ELEG/FSAN 817 Large Scale Machine Learning

Credits: 3
Fall 2020
Meeting:TBD, **Location:** TBD.

Instructor: Austin J. Brockmeier
Assistant Professor
Department of Electrical and Computer Engineering
Department of Computer and Information Sciences
Resident Faculty, Data Science Institute

Email address: ajbrock@udel.edu
Office location: Evans 306
Office hours: TBD
Website: www.eecis.udel.edu/~ajbrock

Description

Large-scale machine learning is an introduction to the analysis and processing of massive high-dimensional data. Massive data sets generally involve growth not only in the number of individuals represented but also in the number of descriptive parameters of the individuals, leading to exponential growth in the number of hypotheses considered. New approaches to address these problems exploit sparsity prior concepts from optimization theory, signal processing, statistics, and machine learning.

Prerequisites

Previous machine learning course such as FSAN/ELEG 815 or equivalent; Previous experience with a programming language suitable for data science. This course is meant to build on previous experience in machine learning and data science methodology and theory.

Topics

"I learned very early the difference between knowing the name of something and knowing something." R. Feynman

- **Experiment design, evaluation metrics, cross-validation, and model selection**
  - Bag of little bootstraps, stability, online learning
- **Handling high dimensionality**
  - Feature selection and regularization: James-Stein estimator, shrinkage, sparsity, LASSO
  - Representations and algorithms: bags of features, sparse matrices, hashing trick, Johnson-Lindenstrauss lemma, kernel trick, random forests, naive Bayes, k-nearest neighbors
- **Approximating and completing large matrices**
  - Matrix sketching, low-rank decompositions, CUR decomposition, matrix completion, Nyström approximation for kernel matrices, subspace clustering, multiway/tensor generalizations
- **Generalized machine learning paradigms and structured data**
  - Multilabel, multiclass, multiview, multitask, and multi-instance learning
  - Group LASSO, matrix-variate and tensor-variate features and responses
  - Semi-supervised learning, active learning
- **Uncertainty in high dimensional space**
  - Gaussian processes, hyperparameter exploration, Bayesian optimization, reinforcement learning
- **Neural networks**
  - Auto-encoders, generative adversarial networks

Learning Outcomes

- At the completion of this course an engaged student will be able to
    a. mathematically formulate data science and machine learning tasks (problem framing), with clear objective or cost function, assumptions, constraints, and the mathematical characteristics of input and output.
    b. analyze the convexity or concavity of objective or cost functions and constraint sets.
    c. choose an appropriate regularization, model selection criterion, and valid experimental design (including hyperparameter selection) to ensure generalizability and reproducibility, especially to scale to high-dimensional problem size of interest.
    d. summarize and critique descriptions of machine learning models, experimental design, result discussions (including statistical tests) using sound reasoning in a peer review setting with constructive feedback.
    e. select and justify appropriate algorithms, data structures, and relaxations for large-scale problems such that computation can be successfully executed with an understanding of the trade-off between approximation and complexity.
    f. list the challenges, errors, and uncertainties inherent with large-scale data.
    g. formulate and implement appropriate for large-data including sparse matrices and graphs, low-rank matrices, block models, tensor formulations, hierarchical models, tree structures, and neural networks with weight sharing, convolutional, and recurrent architectures.
    h. explain and statistically justify randomization and sampling techniques useful to large-scale data.
    i. compare and contrast different optimization techniques and specific algorithms, such as constraint relaxation, greedy algorithms, and distributed optimization.

Course Elements and Assessment

"For the things we have to learn before we can do them, we learn by doing them." Aristotle

- Weekly assigned readings of journal articles and book chapters with summaries (10%).
- Homework assignment to prototype algorithms for large scale machine learning (20%).
- Midterm examination (25%).
- Projects to explore scaling of different algorithms on large data sets (35%).
- Peer feedback of project reports (10%).

The project will be broken into a series of assessment due in sequence (35%)

- Abstract and predictions on outcomes (10%)
- Formulation, experimental design, and presentation of results (10%)
- Final description and discussion presented on standardized slides (15%)
- *Formulate, analyze, design/choose, analyze, implement, experiment, analyze, discuss*

Paper summaries are for 8 assigned readings (10%)

- Paper summary should provide a short description 3–6 sentences in your own words, a description of your own insights from reading the paper, a list of any resources (algorithms, data sets, experimental designs, proof techniques, statistical tests, theorems, visualizations) you would find useful for your own or future research, and a perspective on how it can be applied to your own research or project
- *Comprehend, summarize, critique, extract insight, catalog resources*

Each student will provide peer feedback on three other students projects in the style of conference peer review (10%)

- Peer feedback on project abstract (2.5%)
- Peer feedback on formulation, experimental design, and presentation of results (2.5%)
- Peer feedback on final report (5%)

Important Dates

- 10/14 Midterm examination
- 10/26 Project abstract due
- 11/2 Revised project abstract due
- 11/9 Peer feedback on project abstract due
- 11/16 Project formulation, experimental design, and design of presentation of results due
- 11/30 Peer feedback on above due
- 12/7 Project final description due
- 12/14 Peer feedback on final project due

Learning Resources

Canvas:

- Weekly readings, homework, project details will be posted there

Technology

Projects will require access to computer programming environment. Suggested languages: R, python, MATLAB/ Octave, Julia

Course Policy Document

Communication

Content-related questions should be discussed in class or during office hours. Please limit the use of email to times when these options are not suitable, as questions sent by email may not be answered before the next class, especially questions whose response could benefit the entire class.

Academic Integrity

Please familiarize yourself with UD policies regarding academic dishonesty. To falsify the results of one's research, to steal the words or ideas of another, to cheat on an assignment, to re-submit the same assignment for different classes, or to allow or assist another to commit these acts corrupts the educational process. Students are expected to do their own work and neither give nor receive unauthorized assistance. Complete details of the university's academic integrity policies and procedures can be found at http://www1.udel.edu/studentconduct/policyref.html Office of Student Conduct, 218 Hullihen Hall, (302) 831-2117. E-mail: student-conduct@udel.edu

Harassment and Discrimination

The University of Delaware works to promote an academic and work environment that is free from all forms of discrimination, including harassment. As a member of the community, your rights, resource and responsibilities are reflected in the non-discrimination and sexual misconduct policies. Please familiarize yourself with these policies at www.udel.edu/oei . You can report any concerns to the University's Office of Equity & Inclusion, at 305 Hullihen Hall, (302) 831-8063 or you can report anonymously through UD Police (302) 831-2222 or the EthicsPoint Compliance Hotline at www1.udel.edu/compliance. You can also report any violation of UD policy on harassment, discrimination, or abuse of any person at this site: sites.udel.edu/sexualmisconduct/how-to-report/

Faculty Statement on Disclosures of Instances of Sexual Misconduct

If, at any time during this course, I happen to be made aware that a student may have been the victim of sexual misconduct (including sexual harassment, sexual violence, domestic/dating violence, or stalking), I am obligated to inform the university's Title IX Coordinator. The university needs to know information about such incidents in order to offer resources to victims and to ensure a safe campus environment for everyone. The Title IX Coordinator will decide if the incident should be examined further. If such a situation is disclosed to me in class, in a paper assignment, or in office hours, I promise to protect your privacy--I will not disclose the incident to anyone but the Title IX Coordinator. For more information on Sexual Misconduct policies, where to get help, and how to reporting information, please refer to www.udel.edu/sexualmisconduct. At UD, we provide 24-hour crisis assistance and victim advocacy and counseling. Contact 302-831-1001, UD Helpline 24/7/365, to get in touch with a sexual offense support advocate.

For information on various places you can turn for help, more information on Sexual Misconduct policies, where to get help, and reporting information please refer to www.udel.edu/sexualmisconduct

Inclusion of Diverse Learning Needs

Any student who thinks he/she may need an accommodation based on a disability should contact the Office of Disability Support Services (DSS) office as soon as possible. The DSS office is located at 240

Academy Street, Alison Hall Suite 130, Phone: 302-831-4643, fax: 302-831-3261, DSS website (www.udel.edu/DSS/). You may contact DSS at dssoffice@udel.edu

Non-Discrimination

The University of Delaware does not discriminate against any person on the basis of race, color, national origin, sex, gender identity or expression, sexual orientation, genetic information, marital status, disability, religion, age, veteran status or any other characteristic protected by applicable law in its employment, educational programs and activities, admissions policies, and scholarship and loan programs as required by Title IX of the Educational Amendments of 1972, the Americans with Disabilities Act of 1990, Section 504 of the Rehabilitation Act of 1973, Title VII of the Civil Rights Act of 1964, and other applicable statutes and University policies. The University of Delaware also prohibits unlawful harassment including sexual harassment and sexual violence.

For inquiries or complaints related to non-discrimination policies, please contact:
Interim Director, Institutional Equity & Title IX Coordinator - Fatimah Stone titleixcoordinator@udel.edu, 305 Hullihen Hall Newark, DE 19716 (302) 831-8063

For complaints related to Section 504 of the Rehabilitation Act of 1973 and/or the Americans with Disabilities Act, please contact: Director, Office of Disability Support Services, Anne L. Jannarone, M.Ed., Ed.S. - ajannaro@udel.edu
Alison Hall, Suite 130, Newark, DE 19716 (302) 831-4643 OR contact the U.S. Department of Education - Office for Civil Rights (wdcrobcolp01.ed.gov/CFAPPS/OCR/contactus.cfm)