

University of Florida
College of Public Health & Health Professions Syllabus
PHC 6097 Statistical Learning with Applications in Health Sciences (3 credits)
Semester: Spring 2023
Delivery Format: On-Campus

Instructor Name: Muxuan Liang
Room 5233 CTB
E-mail: muxuan.liang@ufl.edu
Telephone: (352) 294-5935

Class Meets: Mondays (12:50pm-2:45pm) and Wednesdays (1:55pm-2:45pm)
HPNP G-112
There is no class on Martin Luther King Holiday and Midterm-exam week

Office hours: Mondays (2:45pm-5pm)

Preferred Course Communications: e-mail (muxuan.liang@ufl.edu)

Prerequisites: PHC 6050c Biostatistical Methods I, PHC 6051 Biostatistical Methods II, or the permission of the instructor.

Purpose and Outcome

Course Overview This is an advanced course in statistical learning, which covers a broad range of methods and their applications in high-dimensional data analysis. Many of these methods go far beyond classical statistical methods and are developed for addressing modern problems we encounter in public health and medical science settings.

Relation to Program Outcomes The methods and tools learned from this course will enhance students' abilities in data analysis, method development, and professional advancement. All applications of methods in this course will be implemented using R statistical software. This course is an elective intended for PhD students in biostatistics but may also be suitable for highly advanced MS students in biostatistics. PhD students with a strong mathematic background from other departments are also welcomed to choose the course under the permission of the instructor.

Course Objectives and/or Goals

- Discuss the theoretical foundation of statistical learning methods
- Apply advanced methods, such as kernel smoothing and neural networks, to biomedical and public health research problems
- Weigh approaches when analyzing complex biomedical and public health data using key concepts (e.g., bias–variance tradeoff) in statistical learning
- Implement and revise statistical learning methods in R to address challenges in biomedical and public health research
- Evaluate and determine methods that are appropriate for various purposes

Description of Course Content

Topical Outline/Course Schedule

The course will cover major concepts and a variety of methods in the field of statistical learning.

Week	Date(s)	Topic(s)	Readings
1	1/9-1/11	Introduction to Statistical Learning	
2	1/16	Martin Luther King Jr. Day	
2-4	1/18-1/23, 1/25	Estimation and Inference in Linear Regression: From Low-dimension to High-dimension	1. Tibshirani R. Regression Shrinkage and Selection via the Lasso 2. Yang et al. A General Theory of Hypothesis Tests and Confidence Regions
4-5	1/30, 2/1-2/6	Classification: Linear Discriminant Analysis (LDA) and Empirical Risk Minimization (ERM) Homework 1	3. Haykin et al. Rosenblatt's Perceptron 4. Peter et al. Convexity, Classification, and Risk Bounds
6	2/8-2/13	Basis expansion and Regularization	5. Girosi et al. Regularization Theory
7	2/15-2/20	Model Assessment and Selection	6. Golub et al. Generalized Cross-Validation Ridge
8	2/22-2/27	Model inference and averaging Homework 2	7. Wolpert DH. Stacked Generalization 8. Tibshirani et al. Bumping
9	3/2-3/6	Additive models, trees and related methods	9. Friedman et al. PRIM 10. Friedman JH. MARS
10	3/8	Midterm Exam (take home)	
11	3/13-3/15	Spring Break	
12	3/20-3/22	Boosting and additive trees, Random forests, and Ensemble learning	11. Freund et al. AdaBoost 12. Breiman L. Random Forests 13. Dietterich TG. Ensemble methods
13	4/3-4/5	Other Supervised-Learning Methods: Neural Networks and Deep Learning (with interactive session) Homework 3	14. Lipton et al. Recurrent Neural Networks 15. LeCun et al. Deep Learning Review
14	4/10-4/12	Special Topics in Statistical Learning I: Transfer Learning with its application in Biomedical Research	16. Sai et al. Transfer Learning for High-Dimensional Linear Regression
15	4/17-4/19	Special Topics in Statistical Learning II: Semi-supervised Learning, and Reinforcement Learning	17. Jue et al. Surrogate Assisted Semi-Supervised Inference
16	4/24-4/26	Special Topics in Statistical Learning III: Federated Learning, and Distributed Learning Homework 4	
17	5/1-5/3	Final Projects	

Course Materials and Technology

The course is developed based on two required textbooks, both of which are freely available online:

- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: Springer. (Chapters 1-5,7,8-10,12)
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press. (Chapters 6,9,10,15)

Full citation information of reading material:

1. Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996): 267-288.
2. Ning, Yang, and Han Liu. "A general theory of hypothesis tests and confidence regions for sparse high dimensional models." *The Annals of Statistics* 45.1 (2017): 158-195.
3. Simon Haykin. "Neural networks and learning machines." Pearson; 3rd edition (2018): Chapter 1.
4. Bartlett, Peter L., Michael I. Jordan, and Jon D. McAuliffe. "Convexity, classification, and risk bounds." *Journal of the American Statistical Association* 101.473 (2006): 138-156.
5. Girosi, F. (1994). Regularization Theory, Radial Basis Functions and Networks. In: Cherkassky, V., Friedman, J.H., Wechsler, H. (eds) From Statistics to Neural Networks. NATO ASI Series, vol 136. Springer, Berlin, Heidelberg.
6. Golub, Gene H., Michael Heath, and Grace Wahba. "Generalized cross-validation as a method for choosing a good ridge parameter." *Technometrics* 21.2 (1979): 215-223.
7. Wolpert, David H. "Stacked generalization." *Neural networks* 5.2 (1992): 241-259.
8. Tibshirani, Robert, and Keith Knight. "Model search by bootstrap "bumping". " *Journal of Computational and Graphical Statistics* 8.4 (1999): 671-686.
9. Friedman, Jerome H., and Nicholas I. Fisher. "Bump hunting in high-dimensional data." *Statistics and computing* 9.2 (1999): 123-143.
10. Friedman, Jerome H., and Charles B. Roosen. "An introduction to multivariate adaptive regression splines." *Statistical methods in medical research* 4.3 (1995): 197-217.
11. Freund, Yoav, and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting." *Journal of computer and system sciences* 55.1 (1997): 119-139.
12. Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.
13. Dietterich, Thomas G. "Ensemble methods in machine learning." *International workshop on multiple classifier systems*. Springer, Berlin, Heidelberg, 2000.
14. Lipton, Zachary C., John Berkowitz, and Charles Elkan. "A critical review of recurrent neural networks for sequence learning." *arXiv preprint arXiv:1506.00019* (2015).
15. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553 (2015): 436-444.
16. Li, Sai, T. Tony Cai, and Hongzhe Li. "Transfer learning for high-dimensional linear regression: Prediction, estimation, and minimax optimality." *arXiv preprint arXiv:2006.10593* (2020).
17. Hou, Jue, Zijian Guo, and Tianxi Cai. "Surrogate Assisted Semi-supervised Inference for High Dimensional Risk Prediction." *arXiv preprint arXiv:2105.01264* (2021).

Statistical Software:

We will mainly use R in this course. R is free and you can download R from <http://www.r-project.org/>.

Rstudio is a recommended interface for the R software. It is also free and can be downloaded from

<http://www.rstudio.org>. R packages related to this course can be found under <https://cran.r-project.org/web/packages/ElemStatLearn/index.html>

For technical support for this class, please contact the UF Help Desk at:

- Learning-support@ufl.edu
- (352) 392-HELP - select option 2
- <https://lss.at.ufl.edu/help.shtml>

Academic Requirements and Grading

Grading The course will be evaluated based on homework (20%), midterm exam (30%), and a final project (40%). Professionalism (e.g., attendance and being involved in discussions) will account for 10% of the final grade.

Homework There are total 4 homework assignments, which are based on the exercises in the textbook. Discussion on homework problems is allowed, but plagiarism is prohibited. Full credit will be given for assignments turned in on the due date (by 11:59pm) or late submissions approved by the instructor on or before the due day due to reasons consistent with university policies in the Graduate Catalog (<https://gradcatalog.ufl.edu/graduate/regulations/>). Reduced credit (50% point reduction if submitted after the due date) will be given for assignments turned in after the due date; the assignments submitted one-week after the due date will not be graded.

Midterm exam There will be one midterm exam. The midterm exam will be take-home exam. Students will prepare neatly typed exams in Word or LaTeX that can be submitted online in Canvas or email to the instructor. The take-home exam will consist of 8-10 questions and one bonus question covering content during lectures in Weeks 1-9. For proofs or derivations presented during lectures, students are expected to repeat the proof or derivations in simple examples. At least one of the questions will be the homework question. Students are not allowed to work together on the midterm exam. Questions about the midterm exam should be directed to the instructor as early as possible, at least 24 hours before the exam is due.

Final Project For the final project, each student will work on and present a project based on 1) the analysis of a real dataset using existing methods/software, or 2) comparing existing methods by simulations. Each presentation is expected to consist of 25-minutes for presentation and 5-minutes for questions. The grade is given based on the quality of the project (e.g., an innovative way of using an existing method or comprehensively compare several methods), and the presentation of the project (i.e., delivering an easy-to-follow and informative talk). The quality of the project and the presentation of the project are of equal importance (50% of the total point for each).

Professionalism Full points (i.e., 10 points) will be given to students who attend the class on time, and reduced points will be given to students who miss classes (50% of the total point will be deducted for each class). Cell phones should be silenced, and laptops should be turned off during class unless needed. Students are also encouraged to be actively engaged in group discussions and ask questions during class.

Requirement	Due date	Points or % of final grade (% must sum to 100%)
Homework 1 (Lectures 1-4)	2/13	5%
Homework 2 (Lectures 5-7)	3/6	5%
Homework 3 (Lectures 8-10)	4/5	5%
Homework 4 (Lectures 11-12)	4/23	5%
Midterm Exam	3/8	30%
Final project	4/24-4/26	40%
Professionalism	1/11-4/19	10%

The numerical final score will be converted to the letter grades according to the following scale and cutoffs:

Points Earned	Letter Grade
93-100	A
90-92	A-
87-89	B+
83-86	B
80-82	B-
77-79	C+
73-76	C
70-72	C-

More information on UF grading policy may be found at:

<http://gradcatalog.ufl.edu/content.php?catoid=10&navoid=2020#grades>

Policy Related to Make up Work and Professionalism

Policy Related to Make up Work

Please note: Any requests for make-ups due to technical issues MUST be accompanied by the UF Computing help desk (<http://helpdesk.ufl.edu/>) correspondence. You MUST e-mail me within 24 hours of the technical difficulty if you wish to request a make-up.

Policy Related to Professionalism

Students are required to attend the class on time. Cell phones should be silenced and laptops should be turned off during class unless needed. Students are also encouraged to be actively engaged in classes, asking questions and being involved in discussions.

Policy Related to Required Class Attendance

Students are expected to be present for **all** classes since much material will be covered only once in class. Attendance will be taken during each class.

As a matter of mutual courtesy, please let the instructor know when you're going to be late or if you need to leave early. Please try to do any of these as little as possible. Students who have extraordinary circumstances preventing attendance or who must leave early should explain these circumstances to the course instructor prior to the scheduled class, or as soon as possible thereafter. The instructor will then try to accommodate reasonable requests. All excused absences must be consistent with university policies in the Graduate Catalog (<https://gradcatalog.ufl.edu/graduate/regulations/>).

If the requirement to attend this course conflict with religious observances, affected students could contact the instructor in advance of the holiday so that any needed arrangements to make up material or activities covered in their absence could be made. Students will not be penalized due to absence from class because of their religious observances.

Student Expectations, Roles, and Opportunities for Input

Expectations Regarding Course Behavior

Students are expected to show up for class prepared and on time. Cell phones are to be silenced during class unless there is an emergency, in which case please inform the instructor.

Communication Guidelines

The preferred methods of communication for the course are messages in e-learning or e-mail.

Academic Integrity

Students are expected to act in accordance with the University of Florida policy on academic integrity. As a student at the University of Florida, you have committed yourself to uphold the Honor Code, which includes the following pledge:

“We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity.”

You are expected to exhibit behavior consistent with this commitment to the UF academic community, and on all work submitted for credit at the University of Florida, the following pledge is either required or implied:

“On my honor, I have neither given nor received unauthorized aid in doing this assignment.”

It is your individual responsibility to know and comply with all university policies and procedures regarding academic integrity and the Student Honor Code. Violations of the Honor Code at the University of Florida will not be tolerated. Violations will be reported to the Dean of Students Office for consideration of disciplinary action. For additional information regarding Academic Integrity, please see Student Conduct and Honor Code or the Graduate Student Website for additional details:

<https://www.dso.ufl.edu/sccr/process/student-conduct-honor-code/>

<http://gradschool.ufl.edu/students/introduction.html>

Please remember cheating, lying, misrepresentation, or plagiarism in any form is unacceptable and inexcusable behavior.

Recording Within the Course:

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A “class lecture” is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To “publish” means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform,

including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.

Online Faculty Course Evaluation Process

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at <https://gatorevals.aa.ufl.edu/students/>. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via <https://ufl.bluera.com/ufl/>. Summaries of course evaluation results are available to students at <https://gatorevals.aa.ufl.edu/public-results/>.

Policy Related to Guests Attending Class

Only registered students are permitted to attend class. However, we recognize that students who are caretakers may face occasional unexpected challenges creating attendance barriers. Therefore, by exception, a department chair or his or her designee (e.g., instructors) may grant a student permission to bring a guest(s) for a total of two class sessions per semester. This is two sessions total across all courses. No further extensions will be granted. Please note that guests are **not** permitted to attend either cadaver or wet labs. Students are responsible for course material regardless of attendance. For additional information, please review the Classroom Guests of Students policy in its entirety. Link to full policy: <http://facstaff.php.ufl.edu/services/resourceguide/getstarted.htm>

Policy Related to Guests Attending Class

Only registered students are permitted to attend class. However, we recognize that students who are caretakers may face occasional unexpected challenges creating attendance barriers. Therefore, by exception, a department chair or his or her designee (e.g., instructors) may grant a student permission to bring a guest(s) for a total of two class sessions per semester. This is two sessions total across all courses. No further extensions will be granted. Please note that guests are **not** permitted to attend either cadaver or wet labs. Students are responsible for course material regardless of attendance. For additional information, please review the Classroom Guests of Students policy in its entirety. Link to full policy: <http://facstaff.php.ufl.edu/services/resourceguide/getstarted.htm>

SUPPORT SERVICES

Accommodations for Students with Disabilities

If you require classroom accommodation because of a disability, it is strongly recommended you register with the Dean of Students Office <http://www.dso.ufl.edu> within the first week of class or as soon as you believe you might be eligible for accommodations. The Dean of Students Office will provide documentation of accommodations to you, which you must then give to me as the instructor of the course to receive accommodations. Please do this as soon as possible after you receive the letter. Students with disabilities should follow this procedure as early as possible in the semester. The College is committed to providing reasonable accommodations to assist students in their coursework.

Counseling and Student Health

Students sometimes experience stress from academic expectations and/or personal and interpersonal issues that may interfere with their academic performance. If you find yourself facing issues that have the potential to or are already negatively affecting your coursework, you are encouraged to talk with an instructor and/or seek help through University resources available to you.

- The Counseling and Wellness Center 352-392-1575 offers a variety of support services such as psychological assessment and intervention and assistance for math and test anxiety. Visit their web site for more information: <http://www.counseling.ufl.edu>. On line and in person assistance is available.
- You Matter We Care website: <http://www.umatter.ufl.edu/>. If you are feeling overwhelmed or stressed, you can reach out for help through the You Matter We Care website, which is staffed by Dean of Students and Counseling Center personnel.
- The Student Health Care Center at Shands is a satellite clinic of the main Student Health Care Center located on Fletcher Drive on campus. Student Health at Shands offers a variety of clinical services. The clinic is located on the second floor of the Dental Tower in the Health Science Center. For more information, contact the clinic at 392-0627 or check out the web site at: <https://shcc.ufl.edu/>
- Crisis intervention is always available 24/7 from:
Alachua County Crisis Center:
(352) 264-6789
<http://www.alachuacounty.us/DEPTS/CSS/CRISISCENTER/Pages/CrisisCenter.aspx>
- **University Police Department:** [Visit UF Police Department website](#) or call 352-392-1111 (or 9-1-1 for emergencies).
- **UF Health Shands Emergency Room / Trauma Center:** For immediate medical care call 352-733-0111 or go to the emergency room at 1515 SW Archer Road, Gainesville, FL 32608; [Visit the UF Health Emergency Room and Trauma Center website](#).

Do not wait until you reach a crisis to come in and talk with us. We have helped many students through stressful situations impacting their academic performance. You are not alone so do not be afraid to ask for assistance.

Inclusive Learning Environment

Public health and health professions are based on the belief in human dignity and on respect for the individual. As we share our personal beliefs inside or outside of the classroom, it is always with the understanding that we value and respect diversity of background, experience, and opinion, where every individual feels valued. We believe in, and promote, openness and tolerance of differences in ethnicity and culture, and we respect differing personal, spiritual, religious and political values. We further believe that celebrating such diversity enriches the quality of the educational experiences we provide our students and enhances our own personal and professional relationships. We embrace The University of Florida's Non-Discrimination Policy, which reads, "The University shall actively promote equal opportunity policies and practices conforming to laws against discrimination. The University is committed to non-discrimination with respect to race, creed, color, religion, age, disability, sex, sexual orientation, gender identity and expression, marital status, national origin, political opinions or affiliations, genetic information and veteran status as protected under the Vietnam Era Veterans' Readjustment Assistance Act." If you have questions or concerns about your rights and responsibilities for inclusive learning environment, please see your instructor or refer to the Office of Multicultural & Diversity Affairs website: www.multicultural.ufl.edu