ENEE 729T: Topics in Communication
Information Theoretic Methods in Learning
Spring 2021

TIME/VENUE

MW 3:30 – 4:50 pm EDT
ONLINE

INSTRUCTOR

Prakash Narayan
Room AVW 2353
Phone: (301) 405 3661
E-mail: prakash@umd.edu

OFFICE HOURS

MW 2:00 – 3:15 pm (But please, please make an appointment to avoid idling Zoom sessions.) Also at other times by appointment.

COURSE WEBSITE (`coming soon`)

http://www.ece.umd.edu/class/enee729T

OVERVIEW

The course covers an assortment of information theoretic methods of interest in statistical inference and learning. Topics include: (i) information geometry leading to the EM algorithm and application to maximum likelihood estimation; (ii) measure concentration methods; (iii) correlated multiarmed bandits including in parameter estimation and probability distribution learning from partially sampled observations (finding an arm or a small set of arms that yield information about other correlated arms); and (iv) data privacy *vs* function computation utility tradeoffs.

PREREQUISITES

ENEE 620 (Random processes) or equivalent, ENEE 627 (Information theory), or permission of the instructor.

COURSE OUTLINE

I. Information geometry

Kullback-Leibler divergence, Pythagorean inequality, I-projection on linear families, iterative algorithm for finding the minimum divergence between two convex sets of distributions, EM algorithm, application to maximum likelihood estimation.

II. Concentration of measure

Concentration inequalities as basic tools: Markov, Chebyshev, Chernoff bounding, Hoeffding, Bennett, Bernstein, Efron-Stein, Azuma; Entropy method, tensorization, Han's inequalities for entropy and divergence, log Sobolev inequalities, Vapnik-Chevronenkis dimension and entropy. Some applications.

III. Correlated multiarmed bandits

Prediction and parameter estimation problems will be considered in which an estimator is able to assess self-reward or -loss based on local measurements but is unaware of the consequences of estimation from unseen measurements. However, correlation among measurements affords a new latitude in learning. Stochastic and nonstochastic (or adversarial) multiarmed bandit problems will be considered.

IV. Data privacy

Various notions of data privacy: Differential privacy, distribution privacy, divergence-based privacy; data privacy vs function computation utility tradeoffs.

COURSE GRADE

The course grade will be determined on the basis of individual or (small) team projects and in-class presentations. Specifically, an individual's or a team's performance will be evaluated by means of a (i) a midterm assessment of progress on a term project, and (ii) a final assessment of the completed term project. A term project can consist of (a) work on a chosen or assigned topic involving open issues or (b) a critical examination of a pertinent topic or topics in the existing literature – in both cases combined with a comprehensive oral presentation at the end of the semester.

Problems to be addressed in a term project will be fixed at the end of approximately four (4) weeks into the semester.

REFERENCES

There is no required or recommended text. The course material will be drawn largely from a selection of books and publications that are listed below. Some of this material will be posted at the course website (marked by (*)).

0. Sine qua non

The following two classics are shining sources of enlightenment in information theory.

(*) I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic, Cambridge U Press, 2011. (`A PDF version of the first edition (1981) will be posted at the course website.`)

T.M. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

I. Information geometry

D. Barber, *Bayesian Reasoning and Machine Learning*, Chapter 11, Cambridge U Press, 2012.

(*) I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statistics and Decisions*, Suppl. Issue 1, pp. 205-237, 1984.

(*) I. Csiszár and P. Shields, "Information Theory and Statistics: A Tutorial," *Foundations and Trends in Communications and Information Theory*, Volume 1, Issue 4, 2004.

(*) A. Kumar and R. Sundaresan, "Minimization Problems Based on Relative α-Entropy I: Forward Projection," *IEEE Trans. Information Theory*, vol. 61, no. 9, pp. 5063 - 5080, September 2015.

(*) A. Kumar and R. Sundaresan, "Minimization Problems Based on Relative α-Entropy II: Reverse Projection," *IEEE Trans. Information Theory*, vol. 61, no. 9, pp. 5081 - 5095, September 2015.

S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning, From Theory to Algorithms*, Chapter 24, Cambridge U Press, 2014.

II. Concentration of measure

D. P. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms*, Cambridge U Press, 2009.

(*) S. Boucheron, G. Lugosi and O. Bousquet, "Concentration Inequalities," *Advances in Machine Learning*, O. Bousquet, U. v. Luxburg and G. Rätsch (Eds.), pp. 208 - 240, Springer, 2004.

(*) S. Boucheron, G. Lugosi and O. Bousquet, "Introduction to Statistical Learning Theory," *Advances in Machine Learning*, O. Bousquet, U. v. Luxburg and G. Rätsch (Eds.), pp. 169 - 207, Springer, 2004.

S. Boucheron, G. Lugosi and P. Massart, *Concentration Inequalities*," Oxford U Press, 2013.

(*) M. Raginsky and I. Sason, *Concentration of Measure Inequalities in Information Theory, Communications, and Coding*, 3rd Ed., Foundations and Trends in Communications and Information Theory, NoW, 2019.

(*) I. Sason and S. Verdú, "Arimoto-Rényi Conditional Entropy and Bayesian M-Ary Hypothesis Testing," *IEEE Trans. Information Theory*, vol. 64, no. 1, pp. 4 - 25, January 2018.

III. Multiarmed bandits

J. Y. Audibert, S. Bubeck and R. Munos, "Best Arm Identification in Multiarmed Bandits," *Conference on Learning Theory*, pp. 41 53, 2010.

(*) S. Bubeck and N. Cesa-Bianchi, "Regret Analysis of Stochastic and Nonstochastic Multiarmed Bandit Problems," https://arxiv.org/abs/1204.5721v2.

N. Cesa-Bianchi, and G. Lugosi, *Prediction, Learning, and Games*, Cambridge U Press, New York, 2006.

C. Guestrin, A. Krause, and A. P. Singh, "Near-Optimal Sensor Placements in Gaussian Processes," International Conference on Machine Learning, pp. 265 272, 2005.

(*) S. Gupta, S. Chaudhari, S. Mukherjee, G. Joshi and O. Yagan, "A Unified Approach to Translate Classical Bandit Algorithms to the Structured Bandit Setting," *IEEE Trans. Information Theory*, vol. 1, no. 3, pp. 840 - 853, November 2020.

(*) S. Gupta, G. Joshi and O. Yagan, "Best-Arm Identification in Correlated Multiarmed Bandits," *preprint*, 2020.

(*) K. Jamieson and R. Nowak, "Best-arm Identification Algorithms for Multiarmed Bandits in the Fixed Confidence Setting," 2014 48th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, 2014, pp. 1-6, doi: 10.1109/CISS.2014.6814096.

E. Kaufmann, O. Cappé and A. Garivier, “On the Complexity of Best Arm Identification in Multiarmed Bandit Models,” *Journal of Machine Learning Research*, 2015.

T. Lattimore and C. Szepesvári, *Bandit Algorithms*, https://tor-lattimore.com/downloads/book/book.pd

C. Y. Liu and S. Bubeck, “Most Correlated Arms Identification,” *Conference on Learning Theory*, pp. 623 637, 2014.

L. A. Prashanth, “Multi-Armed Bandits,” *CS 6046, Course notes*, March 2018.

IV. Data privacy

(*) S. Asoodeh, M. Diaz, F. Alajai and T. Linder, “Estimation Efficiency Under Privacy Constraints,” *IEEE Trans. Information Theory*, vol. 65, no. 3, pp. 1512 - 1534, March 2019.

(*) R. Bassily, A. Groce, J. Katz and A. Smith, “Coupled-Worlds Privacy: Exploiting Adversarial Uncertainty in Statistical Data Privacy,” FOCS 2013.

(*) C. Dwork, “Differential Privacy,” in Bugliesi M., Preneel B., Sassone V., Wegener I. (eds), *Automata, Languages and Programming, ICALP 2006, Lecture Notes in Computer Science*, vol 4052. Springer, Berlin, Heidelberg.

(*) C. Dwork, F. McSherry, K. Nissim and A. Smith, “Calibrating Noise to Sensitivity in Private Data Analysis,” *Journal of Privacy and Confidentiality* (2016-2017)7, Number 3, 1751.

(*) Q. Geng and P. Viswanath, “The Optimal Noise-Adding Mechanism in Differential Privacy,” *IEEE Trans. Information Theory*, vol. 66, no. 3, pp. 925 - 951, February 2016.

(*) Q. Geng, P. Kairouz, S. Oh and P. Viswanath, “The Staircase Mechanism in Differential Privacy,” *IEEE Journal Selected Topics in Signal Processing*, vol. 9, No. 7, pp. 1176 - 1184, October 2015.

(*) C. Huang, P. Kairouz, X. Chen, L. Sankar and R. Rajagopal, “Context-Aware Generative Adversarial Privacy,” *Entropy 2017, 19, 656*, pp. 1 - 35.

(*) I. Issa, A. B. Wagner and S. Kamath, “An Operational Approach to Information Leakage,” *IEEE Trans. Information Theory*, vol. 62, no. 2, pp. 1625 - 1657, March 2020.

(*) A. Nageswaran and P. Narayan, “Data Privacy for a ρ-Recoverable Function,” *IEEE Trans. Information Theory*, vol. 65, no. 6, pp. 3470 - 3488, June 2019.

(*) A. Smith, “Efficient, Differentially Private Point Estimators,” arXiv:0809.4794v1, Sept. 2008.