

University of Western Ontario
Winter 2022

SS9878/CS9878: Analysis of High Dimensional Noisy Data

1 Instructor Information

Prerequisite

Having basic statistics knowledge such as likelihood, conditional expectations, and regression would be important to well appreciate this course.

3 Course Outline

Thanks to the advancement of modern technology in acquiring data, data with diverse features and big volume are becoming more accessible than ever. While the abundant volume

4 Sketch of Topics

PART 1: A BRIEF REVIEW AND PREPARATION

- Matrix Operation
- Convex Optimization
- Conditional Expectation
- Likelihood Method
- Estimating Function

PART 2: HIGH DIMENSIONAL DATA ANALYSIS

- Ridge Regression
- Least Absolute Shrinkage and Selection Operator (LASSO)
- Penalize Likelihood

PART 3: MEASUREMENT ERROR MODELS

- Examples Arising from Distinct Contexts
- Overview of Measurement Error Problems
- Methods of Addressing Measurement Error Effects
 - { Regression Calibration
 - { Simulation-Extrapolation (SIMEX)
 - { Estimation Equation

PART 4: MISSING DATA PROBLEMS

- Introduction and Examples
- Missing Data Mechanisms
- Analysis Methods
 - { Imputation methods
 - { Likelihood-based methods
 - { Inverse Probability Weighted GEE
 - { Doubly robust method

5 Evaluation Scheme

15% participation + 35% presentation + 50% course work

Participation

This includes the student's attendance to classes, the involvement and participation with the class discussions, and the efforts paid to the course.

Presentation

Each student is going to give a presentation on a paper (or a few papers) on a topic concerning high dimensional data analysis, measurement error, or missing data. Students can choose papers on their own or ask the instructor to help them identify papers to present. The presentation length is about 10 to 30 minutes, to be determined after the registration number of the course is finalized. Each presentation will be evaluated by both the instructor and audience. Your presentation slides should be sent to the instructor (at least) a day prior to your presentation.

Time: The last few classes

Course Work

- Extension of An Existing Topic:

You may choose a *specific* topic of your interest and read a relevant research paper (or a couple of research papers if you wish). Extend the development of that research paper by using the knowledge you have learned in this course. Your extensions are expected to be well described and comprehensive with technical details. A paper-format report of length 10 - 30 pages is expected.

- Your Own Topic:

You may identify a *new problem of your own interest* and write a report about it. The topic should be pertinent to high dimensional data analysis, measurement error, or missing data. The report should be in the format of a scientific paper with a length about 10 - 30 pages.

- Your Own Problems and Solutions:

You may create a set of *new problems* on high dimensional data analysis, measurement error, or missing data that you think of or modify from existing sources, together with the solutions of those problems. The number of problems can be as many as you want, but is expected no less than seven.

- Software Package:

You may choose a paper on high dimensional data analysis, measurement error, or missing data, and develop a software package for the public to use. The developed package should be reliable and will be posted at a public platform such as CRAN or GitHub. Check with the instructor before you start.

- Real Application:

You may find an available data set and implement a method (or some methods) you have learned from this course to analyze the data. A complete report of the analysis should be prepared in the scientific paper format.

Note on Course Work

- The course work should be prepared in a self-contained manner with each notation clearly defined. It is expected to be laid out in a research manuscript format, including a title, an abstract, and references, along with the main text. The layout of the contents should be logic and flow smoothly.
- The course work should be prepared neatly in Latex. A .pdf file together with a .tex file is expected to be submitted.
- The course work should be completed and submitted on an individual basis. However, if you think discussing with your peers can help you output more valuable outcomes, you may do so. In this case, please clearly write a statement to point out: (1) how your work is benefited from the discussion, (2) who is involved with the discussion, and (3) what part(s) are identical to your peer(s)' work.

- **Due Date: April 10, 2022**

Important Note: Please be noted that no extensions will be granted to accommodating various requests. It is important that you carefully plan your time throughout the term and do not wait until the last minute to start preparing your course work. Learning to gain time management skills is a secondary goal of this course.

6 Reference Books

- Yi, G.Y. (2017). *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application*. Springer Science+Business Media LLC, New York.
- Yi, G.Y., Delaigle, A. and Gustafson, P. (2021). *Handbook of Measurement Error Models*. Chapman & Hall/CRC.
- Carroll, R.J., Ruppert, D., Stefanski, L.A., and Crainiceany, C.M. (2006). *Measurement Error in Nonlinear Models*, 2nd ed., Chapman & Hall.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical*, hF34 11-32o50()JTJ -39-326(h9o.-3303)a10

http://www.uwo.ca/univsec/pdf/academic_policies/appeals/accommodation_medical.pdf

Academic Policies

The website for Registrarial Services is
<http://www.registrar.uwo.ca>

In accordance with policy,

<http://www.uwo.ca/its/identity/activatenonstudent.html>,

the centrally administered e-mail account provided to students will be considered the individual's official university e-mail address. It is the responsibility of the account holder to ensure that e-mail received from the University at his/her official university address is attended to in a timely manner.

Scholastic offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at this website:

http://www.uwo.ca/univsec/pdf/academic_policies/appeals/scholastic_discipline_undergrad.pdf

Support Services

Please contact the course instructor if you require lecture or printed material in an alternate format or if any other arrangements can make this course more accessible to you. You may also wish to contact Services for Students with Disabilities (SSD) at 661-2111 ext. 82147 if you have questions regarding accommodation.

The policy on Accommodation for Students with Disabilities can be found at
www.uwo.ca/univsec/pdf/academic_policies/appeals/accommodation_disabilities.pdf

The policy on Accommodation for Religious Holidays can be found at
http://www.uwo.ca/univsec/pdf/academic_policies/appeals/accommodation_religious.pdf