Dr. Arne Jonas Warnke
(BASF SE)

Topics in Data Analytics and Machine Learning

Comments

The seminar takes place on June 28-29, 2019, in room 00.010. There will be additional Q&A sessions on May 17, 2019, from 14.00 to 16.00 and on June 7, 2019, from 16.00 to 18.00 (each in room 00.010). A maximum of 15 students can participate in the seminar.

Description

Data science, machine learning and artificial intelligence are leading new technologies which will have a profound impact on our economy and society. In this seminar, we will shed light on different aspects of these technologies and show why social scientist should be aware of them.

We will consider both theoretical

- What is overfitting and how to prevent it via regularization?
- Ways to handle non-linear relationships.
    - E.g. how do tree-based methods work and what are there advantages and disadvantages?
- How does a neural network work?
- What are hyperparameter and how to tune them?

as well as applied topics

- How to write a reproducible code?
- Challenges of implementing a productive analytical solution:
    - What can we learn from Google Flu
- Python vs R for data science: What are the differences?
- Why have (deep) neural networks become so popular recently?
- Advantages and limitations of explainable models
- An introduction into fairness in machine learning
- How did AlphaGo beat professional human Go-player?

Interested students may also opt to work on a forecasting challenge:

- Whose model can predict the level of the Rhine river most accurately?

The first meeting will take place on April 26, 2019, from 14.00 to 16.00 in room 00.010. Further information is available via email. To register for the seminar, please indicate a topic you are interested in (you may also suggest new topics) via email to arne.warnke@gmail.com

The deadline for registration is April 22, 2019.

Content (among others)

A Neural Network Playground: https://playground.tensorflow.org/

Athey, Susan. "The impact of machine learning on economics." In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, 2018.

Athey, Susan, and Guido Imbens. *Lectures on Machine Learning*. NBER Summer Institute Econometric Lectures, 2015.

Bengio, Yoshua, Ian J. Goodfellow, and Aaron Courville. *Deep Learning*. MIT Press (2016). Available at https://www.deeplearningbook.org/

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Vol. 1, no. 10. New York, NY, USA:: Springer series in statistics, 2001. Available at https://web.stanford.edu/~hastie/ElemStatLearn/download.html

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Vol. 112. New York: Springer, 2013. http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf

Koenker, Roger, and Achim Zeileis. "On reproducible econometric research." *Journal of Applied Econometrics* 24, no. 5 (2009): 833-847.

Kuhn, Max, and Kjell Johnson. *Applied predictive modeling*. Vol. 26. New York: Springer, 2013.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. *The parable of Google Flu: traps in big data analysis. Science* 343, no. 6176 (2014): 1203-1205.

Mullainathan, Sendhil, and Jann Spiess. "Machine learning: an applied econometric approach." *Journal of Economic Perspectives* 31, no. 2 (2017): 87-106.

Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser et al. *Mastering the game of Go with deep neural networks and tree search*. *Nature* 529, no. 7587 (2016): 484.

Varian, Hal R. "Big data: New tricks for econometrics." *Journal of Economic Perspectives* 28, no. 2 (2014): 3-28.

Certificates

Your grade will be based on the following: seminar paper (60%), seminar presentation (30%) and active participation (10%) in the seminar discussion.