

University of Western Ontario
Winter 2021

SS9878/CS9878: Analysis of High Dimensional Noisy Data

1 Instructor Information

Grace Y. Yi, Ph.D.

Professor, Canada Research Chair in Data Science (Tier 1)

Department Statistical and Actuarial Sciences

Department of Computer Science

University of Western Ontario

<http://sher.stats.uwo.ca/faculty/yyi/>

<https://covid-19-canada.uwo.ca>

Email: gyi5@uwo.ca

Office Hour: Wednesdays 8:30-9:30pm

Note: The instructor has to handle a large volume of emails on a daily basis. When it

Course Audience

This is a graduate topic course cross-listed by the Department of Statistical and Actuarial Sciences (DSAS) and the Department of Computer Science (DCS). It is open to graduate students in DSAS, DCS, and Data Analytic Master program. This course focuses the discussion on the theory and methods. Hands-on experience on implementations of various methods is not the target, though some implementation software packages are to be discussed.

Prerequisite

Having basic statistics knowledge such as likelihood, conditional expectations, and re-

4 Sketch of Topics

PART 1: A BRIEF REVIEW AND PREPARATION

- Convex Optimization
- Matrix Operation
- 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

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

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 on your own 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 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 logical 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.
-

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.
- 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 Analysis with Missing Data*, 2nd Ed., New York: Wiley.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical learning, with Application in R*. Springer Science+Business Media LLC, New York.

7 Disclaimer

- The lecture materials are only intended for your own use. Some of them might be the on-going research of the instructor and her co-authors that has not been published yet. Please do NOT distribute the lecture notes without the instructor's permission.
- Citation: In case you need to cite some of the lecture material for your future work, you may refer it as

\Yi, G.Y. (2021). *Lecture Notes of \SS9878/CS9878 - Analysis of High Dimensional Noisy Data*", University of Western Ontario.

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