Syllabus: Machine Learning and Big Data in Economics

Course & Instructor Information

Course

Course Title, Prefix, Number, Section: Experimental Course: Machine Learning and Big Data in Economics, ECON 40970-045

Semester and Year: Spring 2025

Number of Credits: 3

Course Component Type: LEC

Class Location: SAD 217

Class Meeting Day(s) & Time(s): TR 12:30-1:50 PM

Instructor

Instructor Name: Stepan Gordeev

Office Location: SCHAR 4108

Office Hours: M 4:30-5:30 PM, R 3:30-4:30 PM

Preferred Method of Contact: Email

Email: s.gordeev@tcu.edu

Response Time: 24 hours on weekdays

Final Evaluative Exercise

Final Portfolio Due May 6

Note for students: The syllabus is your first course reading. It provides an orientation to, overview of the flow, and expectations of the course. You should turn to the syllabus for details on assignments and course policies.

Student Resources & Policy Information

Click or scan QR code for resources to support you as a TCU student. Please note section on Student Access and Accommodation and Academic Conduct & Course Materials Policies.

TABLE OF CONTENTS

COURSE DESCRIPTION

Catalog Description

Section Description

Econometric and machine learning methods for dealing with big data and non-standard data. Introduction to programming in R, linear regression, variable selection, random forests, unsupervised learning, deep learning. The quantity and complexity of data available to economists grows every year. This course will introduce you to the methods of analyzing high-dimensional data that conventional statistical and econometric techniques are not well-suited for. The methods we will cover include linear regression, polynomial regression, resampling, variable selection, trees, random forests, and quantifying non-numerical data. In the process of learning these methods, you will also become proficient in R: a popular programming language used for statistical and econometric analysis.

Prerequisites

MATH 10043 or MATH 30853 or INSC 20153.

COURSE MATERIALS

Recommended Materials

James et al., *An Introduction to Statistical Learning with Applications in R*. ISBN 978-1071614174. Abbreviated as ISL below. You can download a free digital copy from the authors' website: https://www.statlearning.com/

LEARNING OUTCOMES

Course Learning Outcomes

At the conclusion of the course, students will be able to:

i) choose an appropriate econometric or machine learning method for analyzing the data at hand;
ii) implement the method in R;
iii) estimate the economic quantities of interest using the method;
iv) assess the successes and the limitations of the chosen method;
v) describe and interpret the analysis in a compelling way.

COURSE REQUIREMENTS

Assignments

Instructions for assignments will be distributed as .qmd files (Quarto notebooks). Each assignment below is to be submitted as a .qmd file with your work and a "rendered" .html output to the Box folder that will be shared with you, in the sub-folder corresponding to the assignment. Make sure that each .qmd file is fully reproducible (I should be able to run the file and produce the same results). Grades and feedback will be delivered via TCU Online.

In-Class Exercises

In most classes, you will be given an exercise to complete before the end of the same class. You must bring a laptop to class to complete these exercises. You will work on these in pairs, but each student must make their own submission. You cannot use external resources (other than official documentation) or AI assistance for these exercises. Two lowest scores will be dropped. Course learning outcomes covered: ii-iii).

Assignments

You will be given five bigger assignments to complete outside of class. These will build and expand on in-class exercises. These assignments must be completed individually. One lowest score will be dropped. Course learning outcomes covered: i-iv).

Milestone Projects

The assignments will be grouped into three modules devoted to particular sets of methods. Assignments within a single module will build on top of one another, guiding you through completing a significant big data project. You will need to compile your assignment submissions belonging to a single module into a single project, expanding your analysis when requested and ensuring that the shortcomings of individual assignment submissions have been addressed. Each module's project will represent the milestone associated with that module. Course learning outcomes covered: v).

Final Portfolio

Finally, you will assemble the three projects into a single professional portfolio showcasing your proficiency in econometric and machine learning methods for working with big data. It will be due on May 6. Course learning outcomes covered: v).

Grading Policy

Late Work

No late work will be accepted, unless mandated by the Official TCU Absence Policy.

Participation, Engagement & Attendance

There is no explicit grading for attendance and participation. At the same time, if you are not present in class on a particular day, you will not be able to complete the in-class exercise.

Absences described in the Official University Absence Policy will be excused. Excused Absences or Official University Absences are absences described in the Official University Absence Policy and include the following: Title IX related issues, military leave, holy days, and university related absences.

Verified Absences may be excused following the instructor's discretion. Because it is considered an infringement on student privacy for me to have access to student medical records, I cannot accept medical documentation to justify absences. If you have a legitimate reason for your absence and want to provide verification, please access the Absence Documentation Form here.

Course Assignments & Final Grade

Category	Percentage
In-Class Exercises	20
Assignments	40
Milestone Projects	20
Final Portfolio	20
Total	100

Grading Scale

Grade	Percentage
A	94–100
A-	90–93.99
B+	87–89.99
B	84–86.99
B-	80–83.99
C+	77–79.99
C	74–76.99
C-	70–73.99
D+	67–69.99
D	64–66.99
D-	60–63.99
F	0–59.99

Course Policies

Technology Policies

Please turn off or silence cell phones and put them away during class. You may use laptops for the in-class exercises but *not during the lectures*.

Artificial intelligence (AI) and other unapproved assignment-help tools MAY NOT be used for course assignments except as explicitly authorized by the instructor. Specific examples of prohibited activities include, but are not limited to:

- Submitting all or any part of an assignment statement to an AI or unapproved assignment-help tool;
- Incorporating any part of an AI-generated response in an assignment;
- Using AI to brainstorm, formulate arguments, or template ideas for assignments;
- Using AI to summarize or contextualize source materials;
- Submitting your own work for this class to an AI or unapproved assignment-help tool for iteration or improvement.

If you are in doubt as to what constitutes AI, or whether an assignment-help tool is suitable for use in this class, then it is your responsibility to discuss your situation with the instructor.

COURSE SCHEDULE

The calendar represents a tentative schedule. Plans may need to be changed as the semester unfolds.

Dates	Topic	Assigned Materials	Due
Module 1: Data Analysis, Linear Regression for Conventional Data			
Week 1: Jan 14, 16	Syllabus, Intro, R	Read Syllabus	
Week 2: Jan 21, 23	Basics of Programming in R	ISL 2	
Week 3: Jan 28, 30	Reproducible Research		
Week 4: Feb 4, 6	Data Visualization		
Week 5: Feb 11, 13	Linear Regression	ISL 3	Assignment 1
Week 6: Feb 18, 20	Polynomial, Logistic Regr.	ISL 4.3	
Module 2: Variable Selection, Random Forests for High-Dimensional Data			
Week 7: Feb 25, 27	Resampling	ISL 5	Assignment 2
Week 8: Mar 4, 6	Variable Selection	ISL 6	Project 1
Week 9: Mar 11, 13	Trees	ISL 8.1	
Week 10: Mar 18, 20	Spring Break		
Week 11: Mar 25, 27	Random Forests	ISL 8.2	Assignment 3
Week 12: Apr 1, 3	Principal Components	ISL 12.3	Assignment 4
Week 13: Apr 8, 10	Clustering	ISL 12.4	
Module 3: Deep Learning for Unstructured Data			
Week 14: Apr 15, 17	Deep Learning	ISL 10.1, 10.2	Project 2
Week 15: Apr 22, 24	Text Data		Assignment 5
Week 16: Apr 29	Portfolio Workshop		Project 3
May 6			Final Portfolio