UCLA FIELDING SCHOOL OF PUBLIC HEALTH

BIOSTAT 274: Topics in Statistical Machine Learning
Term: Spring 2022
Credits: 4

COURSE SYLLABUS

A. Overview

Course Description	This course aims to introduce students to advanced topics in machine learning with focus on statistics and algorithms. The course will dive into the math and statistics behind various machine learning methods, from the classic Linear Discriminant Analysis to the cutting-edge deep neural networks. Of equal importance, the other part of the course is to introduce the algorithms and implementations in R and Python of the methods.
Prerequisites	Biostat 200 AB, or consent of instructor
Instructor	Zhe Fei Assistant Professor In-Residence Department of Biostatistics UCLA Fielding School of Public Health (FSPH) Office: 21-293D Phone: 310-825-9786 Email: feiz@g.ucla.edu
Class Days, Times, Location	M/W 1-2:50 p.m. CHS 71-257
Office Hours	M 12 – 1 p.m. CHS 21-293D
Course Texts	There is no required textbook. Lecture slides and reading materials will be made available online. The following books are recommended references: • **[ESL]** Hastie, Tibshirani, Friedman. The Element of Statistical Learning. • **[ISL]** James, Witten, Hastie, Tibshirani. An Introduction to Statistical Learning.
Course Format	Lectures: 4 hours per week

Classroom Participation & Attendance	Students are required to attend all lectures.
UCLA ADA Policy	Students needing academic accommodations based on a disability should contact the Center for Accessible Education (CAE) at (310) 825-1501 or in person at Murphy Hall A255. When possible, students should contact the CAE within the first two weeks of the term as reasonable notice is needed to coordinate accommodations. For more information visit www.cae.ucla.edu.
ADA Contact	Nickey Woods Center for Accessible Education A255 Murphy Hall Phone: (310) 825-1501 TTY / TTD: (310) 206-6083 Fax: (310) 825-9656
Inclusivity	UCLA's Office for Equity, Diversity, and Inclusion provides resources, events, and information about current initiatives at UCLA to support equality for all members of the UCLA community. I hope that you will communicate with me or your TA if you experience anything in this course that does not support an inclusive environment, and you can also report any incidents you may witness or experience on campus to the Office of Equity, Diversity, and Inclusion on their website (https://equity.ucla.edu/).

B. Course Assignments & Exams

Grading:

Active participation 20% (includes attending class and arriving on time, asking questions, engaging with other students and the instructor)

Four homework assignments, 20% each. Each assignment consists of the computational part and the statistical part.

Grading Scale:

Grade Point:	4.0	4.0	3.67	3.33	3.0	2.67	2.33	2.0	1.67	1.33	1.0	0.67	0
Final Percentage:	100-98	97-93	92-90	89-88	87-83	82-80	79-78	77-73	72-70	69-68	67-63	62-60	<60
Letter Grade:	A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F

Course Exams Schedule

No exams for this class

C. Course Policies & UCLA Policies

Message about Academic Integrity to all UCLA Students from UCLA Dean of Students: UCLA is a community of scholars. In this community, all members including faculty, staff and students alike are responsible for maintaining standards of academic honesty. 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. You are evaluated on your own merits. Cheating, plagiarism, collaborative work, multiple submissions without the permission of the professor, or other kinds of academic dishonesty are considered unacceptable behavior and will result in formal disciplinary proceedings usually resulting in **suspension** or **dismissal**.

Forms of Academic Dishonesty: As specified in the UCLA Student Conduct Code, violations or attempted violations of academic dishonesty include, but are not limited to, cheating, fabrication, plagiarism, multiple submissions or facilitating academic dishonesty:

Cheating: Unauthorized acquiring of knowledge of an examination or part of an examination

- Allowing another person to take a quiz, exam, or similar evaluation for you
- Using unauthorized material, information, or study aids in any academic exercise or examination – textbook, notes, formula list, calculator, etc.
- Unauthorized collaboration in providing or requesting assistance, such as sharing information
- Unauthorized use of someone else's data in completing a computer exercise
- Altering a graded exam or assignment and requesting that it be regraded

Plagiarism: Presenting another's words or ideas as if they were one's own

- Submitting as your own through purchase or otherwise, part of or an entire work produced verbatim by someone else
- Paraphrasing ideas, data or writing without properly acknowledging the source
- Unauthorized transfer and use of someone else's computer file as your own
- Unauthorized use of someone else's data in completing a computer exercise

Multiple Submissions: Submitting the same work (with exact or similar content) in more than one class without permission from the instructor to do so. This includes courses you are currently taking, as well as courses you might take in another quarter

Facilitating Academic Dishonesty: Participating in any action that compromises the integrity if the academic standards of the University; assisting another to commit an act of academic dishonesty

- Taking a quiz, exam, or similar evaluation in place of another person
- Allowing another student to copy from you

- Providing material or other information to another student with knowledge that such assistance could be used in any of the violations stated above (e.g., giving test information to students in other discussion sections of the same course)

Fabrication: Falsification or invention of any information in an academic exercise

- Altering data to support research
- Presenting results from research that was not performed
- Crediting source material that was not used for research

While you are here at UCLA, if you are unsure whether what you are considering doing is cheating, **don't take chances**, ask your professor. In addition, avoid placing yourself in situations which might lead your professor to **suspect you of cheating**.

Alternatives to Academic Dishonesty

- **Seek out help** – Meet with your professor, ask for assistance as needed.
- **Ask for an extension** – if you explain your situation to your professor, she/he might be able to grant you an extended deadline for an upcoming assignment.
- **See a counselor** at Student Psychological Services, and/or your school, college or department – UCLA has many resources for students who are feeling the stresses of academic and personal pressures.

If you would like more information, please come see us at the Dean of Students' Office in 1206 Murphy Hall, call us at (310) 825-3871 or visit their website at www.deanofstudents.ucla.edu.

D. Course Outline

This schedule may subject to change as the quarter progresses, according to student enrollment and needs.

Learning objectives:

1. Understand and perform shrinkage methods and algorithms for variable/model selection.
2. Learn and understand the fundamentals of convex optimization and re-sampling methods.
3. Understand the challenges and goals of statistical inference for high dimensional models; learn about recent development and learn to use the methods and their implementations in R software.
4. Learn and use linear methods for classification, Linear Discriminant Analysis, Trees and Random Forests. Know how to implement the methods in R or python.
5. Learn and use splines and kernel smoothing methods, and how kernels are used in Support Vector Machines. Know how to implement SVMs in R or python.
6. Learn and use unsupervised learning methods like clustering and PCA, and how they achieve dimension reduction.
7. Understand the basics and statistical derivation of neural networks; how they are connected to but also different from regression models. Know how to code the algorithms by yourselves.
8. Further learn about more complex networks, including convolutional neural networks, recurrent neural networks, autoencoders, etc.

Week	Lecture	Materials (in addition to lecture notes)	Assignment
1	Introduction Course learning objectives and expectations Linear regression	ESL Chapters 1, 2 ESL Chapter 3	None
2	Shrinkage methods and algorithms Convex optimization	ESL Chapter 3 Boyd, Stephen, and Lieven Vandenberghe. Convex Optimization. Cambridge University Press, 2004. free online Lange. Optimization. Springer, New York, NY. 2004	Assignment No. 1
3	Resampling methods High dimensional inference	ESL Chapters 7, 8 Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76(1), 217–242. Van de Geer, S., P. B¨uhlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. The Annals of	None Reading

		Statistics 42(3), 1166–1202.	
4	High dimensional inference: debiased estimators; split and smoothing; algorithms and implementations	Javanmard, A. and A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. Journal of Machine Learning Research 15(1), 2869–2909. Dezeure, R., P. Buhrmann, L. Meier, and N. Meinshausen (2015). High-dimensional inference: confidence intervals, p-values and r-software hdi. Statistical Science 30(4), 533–558. Zhe Fei, Ji Zhu, Moulinath Banerjee, and Yi Li. Drawing inferences for high-dimensional linear models: A selection-assisted partial regression and smoothing approach. Biometrics, 75(2):551-561, 2019.	Reading
5	Linear methods for classification Trees and Random Forests	ESL Chapter 4 ESL Chapter 15	Assignment No. 2
6	Splines and kernel smoothing methods Support vector machines	ESL Chapters 5, 6 ESL Chapter 12	None
7	Unsupervised Learning: clustering Dimension reduction	ESL Chapter 13 ESL Chapter 14	Assignment No. 3
8	Ensemble learning Model based clustering: Gaussian mixture models	ESL Chapter 16 https://simons.berkeley.edu/talks/tamara-broderick-michael-jordan-01-25-2017-2	None
9	Neural networks	ESL Chapter 11 Mei, S., Montanari, A., & Nguyen, P. M. (2018). A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, *115*(33), E7665-E7671.	Assignment No. 4 Reading
10	Deep learning	Wang, H., & Raj, B. (2017). On the origin of deep learning. *arXiv preprint arXiv:1702.07800*. Sun, R. (2019). Optimization for deep learning: theory and algorithms. *arXiv preprint arXiv:1912.08957*.	Reading

E. Competencies

MPH competencies

1.Demonstrate mastery of fundamental concepts of statistical analysis for datasets from health studies.
2.Employ computational methods for analysis of public data sets.
3.Recommend research study designs to support public-health-relevant data analyses.
4.Contribute to the analysis of public health studies in collaborative multidisciplinary teams
5.Prepare written and oral presentations providing public health insights based on statistical analyses.

MS competencies

1.Demonstrate mastery of the foundations of probability theory and biostatistical concepts. 2.Examine foundations of linear and generalized linear statistical models. 3.Employ computational methods of applied regression to analysis of biomedical data sets .4.Provide effective biostatistical advice in collaborative research projects 5.Communicate results of biostatistical research both orally and in writing

PhD competencies

1.Demonstrate mastery of advanced theory and applications of statistical models.2.Develop algorithms to implement advanced biostatistical methodologies. 3.Present effective seminars on biostatistical research and research in public health sciences.4.Promote effective use of biostatistics in collaborative team research on public health problems 5.Develop original research in the theory/methodology of biostatistics and demonstrate its application in a substantive field.