EE 529: Data Analytics in ECE, Spring 2021

Overview

An introduction to data analytics covering topics relevant to electrical and computer engineers.

Class information

Instructor:	Aleksandar Dogandžić ald@iastate.edu	Coover 3119 Office hours: M W 1–2
Lectures:	M W 4:25–5:45	1227 Hoover

Textbook

We will mostly follow our course lecture notes and slides. A nice supplemental textbook (although somewhat mathematically oriented) is the free online textbook [Blum et al. 2020]. A good undergraduate-level book with relevant content on clustering and least-squares regression and classification is [Boyd and Vandenberghe 2018]. Also useful and free: [Bishop 2006; Hastie, Tibshirani, and Friedman 2009; Hastie, Tibshirani, and Wainwright 2015; Leskovec et al. 2020; Theodoridis 2020; Zaki and Meira 2020] . Programming aspects (ISU free access, code on Github) [Müller and Guido 2016; Raschka and Mirjalili 2019].

Prerequisites

Familiarity with linear algebra and coding/programming. EE 322 or equivalent course in probability and random processes. Familiarity with optimization will be a plus.

Syllabus (tentative)

- Math basics, inner product, vectors, matrices,
- Modeling data in high dimensions,
- Regression: linear, logistic,
- Classification: nearest neighbors, SVMs, kernel methods, neural nets,
- Visualization and Dimensionality Reduction: PCA, clustering, Isomap, MDS,
- Graphs: Random walks, PageRank,
- Big Data: Compressed sensing and matrix completion,
- Introduction to streaming algorithms.

Homework

Homework-problem sets will be handed out during the semester. Problem sets will involve a mix of theory and practical implementation. Theoretical problems may involve a short mathematical derivation, an analysis of a particular data-processing technique, or a construction of a new method. Implementation problems will involve some degree of programming. Feel free to use a scientific programming environment that you are most comfortable with. Python/Matlab/R/Julia will suffice for most problems.

Collaboration policy

You are encouraged to collaborate on homework assignments. However, you must (a) clearly acknowledge your collaborator, and (b) compose your final writeup and/or code by yourself. If two assignments are obviously identical to each other, then both will automatically receive a score of zero (0). Please talk to me in advance for any clarifications.

Course project

Final exam consists of a course project. The goal of the project is open-ended; it can involve either: (a) conducting research on a specific topic of your choice, or (b) coding up a technique and testing it on a real-world dataset, or (c) both. The only requirement is that it should involve (any combination of) analysis, design, or implementation of a data analytics technique.

Projects are conducted in groups of at most two (2). It will be particularly beneficial for you (and the rest of the class) if you can integrate the project with your own research interests. Start thinking of project ideas early and discuss them with me before finalizing. A two-page project proposal will be due on March 25. Project presentations will be carried out during Dead Week. A final report will be due on May 6. More details will be given out in the coming weeks.

Grading policy

The final grade will be calculated on a score of 100 (homework: 60%, project: 40%). No late submissions please!

Student Accommodations

Meet with me if you have a documented disability and anticipate needing accommodations in this course, Please request that a Disability Resources (DR) staff send a Student Academic Accommodation Request (SAAR) form verifying your disability and specifying the accommodations you need. DR is located in Room 1076 of Student Services.

References

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Blum, Avrim, John Hopcroft, and Ravi Kannan (2020). *Foundations of Data Science*. New York: Cambridge Univ. Press.

Boyd, Stephen and Lieven Vandenberghe (2018). *Applied Linear Algebra: Vectors, Matrices, and Least Squares*. New York: Cambridge Univ. Press. URL: `vmls-book.stanford.edu/`.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer.

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Boca Raton, FL: CRC Press.

Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman (2020). *Mining of Massive Data Sets*. 3rd ed. Cambridge Univ. Press. URL: `www.mmds.org/`.

Müller, Andreas C. and Sarah Guido (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media. URL: `github.com/amueller/introduction_to_ml_with_python`.

Raschka, Sebastian and Vahid Mirjalili (2019). *Python Machine Learning*. 3rd ed. Birmingham, UK: Packt Publishing. URL: `github.com/rasbt/python-machine-learning-book-3rd-edition`.

Theodoridis, Sergios (2020). *Machine Learning: A Bayesian and Optimization Perspective*. 2nd ed. New York: Elsevier.

Zaki, Mohammed J. and Wagner Meira Jr. (2020). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. 2nd ed. New York: Cambridge Univ. Press.