

ADVANCED DATA ANALYSIS TECHNIQUES
(Statistical Learning Techniques for Engineering and Science)
CVEN 6833
Spring 2025

Instructor

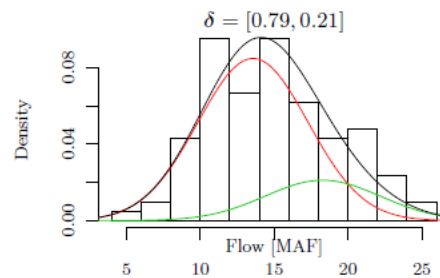
Prof. R. Balaji ECOT 452
Phone: (303) 492-5968
E-mail: balajir@colorado.edu
Lectures: Tuesdays and Thursdays 10:00 – 11:15AM
In-person: SEEC N124
Office hours: (*anytime* on E-mail and by appointment)
Class page: <http://civil.colorado.edu/~balajir/CVEN6833>

Prerequisites

Familiarity and comfort with topics covered in introductory graduate course in probability and statistics (such as [CVEN 5454](#)), calculus, linear algebra

Course Objectives

Lots of data everywhere, but little knowledge!. We face this conundrum in the age of *big data*. The objective of this course is to provide a good exposure to a variety of statistical learning techniques - both traditional and modern – to help *extract knowledge from data*. Examples from hydrology, hydroclimatology, environmental engineering and construction safety will be presented - the techniques are general in nature that they could be easily applied to data analysis problems from *any other fields*. The course will have a significant hands-on component on the powerful data analysis tool **R**¹ (<http://www.r-project.org>).



Course Format

1. Formal lectures with exposure to **R**.
2. There will be ~4 long home works (covering the topics) and a project that will require extensive use of **R**
3. Students have to do a project using data sets from their research and produce a research paper/report. *Many of the student reports have resulted in journal publications over the years.*
4. There are no comprehensive book(s) available that cover all the proposed topics - hence, material from a range of sources (books, research papers etc.) will be used and placed on the class web page.

Planned Topics

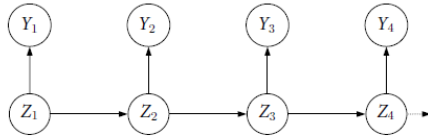
1. Regression (continuous, discrete and binary variables) – **Linear and Nonlinear**
 - Revision of linear regression
 - Generalized Linear Modeling (GLM)
 - Regularization – Ridge and Lasso regressions
 - Nonparametric Regression - Local Polynomials
 - Splines and Generalized Additive Models (GAM)
2. Spatial Regression Models – Kriging
3. Bayesian Hierarchical Modeling

¹ <http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html>

$$\Pr(Z_t|Z_1, \dots, Z_{(t-1)}) = \Pr(Z_t|Z_{t-1})$$

$$\Pr(Y_t|Y_1, \dots, Y_{(t-1)}, Z_1, \dots, Z_{(t-1)}) = \Pr(Y_t|Z_t)$$

$$\Pr(Y_t = y_t|Z_t = i) = p_i(y_t)$$



Supervised Learning

- CART; Random Forest
- PCA-regression; SVM

5. Copulas – Modeling multivariate data and Multivariate Extremes

6. Hidden Markov Models

7. Introduction to Deep Learning

- Convolution Neural Networks (CNN)

8. Time Series Analysis (Modeling/Simulation/Forecasting):

- ARMA (parametric)
- K-nearest neighbor Bootstrap & Block Bootstrap (nonparametric)
- Frequency domain analysis:
 - Wavelet Spectral methods for computing spectrum of time series
 - Singular Spectrum Analysis (SSA)

Grading

Grading will be based entirely on the long homeworks (40%) project & report (50%), project presentation and active class participation (10%).

Suggested References

- An Introduction to Statistical Learning with Applications in R by G. James, D. Witten, T. Hastie and R. Tibshirani – Springer
- The Elements of Statistical Learning by T. Hastie, R. Tibshirani and J. Friedman – Springer
- Applied Spatial Data Analysis with R by Bivand, Roger S., Pebesma, Edzer, Gómez-Rubio, Virgilio – Springer
- Bayesian Data Analysis by A. Gelman, Chapman and Hall, CRC Press, Inc
- Statistical Rethinking – A Bayesian course with examples in R and STAN, Richard McElreath, CRC Press
- Local Regression and Likelihood by C. Loader – Springer
- Wikle, C. K., Zammit-Mangion, A., and Cressie, N. (2019), Spatio-Temporal Statistics with R, Chapman & Hall/CRC
- Graphical Neural Networks – Foundations, Frontiers and Applications, L. Wu, P. Cui, J. Pei and L. Zhao, Springer
- Hidden Markov Models for Time Series by Walter Zucchini and Iain L. MacDonald – Chapman and Hall/CRC
- Statistical Analysis in Climate Research by Hans von Storch and F.W. Zwiers - Cambridge Univ. Press, U.K.
- Statistical Methods in the Atmospheric Sciences: An Introduction by Daniel S. Wilks - Academic Press

4. Multivariate data analyses (Identifying patterns/signals from multivariate data sets/forecasting)

Unsupervised Learning

- Principal Component Analysis
- Singular Value Decomposition (SVD) analysis
- Canonical Correlation Analysis (CCA)
 - Clustering – K-means; Heirarchical; Extremes
 - Self Organizing Maps (SOM)

SYLLABUS STATEMENTS

CLASSROOM BEHAVIOR:

Both students and faculty are responsible for maintaining an appropriate learning environment in all instructional settings, whether in person, remote or online. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. For more information on classroom behavior, Student Conduct & Conflict Resolution policies and others, see this [page](#).

COVID-19 updates

General information

Check for ongoing updates. CU Boulder employs a framework for public health and COVID-19 decision-making that follows Centers for Disease Control and Prevention (CDC) guidance. The campus will continue to work closely with Boulder County Public Health as it considers public health requirements in the Boulder community. If Boulder County Public Health adopts or modifies any requirements, the campus will immediately inform the CU Boulder community. For ongoing updates on campus COVID-19 protocols, check the [COVID-19 webpage](#) and [CU Boulder Today's COVID-19 updates](#).

ACCOMMODATION FOR DISABILITIES

If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the [Disability Services website](#). Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition, see [Temporary Medical Conditions](#) on the Disability Services website.

PREFERRED STUDENT NAMES AND PRONOUNS

CU Boulder recognizes that students' legal information doesn't always align with how they identify. Students may update their preferred names and pronouns via the student portal; those preferred names and pronouns are listed on instructors' class rosters. In the absence of such updates, the name that appears on the class roster is the student's legal name. *Please do not hesitate to correct me if I get a pronoun incorrect (likely someone else's because this is often a 3rd person issue, not a 2nd person issue).*

HONOR CODE

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code academic integrity policy. Violations of the Honor Code may include, but are not limited to: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code (honor@colorado.edu); 303-492-5550). Students found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code as well as academic sanctions from the faculty member. Additional information regarding the Honor Code academic integrity policy can be found on the [Honor Code website](#). *I have reported several students to the Honor Code Office. While it*

is my least favorite thing to do, it is one of the most important responsibilities I have as an instructor. Thus, I do not hesitate to do so when warranted.

SEXUAL MISCONDUCT, DISCRIMINATION, HARASSMENT AND/OR RELATED RETALIATION

The University of Colorado Boulder (CU Boulder) is committed to fostering an inclusive and welcoming learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct (harassment, exploitation, and assault), intimate partner violence (dating or domestic violence), stalking, or protected-class discrimination or harassment by or against members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or email cureport@colorado.edu. Information about OIEC, university policies, [reporting options](#), and the campus resources can be found on the [OIEC website](#).

Please know that faculty and graduate instructors have a responsibility to inform OIEC when made aware of incidents of sexual misconduct, dating and domestic violence, stalking, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about their rights, support resources, and reporting options.

RELIGIOUS HOLIDAYS

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, please contact me at least 1 week ahead of time to make arrangements for any possible assignment submission changes, etc. See the [campus policy regarding religious observances](#) for full details.