ILLINOIS
Economics
COLLEGE OF LIBERAL ARTS & SCIENCES

ECONOMETRICS LAB
QUANTITATIVE RESEARCH @ UIUC

ECO 491
Applied Machine Learning

Course Syllabus

Credits: 3 (three) undergraduate hours or 4 (four) graduate hours
Semester: Fall 2023
Meeting Times: Mondays and Wednesdays, from 9:30 am to 10:50 am
Instructor: Marcelo C. Medeiros
Email: marcelom@illinois.edu
Office Hours: Mondays, from 1:00 pm to 2:30 pm

Course Description

This course covers the most recent advances in **Statistical Machine Learning Theory** and explores several applications in finance, economics, and related areas. The methods presented during the course can be applied to prediction problems, causal inference, and text analysis. Topics covered include (1) dimensionality reduction, principal components, and factor models; (2) shrinkage estimation methods, such as Ridge and Lasso (with extensions); (3) nonlinear models, such as Random Forests and Neural Networks; (4) Deep Learning; (5) Ensemble methods, like bagging and boosting; (6) clustering; (7) text analysis and unstructured data; and (8) reinforcement learning. The course will be hands-on, and the theory will be illustrated with empirical applications in economics, finance, and related areas.

Prerequisites

It is strongly recommended that the students know calculus, linear algebra, and regression analysis. A quick review of the main concepts of the linear regression model will be provided at the beginning of the course. **Intermediate knowledge of a programming language like R, Python, Matlab, or Julia is mandatory**.

Recommended prerequisites: ECO 203, MATH 112 (MATH 225 would be better), MATH 220 (or MATH 221).

Learning Outcomes

This is a hands-on course. At the end of the semester, students must be able to successfully apply cutting-edge machine-learning techniques to a wide range of potential practical problems. Examples are economic/financial forecasting, causal inference and policy evaluation, optimal policy design, optimal pricing, portfolio optimization, and risk analysis. Furthermore, the students will learn how to prepare a presentation summarizing their main findings.

Learning Resources and Bibliography

Lecture notes (slides) are the primary reference for this course and will be distributed through **Canvas** (https://canvas.illinois.edu/). For each lecture, we will point to the relevant bibliographical references and any other resource of potential interest.

For the interested reader, the following books are recommended:

1. Hastie, T., R. Tibshirani and J. Friedman (2009). **The Elements of Statistical Learning: Data Mining, Inference and Prediction.** Springer. In the lecture notes (slides), I will use the acronym HTF (2009) to refer to this book. An official PDF of the book can be obtained from Trevor Hastie's webpage at Stanford University: http://web.stanford.edu/~hastie/pub.htm

2. James, G., D. Witten, T. Hastie and R. Tibshirani (2021). **An Introduction to Statistical Learning with Applications in R.** Second Edition. Springer. In the lecture notes (slides), I will use the acronym JWHT (2021) to refer to this book. https://www.statlearning.com/

3. James, G., D. Witten, T. Hastie, R. Tibshirani and J. Taylor (2023). **An Introduction to Statistical Learning with Applications in Python.** First Edition. Springer. In the lecture notes (slides), I

will use the acronym JWHTT (2023) to refer to this book. https://www.statlearning.com/

4. Murphy, K. P. (2022). Probabilistic Machine Learning: An Introduction. MIT Press. In the lecture notes (slides), I will refer to this book by the acronym M (2022). https://probml.github.io/pml-book/book1.html

5. Murphy, K. P. (2023). Probabilistic Machine Learning: Advanced Topics. MIT Press. In the lecture notes (slides), I will refer to this book by the acronym MA (2023). https://probml.github.io/pml-book/book2.html

For the ones interested in going deeper into the theory, there are also several recommended books:

6. Wainwright, M. (2019). **High-Dimensional Statistics. A Non-Asymptotic Viewpoint**. Cambridge University Press. In the lecture notes (slides), I will refer to this book by the acronym W (2019).

7. Fan, J., Li, R., Zhang, C.-H., and Zou (2020). **Statistical Foundations of Data Science**. CRC Press. In the lecture notes (slides), I will use the acronym FLZZ (2020) to refer to this book.

8. Hastie, T., R. Tibshirani and M. Wainwright (2015). **Statistical Learning with Sparsity: The Lasso and Generalizations**. CRC Press. In the lecture notes (slides), I will use the acronym HTW (2015) to refer to this book.

9. Bülhmann, P., and S. van der Geer (2011). **Statistics for High-Dimensional Data.** Springer. In the lecture notes (slides) we will use the acronym BvdG (2011) to refer to this book.

The empirical applications discussed in the lectures will be partially based on academic papers that will be distributed at the beginning of the course.

Grading

This class uses a plus/minus grading system. Letter grades will be assigned only at the end of the semester based on the overall score. **There is no curve in this class**.

Students taking the course for three credits.

Homework assignments are 40%, the midterm exam is 30%, and the take-home final is 30% of the grade. The final will be a short empirical project. Class participation will be the deciding criterion for students who fall close to a grade cutoff.

Students taking the course for four credits.

Homework assignments are 20%, the midterm exam is 30%, the take-home final is 30%, the and final project is 20%. The final will be a short empirical project. Students taking the course for four credits must deliver a longer report (essay) based on the findings of the final project. Class participation will be the deciding criterion for students who fall close to a grade cutoff.

Plus/Minus Grade Cutoffs:

A+	97-100	B+	87-89	C+	77-79	D+	67-69	F	0-59
A	93-96	B	83-86	C	73-76	D	63-66		
A-	90-92	B-	80-82	C-	70-72	D-	60-62		

Assignments, Exams, and Final Essay

Homework Assignments: There will be **four homework assignments**. Assignments will be posted on the Canvas course site. Unless indicated otherwise, your answers to the homework assignments must be completed **in groups of four**. Your answers must be typed (or handwritten and scanned) **and uploaded in Canvas in PDF format** by the due date and time indicated on the schedule below. Late homework will be penalized. Please use the following convention to name your PDF file:

ECO491_HW[number]_LastName_FirstName.pdf

The assignments will cover the material presented during the lectures and will be a mix of computer implementations and theoretical questions to complement the results shown by the instructor.

The schedule of the assignments is as follows:

Assignment 1. Delivery to the students: Aug 21. Due date: Sep 13. Review of regression and discrete-choice models and applications of factor models. It can be delivered as Python notebooks, R markdown, Matlab report generator, or typed in a word processing software. The students may choose the programming language that they are more comfortable with.

Assignment 2. Delivery to the students: Sep 13. Due date: Oct 11. This assignment will consist of practical questions about factor models, penalized regression techniques (Ridge, Lasso, and extensions), and their uses for causal inference.

Assignment 3. Delivery to the students: Oct 11. Due date: Nov 08. Applications of nonlinear machine-learning models.

Assignment 4. Delivery to the students: Nov 08. Due date: Dec 06. This assignment will be about clustering methods.

Midterm exam. October 25, in class. The midterm in this class is open notes/book. You can consult any other sources during the exam, as long as it is hard copy. You can only use hard copy material, not digital. The exam will consist of 5 (five) questions that must be answered in 90 minutes. A practice lecture will take place on October 23.

Final exam. Delivery to the students: Dec 06. Due date: Dec 13. The final will consist of a short empirical project. Students taking the course of 3 (three) credits can deliver the answers in Python notebooks, R markdown, Matlab report generator, or type in a word processing software. **Students taking the course for 4 (four) credits must also write a longer report in a format of an academic paper. The report must have the following sections: Introduction, Problem Statement, Methodology, Empirical Results, and Conclusion. The final essay cannot be longer than 30 pages including figures, tables and references.**

Academic Assistance

Students are encouraged to utilize the many resources we have throughout campus to assist with academics.

We recommend that you seek them out starting early in the semester, not just in times of academic need, to develop good study habits and submit work that represents your full academic potential. Many resources are found on the Economics Website, including details about the Economics Tutoring Center, Academic Advising, and other academic support options: https://economics.illinois.edu/academics/undergraduate-program/academic-student-support

Academic Integrity

According to the Student Code, `It is the responsibility of each student to refrain from infractions of academic integrity, from conduct that may lead to suspicion of such infractions, and from conduct that aids others in such infractions.' Please know that it is my responsibility as an instructor to uphold the academic integrity policy of the University, which can be found here: https://studentcode.illinois.edu/article1/part4/1-401/ Academic dishonesty may result in a failing grade. Every student is expected to review and abide by the Academic Integrity Policies. It is your responsibility to read this policy to avoid any misunderstanding. Do not hesitate to ask the instructor(s) if you are ever in doubt about what constitutes plagiarism, cheating, or any other breach of academic integrity. **Read the full Student Code at** https://studentcode.illinois.edu/

Students with Disabilities

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES you may visit 1207 S. Oak St., Champaign, call 333-4603 (V/TTY), or e-mail a message to disability@illinois.edu. DRES Website: www.disability.illinois.edu/

Community of Care

As members of the Illinois community, we each have a responsibility to express care and concern for one another. If you come across a classmate whose behavior concerns you, whether in regards to their well-being or yours, we encourage you to refer this behavior to the Student Assistance Center (217-333-0050 or http://odos.illinois.edu/community-of-care/referral/). Based on your report, the staff in the Student Assistance Center reaches out to students to make sure they have the support they need to be healthy and safe. Further, we understand the impact that struggles with mental health can have on your experience at Illinois. Significant stress, strained relationships, anxiety, excessive worry, alcohol/drug problems, a loss of motivation, or problems with eating and/or sleeping can all interfere with optimal academic performance. We encourage all students to reach out to talk with someone, and we want to make sure you are aware that you can access mental health support at the Counseling Center (https://counselingcenter.illinois.edu/) or McKinley Health Center (https://mckinley.illinois.edu/). For mental health emergencies, you can call 911 or walk into the Counseling Center, no appointment needed.

Disruptive Behavior

Behavior that persistently or grossly interferes with classroom activities is considered disruptive behavior and may be subject to disciplinary action. Such behavior inhibits other students' ability to learn and an instructor's ability to teach. A student responsible for disruptive behavior may be required to leave class pending discussion and resolution of the problem and may be reported to the Office for Student Conflict Resolution for disciplinary action.

Emergency Response Recommendations

Emergency response recommendations can be found at the following website: http://police.illinois.edu/emergency-preparedness/. I encourage you to review this website and the campus building floor plans website within the first 10 days of class. http://police.illinois.edu/emergency-preparedness/building-emergency-actionplans/.

Religious Observances

The Religious Observance Accommodation Request form is available at https://odos.illinois.edu/community-of-care/resources/students/religious-observances/. Submit the form to the instructor and to the Office of the Dean of Students (helpdean@illinois.edu) by the end of the second week of the course; in the case of exams or assignments scheduled after this period, students should submit the form to the instructor and to the Office of the Dean of Students as soon as possible.

Family Educational Rights and Privacy Act (FERPA)

Any student who has suppressed their directory information pursuant to Family Educational Rights and Privacy Act (FERPA) should self- identify to the instructor to ensure protection of the privacy of their attendance in this course. See http://registrar.illinois.edu/ferpa for more information on FERPA. Student information and records will only be released to the student if the student has provided written approval or as required by law.

Sexual Misconduct Reporting Obligation

The University of Illinois is committed to combating sexual misconduct. Faculty and staff members are required to report any instances of sexual misconduct to the University's Title IX and Disability Office. In turn, an individual with the Title IX and Disability Office will provide information about rights and options, including accommodations, support services, the campus disciplinary process, and law enforcement options. A list of the designated University employees who, as counselors, confidential advisors, and medical professionals, do not have this reporting responsibility and can maintain confidentiality, can be found here: http://www.wecare.illinois.edu/resources/students/#confidential.
Other information about resources and reporting is available here: http://wecare.illinois.edu/.

Student Support

The Counseling Center is committed to providing a range of services intended to help students develop improved coping skills in order to address emotional, interpersonal, and academic concerns. Please visit their website to find valuable resources and services: https://counselingcenter.illinois.edu/.
Counseling Center Information: 217-333-3704. Location: Room 206, Student Services Building (610 East John Street, Champaign IL)

McKinley Mental Health Information: 217-333-2705
Location: 3rd Floor McKinley Health Center 1109 South Lincoln, Urbana, IL

Emergency Dean: The Emergency Dean may be reached at (217) 333-0050 and supports students who are experiencing an emergency situation after 5 pm, in which an immediate University response is needed and which cannot wait until the next business day. The Emergency Dean is not a substitute for trained emergency personnel such as 911, Police or Fire. If you are experiencing a life threatening emergency, call 911. Please review the Emergency Dean procedures: http://odos.illinois.edu/emergency/

Academic Dates and Deadlines

Students should make note of important academic deadlines for making changes to their courses (add, drop, credit/no-credit, grade replacement, etc.). https://registrar.illinois.edu/academic-calendars
Please check with your academic department regarding specific procedures and policies.

Course Schedule (subject to change with advance notice)

Week	Date	Description	References
1	08/21	**Introduction to big data and machine learning** What is big data? What are the challenges that large datasets impose? Big data in finance and economics. What is machine learning? Is machine learning different from statistics? Types of machine learning: supervised, unsupervised, reinforcement. Examples of machine learning techniques.	Slides M (Ch. 1)
	08/23	**Review of linear regression** Conditional expectation Correlation versus causality The ordinary least-squares (OLS) estimator Inference Empirical examples	Slides HFT (Chs. 2 and 3.1-3.2) JWHT (Ch. 3) JWHTT (Ch. 3) M (Ch. 11.2)
2	08/28	**Review of discrete-choice models** Linear discriminant and classification Probit/Logit models Maximum likelihood estimation Inference Empirical examples	Slides JWHT (Ch. 4) JWHTT (Ch. 4) M (Chs. 9 and 10)
	08/30	**Factor models** Principal component analysis (PCA) Overview of the theory Selection of principal components PCA and factor models Empirical examples	Slides JWHT (Chs. 6.3 and 12.2) JWHTT (Chs. 6.3 and 12.2) M (Chs. 20.1 and 20.2)

3	09/04	**Labor Day**	
	09/06	**Factor models** Principal component regression Empirical examples	Slides JWHT (Chs. 6.3 and 12.2) JWHTT (Chs. 6.3 and 12.2) M (Chs. 20.1 and 20.2)
4	09/11	**Ridge regression** The bias-variance trade-off and James-Stein's estimator Introduction to the ridge regression Ridge in low dimensions and its relationship with principal components	Slides HFT (Ch. 3.4) JWHT (Ch. 6) JWHT (Ch. 6) M (Ch. 11.3)
	09/13	**Ridge Regression** Ridge in high dimensions Kernel Ridge Empirical examples	Slides HFT (Ch. 3.4) JWHT (Ch. 6) JWHT (Ch. 6) M (Ch. 11.3)
5	09/18	**Lasso** Definition and geometry of the LASSO problem Algorithms Theory overview	T Slides HFT (Ch. 3.4) JWHT (Ch. 6) JWHT (Ch. 6) M (Ch. 11.4)
	09/20	**Extensions of the Lasso** Adaptive Lasso Elastic Net Group Lasso Other methods Empirical applications	Slides M (Ch. 11.4)

6	09/25	**Combining Factor and Sparse Models** FarmPredict Network analysis Empirical applications	Slides
	09/27	**Causal Inference** Review of instrumental variables (IV) Review of methods for treatment evaluation	Slides
7	10/02	**Causal Inference with LASSO, Ridge, and Factors** Post model selection inference Double selection Desparsified Lasso	Slides
	10/04	**Causal Inference with LASSO, Ridge, and Factors** IV, factors and LASSO Empirical applications	Slides
8	10/09	**Causal Inference with LASSO, Ridge, and Factors** Panel-based methods for causal inference Difference-in-Differences Synthetic Controls and extensions	Slides
	10/11	**Neural Networks** Sieves and neural networks Shallow feed-forward neural networks Model determination	Slides HFT (Chs. 6 and 11) JWHT (Chs. 7 and 10) JWHTT (Chs. 7 and 10) M (Ch. 13)

9	10/16	**Neural Networks** Deep Neural Networks Overview of the theory Empirical application	Slides HFT (Chs. 6 and 11) JWHT (Chs. 7 and 10) JWHTT (Chs. 7 and 10) M (Ch. 13) MA (Ch. 16)
	10/18	**Neural Networks** Convolutional neural networks Long-short-term memory neural networks	Slides M (Chs. 14 and 15) MA (Ch. 16)
10	10/23	**Midterm Practice**	
	10/25	**Midterm**	
11	10/30	**Bagging** Introduction to bagging and the bootstrap Bagging and unstable estimators Empirical example	Slides
	11/01	**Boosting** Introduction do boosting L2-Boosting Component-wise boosting Boosting versus Lasso Empirical Example	Slides
12	11/06	**Tree-Based Methods** Regression Trees Random Forests Boosted Trees Empirical application	Slides HFT (Chs. 9 and 10) JWHT (Ch. 8) JWHTT (Ch. 8) M (Ch. 18)
	11/08	**Causal inference with nonlinear models** Double machine learning Synthetic learners Empirical examples	Slides

13	11/13	**Interpreting Machine Learning Methods** Measures of variable importance Partial dependence plots and their extensions Empirical examples	Slides
	11/15	**Interpreting Machine Learning Methods** Shapley values Empirical examples	Slides
14	11/20 11/22	**Fall Break**	
15	11/27	**Clustering Methods** K-mean algorithm Hierarchical clustering Empirical application	Slides JWHT (Ch. 12.4) JWHTT (Ch. 12.4) M (Ch. 21)
	11/29	**Natural Language Processing** Structured versus unstructured data Bag-of-words Dictionary methods Empirical applications	Slides M (Ch. 20.5)
16	12/04	**Natural Language Processing** Latent Dirichlet Allocation Empirical applications	Slides MA (Ch. 28.5)
	12/06	**Introduction to Reinforcement Learning** Sequential treatment Bandit problems Policy optimization	Slides MA (Ch. 35)
17	12/13	**Final Exam Deadline**	

Important Notes

- If you are not familiar with Probability Theory, I strongly recommend M (Chs. 2 and 3)
- If you are not familiar with Statistics, I strongly recommend M (Ch. 4)
- An excellent introduction to Linear Algebra and matrix operations is M (Ch. 7)

- If you are unfamiliar with linear regression, I recommend reading chapters 4 and 5 from Stock, J.H. and M.W. Watson (2003). **Introduction to Econometrics**. Addison Wesley.
- You can use ChatGPT (https://chat.openai.com/) and other AI resources to help with coding and assignments. **USE WISELY**.