

**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 & Markus Müller,
www.uni-potsdam.de/empwifo - seminar@uni-potsdam.de
University of Potsdam, Winter term 2019/20

General information:

This seminar provides a broad overview of the main concepts of machine learning, especially supervised learning, and how they can enhance causal inference. We will not only discuss recent empirical economics papers applying machine learning methods, but also explore how to implement these methods in R. Students will have the chance to get to know R in a Workshop organized by PCQR and/ or through online courses provided by "Datacamp for the classroom".

During the semester students will present one empirical application and complete two problemsets. The final assignment will be in the spirit of a Machine Learning Challenge. Throughout the course, students have the chance to practice public speaking and presenting empirical results intuitively as well as getting hands-on experience in R. Furthermore, this course will enable students to follow-up on new developments in this quickly evolving field on their own.

Registration:

If interested in participating, please register by sending an email to seminar@empwifo.uni-potsdam.de by **October 15, 2019** 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: All sessions start at the exact time as listed (s.t.).

- 15.10.19 Pre-registration for kick-off meeting
- 16.10.19, 10:00-11:30 Kick-off meeting (room: Haus 7, 2.10)
- 23.10.19, midnight Official registration via e-mail with 3 ordered preferences of topics
- 25.10.19 Assignment of topics via e-mail
- 25.10.19, 14:00-17:30 Introduction to Machine Learning (room: Haus 7, 2.10)
- 7./8.11.19 Introduction to R Workshop (PCQR)
- 29.11.19, midnight Block 1: Send your draft slides and office hour sign-up
- 13.12.19, 14:00 - tbd Block 1: Student Presentations (room: Haus 7, 2.10)
- 20.12.19, 14:00 Problemset 1 due
- 03.01.20 midnight Block 2: Send your draft slides and office hour sign-up
- 17.01.20, 14:00 - tbd Block 2: Student Presentations (room: Haus 7, 2.10)
- 24.01.19, 14:00 Problemset 2 due
- 31.01.20, 14:00-15:30 Wrap-up and final discussion (room: Haus 7, 2.10)
- 31.01.20 Empirical assignment sent out
- 7.2.20, midnight Empirical assignment paper due

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 empirical application.
4. Complete two empirical problemsets.
5. Complete the final empirical assignment.
6. Your grade will be determined by how well you do in your presentation, 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.

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

Proposed topics (choose one application):

1. Random forests

Björkegren, D. and Grissen, D. (forthcoming). Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment *World Bank Economic Review*.

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.

Discussion of Methods for the first three applications

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013): Chapter 8, p. 303-336.

4. Lasso for model selection

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

5. Lasso for 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.

Discussion of Methods for topic 4 & 5

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013): Chapter 6.2, p. 214-228.

6. Causal random forests

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* forthcoming.

Davis, J.M., & 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]

7. Causal random forests

Athey, S., & Wager, S. (2019). Estimating Treatment Effects with Causal Forests: An Application. arXiv preprint arXiv:1902.07409.

Discussion of Methods for topic 6 & 7

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353-7360.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228-1242.

Presentation Guidelines

Expected structure of the presentation

Setting 20 minutes + discussion time in which you are the teacher! Teach as you would like to be taught.

Introduction What method is discussed? Formulate 3 takeaways for this method and/ or the empirical example you present.

Method Explain the method in your own words referring to the textbook or other general references. How does this ML-method work? What are advantages, what are disadvantages? Create your own sketches or rely on graphs / figures, etc. from James et al. 2013 or other books (state the reference!) to explain the method.

Application How is this ML-method applied in your assigned paper? Illustrate this as intuitively as possible using tables, graphs, etc. How did the actual implementation take place?

Discussion Compare and contrast the method to an econometric method you already know. What is the advantage of the machine learning method over the econometric one? When would you prefer the econometric technique?

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.