EE 529 Course Information

Overview

Introduction to a variety of foundational principles in data analytics – particularly those relevant to electrical and computer engineers. Topics include theory and algorithms for data classification, visualization, and parameter estimation, with applications to signals, images, matrices, and graphs.

Class information

Instructor:	Aleksandar Dogandžić ald@iastate.edu	Coover 3119 Office hours: Tue Thu 1–2
Lectures:	MW 4:10pm–5:30pm	Coover 1012

Textbook

We will mostly follow our course lecture notes and slides. A nice supplemental textbook (although somewhat mathematically oriented) is the free online textbook, "Foundations of Data Science" by Blum, Hopcroft, and Kannan [Blum et al. 2020]. A good undergraduate-level book with relevant content on clustering and least-squares regression and classification is "Introduction to Applied Linear Algebra" by Boyd and Vandenberghe, also available online [Boyd and Vandenberghe 2018]. Also useful and free: [Deisenroth et al. 2020; Hastie et al. 2009; Leskovec et al. 2020]. Programming aspects (ISU free access, code on Github) [Raschka and Mirjalili 2019]: Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. Also Müller & Guido: Introduction to Machine Learning with Python.

Prerequisites

EE 322 or equivalent course in probability and random processes. Familiarity with linear algebra and optimization.

Syllabus (tentative)

- Modeling data in high dimensions
- Classification: Nearest neighbors, kernel methods, neural methods, …
- Regression: linear, logistic, matrix factorization
- Visualization: Clustering, nonlinear dimensionality reduction
- Graphs: Random walks, spectral analysis
- Big Data: Compressive sensing and matrix completion

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! (Your grade will be reduced by 25% each day it is late.)

Student Accommodations

If you have a documented disability and anticipate needing accommodations in this course, please meet with me. 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

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/`.

Deisenroth, Marc Peter, A Aldo Faisal, and Cheng Soon Ong (2020). *Mathematics for Machine Learning*. New York: Cambridge Univ. Press. URL: `mml-book.github.io/`.

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

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

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`.