Keck School of Medicine of USC

PM579 Statistical Analysis of High-Dimensional Data

Units: 4
Term: Summer 2022
Time: Wed 1-4pm
Location: SSB-114

Instructor: Kimberly Siegmund
Office: SSB-202K
Office Hours: Tues 11am-12 and by appt
Contact Info: kims@usc.edu
323-442-1310 (office)
Will reply to email within 48 hrs.

Course Description

The course is designed for M.S. and Ph.D. students in the biological or mathematical sciences, and is highly recommended for Biostatistics Ph.D. students in the Statistical Genetics track. The course aims to provide students with a broad overview of current statistical problems and approaches to high-dimensional data analysis. The content will cover methods for classification and class discovery, using data sets for gene expression and DNA methylation. This course will be taught with an emphasis on selecting the appropriate statistical method for data analysis and interpreting the results. We will learn and use the R statistical computing language and Bioconductor, open source software for the analysis of genomic data.

Learning Objectives

After completing this course, the student will be expected to:
1. Create figures to visualize high-dimensional data
2. Perform statistical hypothesis testing in a high-dimensional setting while controlling the error rate
3. Apply the proper statistical analysis method to high-dimensional data
4. Interpret the results from a statistical analysis of high-dimensional data
5. Present analysis results to readers who are not familiar with the statistical methods

Prerequisite(s): PM510 or equivalent
Recommended Preparation: PM511a

Teaching Methods & Assessments

Teaching Methods

- Assigned reading/writing (texts)
- Assigned reading (journal articles)
- Computer lab
- Group Activity
- Recorded Lecture

Assessment Methods

- Quiz
- Weekly assignments
- Oral presentation

- Term paper

Communication Policies
Students are encouraged to contact the instructor by email and during office hours. The instructor will reply to emails within 48 hours, 72 hours over a weekend, and the work day following a holiday.

Course Notes
The instructor will use Blackboard, GoogleDrive and GitHub for file sharing.

Technological Proficiency and Hardware/Software Required
Students will be asked to bring a laptop to each class. They will learn R programming and the use of Bioconductor packages freely available from the following websites: http://www.r-project.org/ http://www.bioconductor.org/ Computer code will be shared through GitHub (https://github.com) and GitHub classroom.

Required Readings and Supplementary Materials
The textbook will be Bioconductor Case Studies, Springer Inc., editors Hahn F, Huber W, Gentleman R, Falcon S. 2008, and freely available from USC Library (HHGF)

Grolemund G, Wickham H. R for Data Science, 2016. Available at https://r4ds.had.co.nz/. (GW)

Useful references:

Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S. Eds.
Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Springer Science+Business Media, Inc., 2005.

R scripts for textbook: Bioconductor Case Studies
http://www.bioconductor.org/help/publications/books/bioconductor-case-studies/web-supplement/

Description and Assessment of Assignments
Weekly homework assignments will be given to provide experience in applications of standard methods to real data. During Summer offerings (12 week schedule), no more than 12 homeworks will be assigned. Students will give one oral presentation on a topic of their choice.

In lieu of a final exam, students will be asked to prepare a term paper, due on the last day of class. The term paper could be in the form of
(1) a grant application,
(2) a report on a statistical analysis of high-dimensional data, or
(3) a report on comparing a number of statistical techniques on a (high-dimensional) data set. The data may be your own, or some obtained from the public domain.

Grading Breakdown

Assignment	Points	% of Grade
Class Participation		10
Homework		40
Oral Presentation		25
Written Project		25

Total		100

Grading Scale

Course final grades will be determined using the following scale

A	90-100
A-	85-89
B+	82-84
B	79-81
B-	75-78
C+	71-74
C	67-70
C-	63-66
D+	59-62
D	57-58
F	56 and below

Assignment Submission Policy

Coding assignments will be submitted through GitHub Classroom and reports will be submitted through Blackboard.

Grading Timeline

Homework will be graded within 1 week of handing in.

Late Work

Late assignments will be accepted, but no later than the last day of class. Late work will be penalized by 20% deduction in points for each week the assignment is late unless due to an emergency situation excused by the instructor. Email the instructor as soon as possible to discuss alternate arrangements due to an emergency.

Technology in the classroom

On ground, students will bring a laptop to class each week for computer lab. Online, students will be expected to attend live Zoom sessions through a computer/laptop allowing them to share their screen.

Academic Integrity

A grade of zero will be applied to submitted work that does not comply with the USC standards of academic conduct. Such work may not be resubmitted for a new grade. Academic integrity is included at the end of the syllabus.

Classroom norms

I expect students to treat other with compassion and understanding, and to have a genuine interest in learning.

Expectations on Student Engagement

Students are expected to actively participate in class discussions and work in small groups for in-class exercises.

Course Schedule: A Weekly Breakdown

Date	Topic
Week 1 **5/18**	Introduction to molecular biology and high throughput technologies Learning Objectives: 1. Create a reproducible report 2. Manipulate high-dimensional data in R Reading: Class Material on Blackboard GW. Welcome & Ch 1 Homework 1: Watch Video: http://videolectures.net/cancerbioinformatics2010_baggerly_irrh/ turn in code using GitHub classroom and report through Blackboard (due 5/24)
Week 2 **5/25**	Data Visualization Learning Objectives: 1. Conduct at least 2 different dimension reduction techniques. 2. Create a figure to show the results from a dimension reduction method 3. Write the methods section for a journal article describing the data and your analysis Reading: Class Material on Blackboard Homework 2: DUE 5/31
Week 3 **6/1**	Unsupervised Learning Learning Objectives: 1. Compare and contrast agglomerative versus partitioning unsupervised learning methods 2. Apply at least two different unsupervised learning methods 3. List three decisions/actions of the method that influenced your analysis Reading: Class Material on Blackboard HHGF Ch. 10 Homework 3: DUE 6/14
Week 4 **6/8**	Differential expression Fold-change, volcano plots, moderated t tests, annotation, GSEA Learning Objectives: 1. Explain why moderated t-tests are applied to gene expression data 2. Apply moderated t-tests to gene expression data Reading: HHGF Ch. 3.4.1-3.4.4, 7.1-7.3, 7.5 Class handouts on Blackboard Activities: Apply methods using R
Week 5 **6/15**	Multiple Testing Learning Objective: 1. Interpret the family-wise error rate 2. Interpret the false-discovery rate

	Reading: Class material on Blackboard Reading: Benjamini & Hochberg (1995); Storey & Tibshirani (2003)
Week 6 **6/22**	Multiple testing II Learning Objectives: 1. Interpret multiple comparison adjusted p-values, q-values and the posterior error probability. 2. Describe a simple method to increase power for multiple hypothesis testing 3. Name the statistical quantities required for power calculations Reading: Class Material on Blackboard References: Bourgon et al. (2010, PNAS); Jung (2005) Homework: Select Topic for Student Presentations (6/29)
Week 7 **6/29**	RNA sequencing Learning Objectives: 1. Select the proper statistical model for differential gene expression (DGE) of RNA-seq data 2. Apply a computational pipeline for DGE analysis Reading: Class Material on Blackboard F1000Research 2016, 5:1408 Assignment: Execute code from F1000Research (due 6/29) Homework: Select Topic for class project (due 7/6)
Week 8 **7/6**	Supervised Learning Learning Objectives: 1. Describe the bias-variance tradeoff for classification methods 2. Name two methods for evaluating a classification model Reading: HHGF Ch. 9 Class Material on Blackboard
Week 9 **7/13**	Student Presentations Learning Objective: Present a topic on high-dimensional data to a general audience of scientists Handout on Tips on giving talks Watch: TED talk videos on giving talks
Week 10 **7/20**	Network Analysis Learning Objective: 1. Describe hub genes 2. Explain why hub genes are interesting biologically 3. Describe one approach to identify hub genes Reading: Class Material on Blackboard References: Zhang and Horvath (2005) https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/
Week 11	Data Integration

7/27	Learning Objectives: 1. List three methods for integrating data from separate platforms Reading: Class Material on Blackboard References: Ritchie et al. (2015) Nat Rev Genet; Witten et al. (2009) SAGMB; Shen et al (2013) Ann of Applied Statistics
Week 12 8/3	Data integration with TCGA Learning Objective: Conduct integrative analysis of TCGA colon cancer data In class activity: data analysis Final Project DUE 8/3

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on Research and Scholarship Misconduct.

Students and Disability Accommodations:

USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
studenthealth.usc.edu/counseling
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org
Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call

studenthealth.usc.edu/sexual-assault
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086
eeotix.usc.edu
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
usc-advocate.symplicity.com/care_report
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - (213) 740-0776
osas.usc.edu
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.
USC Campus Support and Intervention - (213) 821-4710
campussupport.usc.edu
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - (213) 740-2101
diversity.usc.edu
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
dps.usc.edu, emergency.usc.edu
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call
dps.usc.edu
Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)
ombuds.usc.edu
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

*Occupational Therapy Faculty Practice - (323) 442-33*40 or otfp@med.usc.edu
chan.usc.edu/otfp
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.