RUTGERS
School of Public Health

Course Title: Statistical Learning for Biomedical Studies

Course Number: **BIST 0663J030**

Course Location: Room 2A, School of Public Health, Piscataway, NJ

Course Date & Time: Thursday, 3:00 – 5:00 PM

Course Instructor: Liangyuan Hu, PhD, Associate Professor,
Biostatistics, Rutgers School of Public Health,
liangyuan.hu@rutgers.edu & 732-235-4664

Office Hours: Thursday, 2:00 - 2:50PM, Room 209

Required Course Text: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. *An introduction to Statistical Learning*, 2013, Springer.

Additional/Supplemental Readings/Resources:

(1) Trevor Hastie, Robert Tibshirani, Jerome Friedman. *The Elements of Statistical Learning*, 2nd edition (2016), Springer
(2) Ian Goodfellow, Yoshua Bengio, Aaron Courville. *Deep Learning*, 2016, The MIT Press.
(3) Fan, Jianqing, Runze Li, Cun-Hui Zhang, and Hui Zou. *Statistical foundations of data science*. 2020, Chapman and Hall/CRC.

Course Description: This is an intermediate level course of Statistical Learning. It emphasizes the theoretical concepts and applications of statistical learning methods for biomedical studies. It is taught at BIST MS level and also suitable for BIST doctoral students. It covers supervised and unsupervised learning methods, including Ridge regression, LASSO, classification, decision tree, random forest, SVM, neural network, deep learning, PCA and clustering models. Model building, model assessment and selection, and interpretation as well as theoretical properties of some models will be taught. If time permits, advanced topics such as Bayesian machine learning will also be covered. R will be used for data analysis.

Selected Concentration Competencies Addressed: The competencies addressed in this course for the MS for the Concentration of *Biostatistics* include:

1. Use statistical computer packages to organize, analyze and report collected data;
2. Apply statistical methods to biomedical, clinical and public health research;
3. Communicate the results of statistical studies both in writing and orally to investigators and lay community members.

Please visit the Concentration webpages on the School of Public Health's website at http://sph.rutgers.edu/ for additional competencies addressed by this course for other degrees and concentrations.

Course Objectives: By the completion of this course, students will be able to:

a. Understand the concepts of machine learning / data mining;
b. Use mathematical expressions to generalize the concepts and methods.
c. Develop the ability to apply these concepts correctly using statistical / machine learning software;
d. Develop the ability to interpret the results of an analysis properly; and
e. Become well-versed in the application ofcore learning techniques (Supervised learning, unsupervised learning, model averaging, parallel computing, etc.)

Competency	Course Objectives(s)	Lessons	Assessment(s)
1	a, b, c	1-14	Assignments 1 – 4, Data analysis Project
2	b, c, d, e	1-14	Assignments 1 – 4, Data analysis Project
3	d, e	1-14	Assignments 1 – 4, Data analysis Project

Special Circumstances During COVID-19, For Fall 2022 (Version Date 8/4/22)
To keep our on-campus communities safe, compliance with all current guidance and policies as set forth in the Guide to Returning to Rutgers is required at all times and without exception. Students, faculty, staff, or visitors who do not comply with these policies will not be permitted to remain on-site. The use of face-coverings indoors *IS* required in classrooms and offices as well as shared spaces (such as hallways and bathrooms). Rutgers employees and students must use the My Campus Pass symptom checker, a self-screening application, each day when traveling to campus or entering a Rutgers building. Please remember to wash your hands, wear a mask while indoors, particularly in crowded spaces and groups, and stay up-to-date on university guidance by consulting the Guide to Returning to Rutgers and the university's COVID-19 website.

Course Requirements and Grading:

- Course evaluation
  1. Homework (4 assignments) 30 %
  2. Lab activities (Online 30) 25 %
  3. Data analysis projects 45 %
  Total: 100 %

- Data analysis project
  The goal is to join a machine learning competition using methods learned in this course, to give presentations (30 minutes) in class and write a report based on results. Students can try out all the different learning approaches (at least 2) that have been covered in class. The project could be a group (2 students per group) project or solo. It will be graded asfollowing:

1. First submission, presentation (5-10 minutes) in week 6, 10%.
2. Appropriateness of methods applied and overall performance, 25%.
3. Final presentation (20 - 30 minutes), 25%
4. Final report, 40%

For the final data analysis project, students should pick a dataset from https://archive.ics.uci.edu/ml/index.php, identify a research question and submit a proposal for approval. The proposal (one page long) should include a full description of the problem and the dataset: dimensions, names of variables with their description. Is it a classification or regression problem? supervised or unsupervised? List names of group members.

- Grading policy
  1. Homework will be collected at the beginning of lecture on due date. Unless notifying the instructor beforehand, later submission of homework will NOT be graded.
  2. On all homework assignments/problem sets, students are encouraged to discuss with one another, but **work should be carried out and written up independently**. If any two identical write-ups are found, both homework assignments are considered failed.
  3. The students are asked to answer each question as accurately and concisely as possible. **If it is necessary to attach the computer output with the homework assignment, ONLY the "essential" segments are required.** DO NOT SUBMIT the complete output section or the log file. Otherwise, 50% of the points will be taken away.

4. Grading Scale:

94 – 100	A
90 – <94	A-
87 – <90	B+
84 – <87	B
80 – <84	B-
77 – <80	C+
70 – <77	C
<70	F

Tentative Course Schedule: This table provides a general plan for the course; some deviations may be necessary.

	Date	Week	Topic	Online 30
Session 1: Introduction	09/08	1	Introduction, linear regression, R basics. Ch. 1, 2, 3	Lab 1.

and review	09/15	2	Model assessment and selection, cross validation Ch. 5 Announcement of data analysis project	Lab 2.
Session 2: Supervised learning	09/22	3	High dimensional data; feature selection; Ridge, Lasso Ch.6; HW1 Assigned	Lab 3.
	09/29	4	Classification; Logistic regression, KNN, LDA, QDA Ch. 4	Lab 4.
	10/06	5	Decision tree, Bagging, boosting, random forest Ch. 8	Lab 5.
	10/13	6	Presentation (50 minutes). Maximal Margin classifier, Support Vector Machines Ch. 9; HW1 Due HW2 Assigned	Lab 6.
	10/20	7	Deep learning, neural networks Ch. 10	Lab 7.
	10/27	8	Bayesian additive regression trees; Missing data; Causal inference Ch. 8 and additional materials	Lab 8.
Session 3: Unsupervised learning	11/03	9	Parallel computing introduction, Parallel computing with R, HPC Handout HW2 Due HW3 Assigned	Lab 9.

	11/10	10	Big data and cloud, Hadoop and Spark, MapReduce Handout	Lab 10.
Session 4: Advanced computing	11/17	11	PCA, K-Means Clustering Ch. 10;	Lab 11.
	11/22	12	Hierarchical Clustering, mixture model Ch. 10 HW3 Due HW4 Assigned	Lab 12.
	11/24		No class, Thanksgiving	
	12/01	13	Presentation I	
	12/08	14	Presentation II	
	12/15	15	Project report due, HW4 Due	

Learning Management System: Canvas will be used extensively throughout the semester for course syllabus, assignments, announcements, communication and/or other course-related activities. It is the student's responsibility to familiarize themselves with Canvas and check it regularly. If you have difficulties accessing Canvas, please inform the instructor and Canvas Support (help@canvas.rutgers.edu). Canvas is accessible at https://tlt.rutgers.edu/canvas.

School of Public Health Honor Code: The School of Public Health Honor Code is found in the student bulletin (sph.rutgers.edu/academics/catalog/index.html). Each student bears a fundamental responsibility for maintaining academic integrity and intellectual honesty in his or her graduate work. For example, all students are expected to observe the generally accepted principles of scholarly work, to submit their own rather than another's work, to refrain from falsifying data, and to refrain from receiving and/or giving aid on examinations or other assigned work requiring independent effort. In submitting written material, the writer takes full responsibility for the work as a whole and implies that, except as properly noted by use of quotation marks, footnotes, etc., both the ideas and the works used are his or her own. In addition to maintaining personal academic integrity, each student is expected to contribute to the academic integrity of the school community by not facilitating inappropriate use of her/his own work by others and by reporting acts of academic dishonesty by others to an appropriate school authority. It should be clearly understood that plagiarism, cheating, or other forms of academic dishonesty will not be tolerated and can lead to sanctions up to and including separation from the Rutgers School of Public Health.

Students with Disabilities: Rutgers University welcomes students with disabilities into all of the University's educational programs. In order to receive consideration for reasonable accommodations, a student must Apply for Services by first completing a Registration Form with the Rutgers Office of Disability Services (ODS) at ods.rutgers.edu. The student will also be required to participate in an ODS intake interview and provide documentation. If reasonable accommodations are granted, ODS will provide you with a Letter of Accommodations which should be shared with your instructors as early in your courses as possible.

Commitment to Safe Learning Environment: The Rutgers School of Public Health is committed to helping create a safe learning environment for all students and for the School as a whole. Free expression in an academic community is essential to the mission of providing the highest caliber of education possible. The School encourages civil discourse, reasoned thought, sustained discussion, and constructive engagement. Provocative ideas respectfully presented are an expected result. An enlightened academic community, however, connects freedom with responsibility. The School encourages all students to disclose any situations where you may feel unsafe, discriminated against, or harassed. Harassment or discrimination of any kind will be not tolerated and violations may lead to disciplinary actions.

Student Well-Being: The School of Public Health recognizes that students may experience stressors or challenges that can impact both their academic experience and their personal well-being. If the source of your stressors or challenges is academic, students are encouraged to discuss these challenges and circumstanceswith their instructor, if they feel they may need additional support or temporary accommodations at the beginning or during this course. The course instructor may consider making reasonable temporary adjustments depending on the student's situation. For personal concerns or if additional support is needed, students may reach out to the Office of Student Affairs (studentaffairs@sph.rutgers.edu) or any of the appropriate referral resources listed on the SPH Student Connect Canvas page.

Reporting Discrimination or Harassment: If you experience any form of gender or sex-based discrimination or harassment, including sexual assault, sexual harassment, relationship violence, or stalking, know that help and support are available. You may report such incidents to the RBHS Title IX Office or to the School of Public Health's Office of Student Affairs. Rutgers University has staff members trained to support survivors in navigating campus life, accessing health and counseling services, providing academic and housing accommodations, and more. If you experience any other form of discrimination or harassment, including racial, ethnic, religious, political, or academic, please report any such incidents to the School's Office of Student Affairs. The School strongly encourages all students to report any incidents of discrimination or harassment to the School. Please be aware that all Rutgers employees (other than those designated as confidential resources such as advocates, counselors, clergy and healthcare providers as listed in Appendix A to Policy 10.3.12) are required to report information about such discrimination and harassment to the School and potentially the University. For example, if you tell a faculty or staff member about a situation of sexual harassment or sexual violence, or other related misconduct, the faculty or staff member must share that information with the RBHS Title IX Coordinator. If you wish to speak to a confidential employee who does not have this reporting responsibility, you can find a list of resources in Appendix A to University Policy 10.3.12. For more information about your options at Rutgers, please visit Rutgers Violence Prevention and Victim Assistance.

Overview of School Policies: Academic and non-academic policies and procedures, such Auditing a Course, Retaking Courses, Grade Grievance and others that cover registration, courses and grading, academic standing and progress, student rights and responsibilities, graduation and more may be found under Policies on the School of Public Health website. Below are select specific policies; however, students are responsible for keeping informed about academic and non-academic policies and procedures beyond those noted on this syllabus.

Graduate Student Computer Policy: Students are required to possess a personal laptop, no older than approximately two years, that must meet minimum requirements which may be found online at: sph.rutgers.edu/student_life/computer_requirements.html

Policy Concerning Use of Recording Devices and Other Electronic Communications Systems: When personally owned communication/recording devices are used by students to record lectures and/or classroom lessons, such use must be authorized by the faculty member or instructor who must give either oral or written permission prior to the start of the semester and identify restrictions, if any, on the use of mobile communications or recording devices.

Policy Concerning Use of Turnitin: Students agree that by taking this course all required papers may be subject to submission for textual similarity review to Turnitin.com (directly or via learning management system, i.e. Canvas) for the detection of plagiarism. All submitted papers will be included as source documents in the Turnitin.com reference database solely for the purpose of detecting plagiarism of such

papers. Use of the Turnitin.com service is subject to the Usage Policy posted on the Turnitin.com site. Students who do not agree should contact the course instructor immediately.

Withdrawal/Refund Schedule: Students who stop attending their course(s) without processing an Add/Drop Course form will receive a failing grade. Furthermore, students dropping to zero credits for the semester are considered withdrawn and must submit a completed Leave of Absence form from the School of Public Health's Office of Student Affairs. The School of Public Health refunds tuition only. Administrative and technology fees are non-refundable. You may find the Withdrawal/Refund Schedule on the School of Public Health website at:
sph.rutgers.edu/academics/registration/school_calendars.html