Course Title: ECE 6970: Statistical Distances for Modern Machine Learning
Author: Ziv Goldfeld, ECE
Authorship or Revision Date: 28/08/2019
Credit Hours: 3 hours
Course Website: http://people.ece.cornell.edu/zivg/ECE6970.html

Catalog Description:
Statistical distances such as optimal transport (particularly, the Wasserstein metric), total variation, Kullback-Leibler (KL) divergence, χ^2 divergence, and others, are used to design and analyze a variety of machine learning systems. Applications include anomaly and outlier detection, ordinal regression, generative adversarial networks (GANs), and many more. This course will establish the mathematical foundations of these important measures, explore their statistical properties (e.g., convergence rates of empirical measures), and focus on GANs and, more generally, on deep neural networks (DNNs) as applications (design and analysis).
The format is based on paper reading and presentation assignments performed by the students. Each student will present a work of hers/his choice from a prescribed list. The course instructor will deliver the first 3-4 lectures, as well as some throughout the semester. The final project will include a scientific assignment based on another chosen article. Choices for project assignments include extension of existing results, implementation tasks, critical summary of a paper, etc. The last 4 lectures will be dedicated to final project synopses presentations.

Course Offering and Frequency:
One-time offering; future openings based on demand.

Prerequisites:
Ph.D. Graduate Standing or permission of instructor.

Corequisites:
None.

Student Preparation Summary:
Knowledge of probability theory, functional analysis and machine learning fundamentals.

Textbook(s) and/or Other Required Materials:
The course is based on selected parts from books and scientific papers:

- Books/lecture notes:
  - C. Villani, “Topics in optimal transportation,” *American Mathematical Society*, 2003.
  - Y. Polyanskiy and Y. Wu, “Lecture notes on Information Theory, ” 2017.
- Scientific papers:
  - I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, “Generative adversarial nets,” In Proceedings of *Advances in Neural Information Processing Systems (NIPS)*, 2014.
  - M. Arjovsky, S. Chintala and L. Bottou, “Wasserstein GAN,” In Proceedings of the *International Conference on Machine Learning (ICML),* 2017.

- R. M. Dudley, "The speed of mean Glivenko-Cantelli convergence," *The Annals of Mathematical Statistics*, Vol. 40, No. 1, pp. 40-50, 1969.
- M. Cuturi "Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances," *Advances in Neural Information Processing Systems (NIPS)*, 2013.
- A. Genevay, L. Chizat, F. Bach, M. Cuturi and G. Peyré, "Sample complexity of sinkhorn divergences," *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.
- G. Mena and J. Weed, "Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem," *arXiv preprint (arXiv:1905.11882)*, 2019.
- Z. Goldfeld, K. Greenewald, Y. Polyanskiy and J. Weed, "Convergence of smoothed empirical measures with applications to entropy estimation," *arXiv preprint (arXiv:1905.13576)*, 2019.
- R. Shwartz-Ziv and N. Tishby, "Opening the black box of deep neural networks via information," *arXiv preprint (arXiv:1703.00810*), 2017.
- A. M. Saxe, Y. Bansal, J. Dapello, M. Advani, A. Kolchinsky, B. D. Tracey and D. D. Cox, "On the information bottleneck theory of deep learning," *International Conference on Learning Representations (ICLR)*, 2018.
- Z. Goldfeld, E. van den Berg, K. Greenewald, I. Melnyk, N. Nguyen, B. Kingsbury and Y. Polyanskiy, "Estimating information flow in deep neural networks," In Proceeding of the *International Conference of Machine Learning (ICML)*, 2019.
- M. I. Belghazi, A. Baratin, S. Rajeswar, S. Ozair, Y. Bengio, A. Courville and R. D. Hjelm, *"MINE: Mutual Information Neural Estimator,"* In Proceeding of the International Conference of Machine Learning (ICML), 2018.
- And (possibly) more…

Lecture, Section, and Laboratory Schedule:

Lectures: Tuesday and Thursday 2:55-4:10pm.

Each students will present a lecture based on their reading assignment. Several introductory and intermediate lectures will be given by course instructor.

Assignments, Exams and Projects:

Paper reading and presentation: 1 paper reading and presentation assignment per student.
Homework: 3-4 assignments per semester. Collaboration with students is encouraged.
Project: Scientific assignment based on a selected paper (from prescribed list or other related works). Possible assignments include extension of an existing result, critical summary, implementation + experiments, etc. Projects may be performed in pairs.
Project presentation: Up to 20 min presentation of project results.

Course Grading Scheme: 30% paper reading and presentation assignment, 20% homework assignments, 40% project, 10% project presentation.

List of Topics Covered:

- Introduction to optimal transport and Wasserstein distances
- Applications of optimal transport to GANs
- Convergence rate of empirical Wasserstein distance and curse of dimensionality

Note: Syllabus subject to change prior to course start.

- Entropic optimal transport
- Gaussian-smoothed optimal transport
- Other statistical distances (total variation, KL and χ^2 divergences) – convergence and smoothing
- Relations between statistical distances
- From statistical distances to information measures
- Information bottleneck principle for deep learning
- Estimating information flows in DNNs
- MINE: Mutual information neural estimator

Student Preparation Summary:

- Math: Students should be comfortable with probability theory. Familiarity with principles of function spaces, convex analysis and statistical estimation is helpful, although necessary background will be given in class.
- Machine learning: Students should be familiar with the supervised/unsupervised learning problem formulations and relevant techniques. In particular, knowledge of DNNs, variational autoencoders (VAEs) and GANs is helpful.
- Programming: Not mandatory, yet knowledge of `Matlab/Python/PyTorch/Tensorflow` would enable more practical final projects.

Academic Integrity:

Students expected to abide by the Cornell University Code of Academic Integrity with work submitted/presented for credit representing the student's own work. Collaboration on reading assignments and projects is permitted (and encouraged) but the final work should represent the student's own work and understanding. Course materials are intellectual property belonging to the authors. Students are not permitted to buy or sell any course materials, or distribute in any form, without the express permission of the instructor. Such unauthorized behavior will constitute academic misconduct.