The University of California - Los Angeles
Department of Statistics

Matrix Algebra and Optimization
Statistics 202B
Professor: Mark S. Handcock

Syllabus and Description

Professor:	Mark S. Handcock	8105C Mathematical Sciences Bldg
Office Hours:	Friday: 3:30pm - 4:15pm	8141 Math. Sci. Bldg
Teaching Assistant:	TBA, Esq.,	
	Weekly office hours:	
	Other times by arrangement. Clearly composed questions sent to the "TBA@ucla.edu" will receive written replies	

Motivation and Synopsis

During the twentieth century, the development of statistical computing played a crucial facilitating role for the growth of the statistics discipline and the adoption of statistical methods within the scientific community and beyond. In the twenty-first century digital age, the amounts of data available for statistical analysis has grown tremendously, yielding new opportunities for statistical computing, as well as new challenges. Statistical computing constitutes an important part of a statistics education, and is highly valuable for statisticians in both academia and industry.

This course is an introduction computational methods that are useful for statistical analysis, with implementations in the statistical package `R`.

Computational methods are essential to both understand and implement modern statistical ideas. Often the computation methods are a translation of the statistical methods into another language, one that computer understand. In this class we will consider this process for key statistical models and techniques. These include multivariate regression, principal component analysis, and multivariate analysis. In doing so, we study the core mathematical

and numerical ideas that make this possible including matrix analysis and deterministic optimization methods.

The course has various parts:

- We'll do some *matrix algebra* (not linear algebra), emphasizing what is available in R.
- We apply the matrix algebra to *multivariate data analysis*, in particular to regression, principal component analysis, canonical analysis, correspondence analysis.
- We discuss *optimization methods*. Most of these will be deterministic, but we will also consider some stochastic variants.

The primary purpose of this course is to provide students with a common set of core knowledge about statistical computing computing for their class work and research. The course will have an applied focus on tools. The course will involve the practical application of the ideas of statistical computing and their implementation through statistical software, particularly `R`.

Structure of the Course

There will be two lectures per week.

Textbooks

[**G**] Gentle, James E
Matrix Algebra: Theory, Computations, and Applications in Statistics (2007).
Springer
Required and available online as a ebook from the library, free for UCLA students.
`http://link.springer.com/book/10.1007%2F978-0-387-70873-7`

[**EH**] Everitt, Brian and Hothorn, Torsten
Introduction to Applied Multivariate Analysis with R (2011).
Springer
Required and available online as a ebook from the library, free for UCLA students.
`http://link.springer.com/book/10.1007%2F978-1-4419-9650-3`

[**I**] Izenman, Alan
Modern Multivariate Statistical Techniques (2008).
Springer
Required and available online as a ebook from the library, free for UCLA students.
`http://link.springer.com/book/10.1007/978-0-387-78189-1`

[**LO**] Lange, Kenneth
Optimization (2013).
Springer
Required and available online as a ebook from the library, free for UCLA students.
`http://link.springer.com/book/10.1007%2F978-1-4614-5838-8`

Students must be connected to the UCLA network to obtain their free download. Students who would like to download the textbook off-campus may do so by connecting the the UCLA network via VPN `https://www.bol.ucla.edu/services/vpn/all.html`.

In addition to these books, there are a multitude of books covering pieces of the course content and from varying perspectives. I suggest you use one or more of the following.

Broad References on Computational Statistics

For computational statistics I suggest

[**GH**] Geof H. Givens and Jennifer A. Hoeting
Computational Statistics, 2nd Ed. (2013).
Wiley series in computational statistics.
Available online as a ebook from the library, free for UCLA students.
`http://onlinelibrary.wiley.com/book/10.1002/9781118555552`

or the massive:

[**JGCS**] Gentle, James E.
Computational Statistics (2009).
Springer: New York.
Required and available online as a ebook from the library, free for UCLA students.
`http://link.springer.com/book/10.1007/978-0-387-75936-4/page/1`

Background on `R`

[**W**] Wickham, Hadley.
Advanced R. (2014).
Chapman and Hall/CRC.
Required and available online as a ebook from the library, free for UCLA students.
`http://www.crcnetbase.com/doi/book/10.1201/b17487`

[**C**] Chang, Winston.
R Graphics Cookbook. (2012).
O'Reilly.
Required and available online as a ebook from the library, free for UCLA students.
`https://goo.gl/IJM16i`

[**M**] Maillardet, Robert, Owen Jones, and Andrew Robinson.
Introduction to Scientific Programming and Simulation Using R. (2014).
Chapman and Hall/CRC.
Required and available online as a ebook from the library, free for UCLA students.
`http://www.crcnetbase.com/doi/book/10.1201/9781420068740`

[**JC**] Chambers, John M.
Software for Data Analysis: Programming with R. (2008).
Springer: New York.
Required and available online as a ebook from the library, free for UCLA students.
`http://link.springer.com/book/10.1007/978-0-387-75936-4/page/1`

Other Resources

You can read the other books with different perspectives online for free from any UCLA account, starting from:

`proquest.safaribooksonline.com/search?q=BOOKTITLE%20r`

Syllabus of the Course

The syllabus of the course will develop on the following weekly schedule. The some later topics may not be reached and we will make choice among them toward the end of the quarter.

Week	Contents discussed
1	Introduction: Why are computation methods important?
2	Fundamentals of Matrices
3	Matrix Decompositions
4	Matrix Approximation
5	Regression and Reduced Rank Methods
6	Multivariate Methods
7	Principles of optimization
8	General optimization
9	Sequential Simplification: MM and friends
10	Additional topics

Course Webpage and Discussion Forum

The course has a webpage through the UCLA *Bruin Learn* system, . The webpage will be continuously updated throughout the course with handouts, homework assignments and solutions. Users sign in to Bruin Learn with their UCLA Logon IDs.

I will be using `campuswire` to provide discussion of issues in class and related questions. For questions that might be of interest to other students, please use `campuswire` rather than solely emailing me. There other students and the TA can answer questions in addition to me. Example of questions are about interesting articles you have seen in the media, problems with access to resources, homework or computer questions. Enjoy!

Please regularly consult this classes Bruin Learn home page, `campuswire` and the archive of the Announcements mailing list. It will contain lecture notes, homework, solutions and course information.

Computer Usage and Software

The computer is the scientific laboratory of the applied researcher in quantitative fields. As such this course requires the student to develop a degree of comfort and competence "in the lab".

Our computer interface to `R` will be the RStudio IDE, which you can download from `www.rstudio.com`.

Course Requirements and Grades

- 60% Homework (6 assignments, 10% each, none are dropped)
- 40% Final project (written report)

Homework

There will be weekly homeworks and exercises both the theory and real data analysis.

None of the homework scores will be dropped. It is your responsibility to verify that your homework assignment successfully uploaded by the deadline. All homework assignments will be posted. Students will submit the solution as `PDF` files electronically via the Bruin Learn *Homeworks* page.

Files must not refer to any resources on the local machine or to files that are not publicly available online. No one should make manual edits to a data file on his or her local machine.

Students are free to discuss homework problems and solutions. Discussing the contents of the course with fellow students can be a valuable element of the learning process, and doing so is therefore generally encouraged. However, each student must hand in their own solutions, and the student should, if asked, be able to explain the solutions.

There will be a final project worth 40% of the grade. For details of the project, see the "Homework" section.

The project can take one of two forms. You should choose one only.

The **first** is to review a topic in multivariate analysis using matrix algebra or optimization. This topic should extend or expand on a topic covered in the class, but not be just that topic. The report should be stand-alone, starting from the level of the lecture notes and expanding out to review the existing literature on that topic. It need not analyze real data, but may use real or simulated data to illustrate the main points. If you choose this form you must pass the topic by me first by sending me a description in an email, chatting with me after class or coming to my office hours.

The **second** is to undertake an analysis of a multivariate data set that can be expressed in matrix form. The analysis should be using one or more of the methods covered in class. You can select any data-set you find interesting, but preferable related to your graduate work or thesis area. You can select any multivariate dataset you like, but preferable related to your graduate work or thesis area. I do not want a quick and routine analysis; a good job will show understanding of the problem and possible solutions and techniques to consider. The technical results should be stated clearly. The report must contain a clearly written conclusion section giving a non-technical summary that is concise and informative. Do *not* merely use data from a textbook - the world is an interesting place! All data sources must be cited, and described.

Late Policy for Homework (silly but necessary)

There is a 10 minute grace period. Submissions up to 10 minutes late will be accepted with no penalty.

Homework assignments submitted 11 minutes late or more will be accepted with penalty. There is a minimum deduction of 1 point for being less than one hour late. An additional 0.5-point deduction will be taken for each additional hour it is late.

Thus, an assignment that is between 11 and 59 minutes late will receive a 1-point deduction. An assignment that is between 1 hour and 1:59 late will receive a 1.5-point deduction. An assignment that is between 2 hours and 2:59 late will receive a 2-point deduction, and so on.

I strongly advise uploading the homework to the Bruin Learn well in advance of the deadline in case there are connectivity problems or server issues.

Academic Integrity

As a student and member of the University community, you are here to get an education and are, therefore, expected to demonstrate integrity in your academic endeavors. All students must uphold University of California Standards of Student Conduct as administered by the Office of the Dean of Students. Students are subject to disciplinary action for several types of misconduct, including but not limited to: cheating, multiple submissions, plagiarism, prohibited collaboration, facilitating academic dishonesty, or knowingly furnishing false information. You may have assignments or projects in which you work with a partner or with a group. For example, you are welcome, and even encouraged, to work with others to solve homework problems. Even though you are working together, the assignment you submit for a grade must be IN YOUR OWN WORDS, unless you receive specific instructions to the contrary. For more information about academic integrity, please go to www.deanofstudents.ucla.edu.

I welcome comments or suggestions about the course at any time, either in person, by letter, or by email. Please feel free to use these ways make comments to me about any aspect of the course.

Support

Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. Students who have experienced sexual harassment or sexual violence can receive confidential support and advocacy at the CARE Advocacy Office for Sexual and Gender-Based Violence, 1st Floor Wooden Center West, CAREadvocate@caps.ucla.edu, (310) 206-2465. You can also report sexual violence or sexual harassment directly to the University's Title IX Coordinator, Kathleen Salvaty, 2241 Murphy Hall, titleix@conet.ucla.edu, (310) 206-3417.

I welcome comments or suggestions about the course at any time, either in person, by letter, or by email. Please feel free to use these ways make comments to me about any aspect of the course.