ADVANCED DATA ANALYSIS TECHNIQUES
(Statistical Learning Techniques for Engineering and Science)
CVEN 6833
Fall 2019

Instructor
Prof. R. Balaji  ECOT 444
Phone:  (303) 492-5968
E-mail  balajir@colorado.edu
Lectures:  Tuesdays and Thursdays 10 – 11:15PM SEEC C315
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 R1 (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. All the material will be available
   on the class web page.

Planned Topics
1. Regression (continuous, discrete and binary variables) – Linear and Nonlinear
   • Revision of parametric linear regression
   • Generalized Linear Modeling (GLM)
   • Nonparametric Regression - Local Polynomials
   • Splines and Generalized Additive Models (GAM)
   • Bayesian Dynamical Linear Models

2. Spatial Regression Models – Kriging

3. Bayesian Hierarchical Modeling

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; Hierarchical; Extremes
- Self Organizing Maps (SOM)

**Supervised Learning**
- CART; Random Forest
- PCA-regression; SVM

5. Copulas – Modeling multivariate data and Multivariate Extremes

6. Time Series Analysis (Modeling/Simulation/Forecasting):
- ARMA (parametric)
- K-nearest neighbor Bootstrap & Block Bootstrap (nonparametric)

7. Hidden Markov Models
- Nonstationary Extremes

8. Frequency domain analysis:
- Wavelet Spectral methods for computing spectrum of time series
- Time series simulation using spectrum - Wavelet + ARMA based approach
- Singular Spectrum Analysis (SSA)

**Grading**

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

**Suggested References**
Multivariate Statistical Modelling Based on Generalized Linear Models by Ludwig Fahrmeir, Gerhard Tutz – Springer
Local Regression and Likelihood by C. Loader - 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
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
Time Series Analysis by Wei, Addison Wesley Publications
Hidden Markov Models for Time Series by Walter Zucchini and Iain L. MacDonald – Chapman and Hall/CRC
Dynamic Linear Models with R by G. Petris and S. Petrone, Springer.
Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations by Bowman and Azzalini – Oxford Publications