MA-FK-600 / MA-W-210 / MA-W-220:
Angewandte empirische Wirtschaftsforschung/Applied Econometrics

Announcement of the Empirical Master Research Seminar

"Topics in Machine Learning and Econometrics"

Dr. Sylvi Rzepka,

www.uni-potsdam.de/empwifo - rzepka@empwifo.uni-potsdam.de

University of Potsdam, Winter term 2020/21

General information:

This seminar has 2 learning goals: 1) introduce the main concepts of machine learning (ML), especially supervised learning, and 2) identify how ML can be used in applied economic research and even enhance causal inference. We will not only discuss text book chapters and empirical economics papers, but also learn how to implement ML methods in R. Students will have the chance to get to know R through online courses provided by "Datacamp for the classroom". During the semester students will present a ML method, an application of ML in economics, and complete problemsets in R. The final assignment will be in the spirit of a competition (with bonus points for the best performing algorithm). Besides technical skills the course promotes a series of important softskills: public speaking, presenting empirical results intuitively, and navigating self-directed learning the field of machine learning.

Registration:

If interested in participating, please register by sending an email to *rzepka@empwifo.uni-potsdam.de* by **November 6, 2020, 12:00** stating your name, matriculation number and a list of courses you have taken at the Chair of Empirical Economics including the grade you achieved. We recommend successful completion of the courses "MA: Microeconometrics" as well as "MA: Public Policy Evaluation". The number of participants in the seminar is limited.

Schedule:

Note: Tentative schedule. Final schedule depends on enrollment. Sessions start at the exact time as listed.

6.11.20 ,14:00 - 15:30	Kick-off meeting
13.11.20, 14:00 - 17:00	Introduction to Machine Learning
15.11.20, midnight	Official registration via e-mail with *Application* preference
16.11.20	Assignment of topics via e-mail
27.11.20, 9:00 - 11:00	Presentation of *ML method block 1*
4.12.20, midnight	Problemset 1 due
11.12.20, 9:00 - 11:00	Presentation of *Application block 1*
18.12.20, midnight	Problemset 2 due
8.1.21, 14:00 - tbd	Presentation of *ML method block 2*
15.1.21, midnight	Problemset 3 due
22.1.21, 14:00 - tbd	Presentation of *Application block 2*
29.1.21, midnight	Problemset 4 due
5.02.21, 14:00-15:30	Wrap-up and final discussion & Empirical assignment sent out
10.2.21, midnight	Empirical assignment paper due (**new date!**)

What we expect from participants:

1. Actively participating in **all** sessions and complying with all deadlines listed in the schedule.
2. Complete the reading assignments for the "Introduction to Machine Learning" Sessions.
3. Present one machine learning method and one application of machine learning to economics.
4. Complete 4 problemsets.
5. Complete the final empirical assignment.
6. Your grade will be determined by how well you do in your presentations, in participating in the discussion, in the problemsets, and in the final empirical assignment.

Introduction to Machine Learning Sessions: Reading Assignments

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013): An Introduction to Statistical Learning with Applications in R New York: Springer. Available online: pdf-version of textbook Chapter 1 and 2 , pp. 1-42. From here on referred to as "ISLR".

Mullainathan and Spiess (2017): Machine Learning: An Applied Econometric Approach, *Journal of Economic Perspectives*, 31(2)87-106.

Machine Learning Methods:

1. Lasso and Ridge Regression, ISLR, Chapter 6.2, p. 214-228.

2. Tree-based methods, ISLR, Chapter 8, p. 303-336.

3. Support Vector Machines ISLR Chapter 9, p. 337-359.

Applications (choose one):

1. Random forests
Björkegren, D. and Grissen, D. (2019). Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment *World Bank Economic Review*:0(0):1-17.

2. Random forests
Glaeser, E. L, Kim, H., and Luca, M. (2019). Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity in Abraham, K.G., Jarmin, R.S., Moyer, B., and Shapiro, M.D. *Big Data for 21st Century Economic Statistics*.

3. Gradient Boosted Decision Trees
Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., Mullainathan, S. (2018). Human Decisions and Machine Predictions *The Quarterly Journal of Economics*, 133 (1), 237-293.

4. Support Vector Machines
Gründler, K. and Krieger, T. (2016). Democracy and growth: Evidence from a machine learning indicator, *European Journal of Political Economy*, Volume 45 (Supplement), 85-107.

5. Support Vector Machines
Gründler, K. and Krieger, T. (2019). Should We Care (More) About Data Aggregation? Evidence from Democracy Indices., *CESifo Working Paper*, 7480.

6. Lasso for Model Selection
Wu, Alice H. 2018. Gendered Language on the Economics Job Market Rumors Forum. AEA Papers and Proceedings, 108 : 175-79.

7. Lasso for Model Selection
Briel and Töpfer (2020): The Gender Pay Gap Revisited: Does Machine Learning offer New Insights? LASER Discussion Papers - Paper No. 116. `http://www.laser.uni-erlangen.de/papers/paper/340.pdf`

8. Lasso for Model and Instrumental Variable Selection
Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects *Journal of Economic Perspectives*, 28(2)29-50.

9. Lasso for Model and Instrumental Variable Selection
Angrist,Joshua and Frandsen, Brigham (2020) Machine Labor, NBER Working Paper No. 26584.

10. Causal Random Forests
Davis, J. M. V. and Heller, S. B. (2017). Using causal forests to predict treatment heterogeneity: An application to summer jobs. American Economic Review, 107(5), 546-50. [SHORT VERSION]
Davis, J. M. V. and Heller, S. B. (forthcoming). Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs *Review of Economics and Statistics*.

Presentation Guidelines
Expected structure of the presentation

Setting 25 minutes + discussion time in which you explain the machine learning application in your assigned paper. Teach as you would like to be taught.

Introduction What method is discussed? Formulate 3 takeaways for this method linked to the application you present.

Application Bulk of the talk! What is the main research question of the paper? Why and how is this ML-method applied? How were "off-the-shelf" ML-methods adapted in this paper? Really zoom in on the ML method. Give details about this methodological aspect of the paper and illustrate it as intuitively as possible using tables, graphs, etc. Since we focus on ML methods in this seminar, results and other methods used in the paper are also interesting, but more as a side note if needed for the flow of your talk.

Conclusion Close with a short discussion. Possible aspects to discuss: Compare and contrast the method to an econometric method you already know. What is the advantage of the machine learning method over the traditional econometric one? What are potential disadvantages? When would you prefer which technique? Refer back to the 3 takeaways you mentioned in the beginning.

Additional advice

- Be precise, avoid jargon, use your own words instead.

- Just like any teacher: Be prepared for comments and questions.

- Make sure the content of your slides is readable (i.e. use large and simple fonts).

- Highlight the intuition behind the ML-technique and link to the key ML concepts (covered in the introduction session) which play a role in your method and particular application.

- Discuss how the authors empirically implement the method to answer a question. Why did they use machine learning and not traditional econometrics?

- There are plenty of additional resources on machine learning online. However, since most of them are from a computer science perspective, you have to be critical when thinking about applying these techniques in the context of economic research.

- Sign up for an office hour if you have specific questions concerning your talk.