Estimation of Monge Maps (neural-based, plug-in estimators)

3. Generalizations (*~weeks 7 to 8*)

- (a) Gromov-Wasserstein distance and its variants (Fused Gromov-Wasserstein, Gromov-Monge)
- (b) Unbalanced and relaxed formulations
- (c) Approximations: sliced distances, low-rank versions

4. Optimal Transport for Machine Learning (*~weeks 9 to 13*)

- (a) General application settings: generative modeling (Wasserstein GAN, OTGAN, etc.), transfer learning, domain adaptation, dataset compression, representation learning and dimensionality reduction, connections to Schrödinger bridges and denoising diffusion models
- (b) Applications in Natural Language Processing: embedding alignment, machine translation, language modeling
- (c) Applications in Computer Vision: image retrieval, texture mixing, and style transfer

- (d) Applications in Computational Biology: medical imaging, single-cell sequencing data
- (e) Applications on Graphs and Networks: approximate graph isomorphism, graph alignment
- (f) Open problems and emerging trends

Grading Policy: Assessment for this course will be based on:

- Problem sets (3, 10% each): These assignments will draw on course material to challenge students to apply optimal transport concepts and techniques in various contexts. These will include theoretical and practical (coding) components.
- Class participation (10%): Students are expected to lead at least one in-class discussion, attend and participate actively in all classes.
- Lecture note editing (10%): Students must sign up to scribe one lecture. Depending on enrollment, this will involve collaborative editing with other students.
- A final project (50%; 10% for proposal, 30% for report, 10% for presentation): For their final project, students will either choose a specific machine learning problem and apply optimal transport methods to solve it, or develop an methodological extension of the optimal transport problem with potential applications in machine learning. They will then present their work in a written report and oral presentation.

In-person attendance will be required to receive the full participation credit.

Submission Logistics: All assignments will be submitted on Gradescope, so students should either have a scanner available or become familiar with LATEX. We will provide latex templates for all assignments. Collaboration is encouraged, but every student should write out their own solutions and make sure they understand everything they submit.

Late Submission Policy: Each student is allotted a total of **72 late hours** (i.e., 3 late days) for use on problem set submissions throughout the semester. These hours may be used in 1-hour increments at the student's discretion—no need to request permission in advance. Once this late budget is exhausted, submissions will still be accepted for up to 48 additional hours, but with a linear penalty applied to the score. Specifically, if a submission is h hours late beyond the allotted budget, the final score will be scaled as:

$$\text{Score} = \text{Raw Score} \times \max\left(0, 1 - \frac{h}{48}\right)$$

After 48 hours past the deadline (and after the late budget is exhausted), submissions will no longer be accepted. All late hours are rounded up to the next full hour. We encourage students to manage their time accordingly and reserve late hours for emergencies or unavoidable scheduling conflicts.

Generative AI Use Policy: This course permits the use of generative AI (GAI) tools, such as ChatGPT, as part of the learning process. **However**, students are strongly encouraged to first attempt all assignments independently, and to use GAI tools only as a secondary aid when they are truly stuck. GAI should be used thoughtfully and sparingly to support problem solving—not to replace it. Use of GAI for *polishing written text* (e.g., improving clarity or grammar) is allowed and encouraged. All use of GAI tools must be clearly acknowledged and cited in submitted work. Students are fully responsible for evaluating the correctness, relevance, and appropriateness of any AI-generated content they submit. Submitting AI-assisted work without meaningful engagement or misrepresenting AI-generated content as entirely one's own may be considered a violation of academic integrity. Please note that AI use policies may vary across Harvard courses. It is each student's responsibility to understand and follow the policy in every course they take.

Accommodation Requests: Harvard University values inclusive excellence and providing equal educational opportunities for all students. Our goal is to remove barriers for disabled students related to

inaccessible elements of instruction or design in this course. If reasonable accommodations are necessary to provide access, please contact the Disability Access Office (DAO). Accommodations do not alter fundamental requirements of the course and are not retroactive. Students should request accommodations as early as possible, since they may take time to implement. Students should notify DAO at any time during the semester if adjustments to their communicated accommodation plan are needed. Any accommodation requests should be pre-approved by your resident dean (undergraduates) or advisor (graduate students), and be sent via email to all teaching staff.