Prediction and machine learning
[EC524/424/]

Winter 2020 Syllabus
https://github.com/edrubin/EC524W20/

Dr. Edward Rubin
Dept. of Economics, University of Oregon

January 20, 2020

	Instructor	GE
	Edward Rubin	Connor Lennon
	edwardr@uoregon.edu	clennon@uoregon.edu
	Use "EC524" in email subject.	
	PLC 519	PLC 430
	Th., 2p–3p; Fr., 1p–2pm	Mo., 1p–2p
	https://edrub.in	

	Lecture	Lab
	Tu. & Th., 10:00a–11:50a	12:00p–12:50p
	105 Peterson Hall	102 Peterson Hall
	Ed	Connor | Ed
	https://github.com/edrubin/EC524W20/	

Cellphone policy

No phones. **You cannot use your phone in class**—texting included. Offenders will lose 1 percentage point off of their final grade for each offense. If you have a concern about this policy, please contact me via email or discuss in office hours during the first week of classes.

The only exceptions to this rule:

1. actual emergencies
2. in-class activities in which I give you permission to use your phone

If you are watching videos or partaking in other distracting activities on laptops/tablets, I will ask you to leave. **Try to be considerate—don't be a jerk.**[1]

[1] *Note:* This is great advice for life—not just the classroom.

Course summary

Description Following the first course on econometrics and causal inference in this sequence, EC524 turns to examining the **tools available and best practices for predicting outcomes**. Put simply, we are now focusing on $\hat{y}_i$ rather than $\hat{\beta}$ from the model $y_i = \alpha + \beta x_i + \varepsilon_i$.

Learning statistical programming is inherent to practicing applied econometrics. Consequently, throughout this course we will also teach the statistical programming language R.

Objectives

1. **Distinguish** between settings that require **causal inference** *vs.* settings that want **prediction**.
2. Understand the main **themes and best practices** in modern **prediction** methods.
3. Develop **familiarity** with common machine-learning algorithms—and their strengths/weaknesses.
4. Build **intuition** for prediction—especially the bias-variance tradeoff.
5. Expand **R expertise**.

Prerequisites This course requires the previous course in our sequence—*i.e.*, Economics 423/523. I also assume you are comfortable in R.

Books

I know you are busy and reading for class is often difficult. However, **if you are actually here to learn, then read these books**.

Note Each book (except one of the recommended books) is available for **free online**. The physical copies are also very reasonably priced—I suggest you buy physical versions for books that you like.

Required books

1. Introduction to Statistical Learning *ISL*
2. The Hundred-Page Machine Learning Book *100ML*
3. Data Visualization *Data Viz*

Suggested books

1. R for Data Science
2. Introduction to Data Science (not available without purchase)
3. The Elements of Statistical Learning (*ESL*, the big brother of *ISL*)

Software and tools

- We will use the statistical programming language **R**.
- We will use **`RStudio`** to interact with R.

Learning R will require time and effort, but it is a powerful and versatile tool that is valued by many employers. Put in the requisite effort and time, and you will be rewarded.

Labs, assignments, projects, and exams

Attend the lab This course includes a lab, which is **integral to learning** the material in (and passing) this course. The lab includes both general econometrics instruction and computing resources necessary to complete the course and learn/master its topics.

Assignments

- You will submit **typed assignments via Canvas**.
- Assignments will typically be due every Tuesday.
- We will grade on a **complete/incomplete scale**. Low-quality work will be returned to be resubmitted as late.

Late submissions Students whose assignments are occasionally late will be penalized half a letter grade. Students whose assignments are frequently late will be penalized a full letter grade.

Group work Feel free to work together on the assignments. Unless explicitly stated, each student is required to write and submit independent answer sheets. This means that word-for-word copies will not be accepted and will be viewed as academic dishonesty. If you work with other students, you must list the students in your study group at the top of your assignment. If you fail to do so, you will receive a score of zero.

Project We will have one major project. Details coming.

Exams

- We will proctor an **in-class final on March 16, 2020 at 8:00am** (likely in 105 Peterson Hall).
- A **take-home final exam will be due March 16, 2020 by 11:59pm**.

If you will be out of town for the exams, you must take the exam at a testing center at a university in whichever town you will be visiting at the same time as the EC524/424 scheduled exam (or take a zero). For in-class exams, you may not wear hats, sunglasses, or hoods.

Recommendations

1. **Take responsibility** for your own education and try to **learn** as much as you can.
2. **Do your own work**.
3. Develop your **intuition**—*e.g.*, why would method x work in one situation and fail in another?
4. **Learn R**. Struggle while you try—and use **Google** to figure things out.
5. Come to **office hours**.[2]

[2] Two related articles from NPR on office hours: *College Students: How to Make Office Hours Less Scary* and *Uncovering A Huge Mystery Of College: Office Hours*.

Honesty and academic integrity

You must do your own work. Do not claim credit for any work other than your own. Cheating or plagiarizing of any sort on any component of this class will result in a failing grade for the term and a report of the offense to the university. Please acquaint yourself with the Student Conduct Code.

Accessibility

If you have a documented disability and anticipate needing accommodations in this course, please make arrangements with me during the first week of the term. Please request that the Accessible Education Center send me a letter verifying your disability.

Grading

Grades will be assigned as follows.[3]

Grade	Assignments	Project	Final exam
A	*Incomplete* on $\leq$ 1 assignment.	$\geq$ *Professional*	$\geq$ 80%
B	*Incomplete* on $\leq$ 2 assignments.	$\geq$ *Minor revision*	$\geq$ 70%
C	*Incomplete* on $\leq$ 3 assignments.	$\geq$ *Moderate revision*	$\geq$ 60%
D	*Incomplete* on $\leq$ 4 assignments.	$\geq$ *Major revision*	$\geq$ 50%

Recall that assignments are graded as *Complete vs. Incomplete*—the standard for *Complete* is much higher than simply submitting.

[3] Undergraduates are allowed to miss one additional assignment in the scheme.

Tentative, predicted outline

0. An introduction to prediction and statistical learning

1. What are we doing? **Readings** *ISL* Introduction, Ch1
2. Prediction *vs.* causal inference **Readings** *Prediction Policy Problems* by Kleinberg *et al.* (2015)
3. Modeling decisions and assessment **Readings** *ISL* Ch3

1. Exploratory data analysis

1. Building insights from graphics **Readings** *Data Viz* Preface, Ch1
2. `ggplot2` **Readings** *Data Viz* Ch3

2. Supervised learning

1. An introduction to machine learning **Readings** *100ML* Preface, Ch1–Ch4; *ISL* 2.1–2.2
2. Resampling methods and other best practices **Readings** *100ML* Ch5; *ISL* Ch5
3. Why don't we stick with regression? **Readings** *ISL* Ch3–Ch4
4. LASSO and Ridge regression **Readings** *ISL* 6.1–6.3, 6.6
5. Classification trees **Readings** *100ML* 3.3; *ISL* 8.1
6. Regression trees **Readings** *100ML* 3.3; *ISL* 8.1
7. SVM **Readings** *100ML* 3.4; *ISL* 9.1–9.4
8. Neural nets **Readings** *100ML* 6
9. Boosting and ensembles **Readings** *100ML* 7.5 and Ch8
10. Random forests **Readings** *ISL*
11. Additional topics **Readings** *100ML* Ch7 anc Ch11

3. Unsupervised learning

1. Introduction to unsupervised learning **Readings** *100ML* Ch9; *ISL* 10.1
2. Principal components analysis **Readings** *ISL* 10.2; *100ML* 9.3
3. Nearest-neighbor matching, K-means, and hierarchical clustering **Readings** *100ML* Ch9; *ISL* 10.3

4. Extensions

1. Bias and fairness **Readings** Hao (2019)