MARUM / GLOMAR Basic Knowledge Course

An Introduction to Data Science and Machine Learning

Hendrik Heuer

Institute for Information Management Bremen (ifib), University of Bremen
hheuer@uni-bremen.de, hen-drik.de

5 June 2018

Objectives

From medical decision support systems to automatic language translation, from sorting and prioritizing news on social networks to autonomous cars: Machine learning is woven into the fabric of daily life. Applying machine learning, data science aims to extract knowledge or insights from data. This application of advanced visualization and machine learning techniques can empower scientists in myriad ways. One important domain where visualization and machine learning support scientists in their research is sociology. Lazer et al. (2009)* characterized this so-called *computational social science* as a field that leverages the capacity to collect and analyse data at a scale to examine patterns of individual and group behaviours and to enhance our understanding of individuals and collectives.
In this course, we will explore how the marine environmental sciences can leverage this.

*D. Lazer et al., "Computational Social Science," Science, vol. 323, no. 5915, pp. 721–723, 2009

Content

The course will provide an introduction to data science and applied machine learning. For this, the programming language Python will be used (and taught). You will learn about the difference between supervised and unsupervised machine learning, and apply your knowledge on the three machine learning tasks: 1. classification (e.g. k-NN, Decision Trees, Support Vector Machines),
2. regression (Linear Regression, Logistic Regression), and
3. clustering (k-means, dimensionality reduction with PCA and t-SNE).
We will also acquire a basic understanding of machine learning subfields like natural language processing for text mining, computer vision, and deep learning. Evaluation, as an integral part of data science, will be taught as well as data processing, feature and model selection, and data mining.

About the lecturer

Hendrik Heuer is a Researcher at the University of Bremen associated with the *Institute for Information Management (ifib)*. The focus of his doctoral research is on the user experience of artificial intelligence. He investigates trust in machine learning from a human-computer interaction perspective. Hendrik studied digital media, human-computer interaction, and machine learning in Bremen, Buffalo, Stockholm, Helsinki, and Amsterdam. His present research group is currently building the first information system that gives a holistic picture of the global welfare state: a next-generation atlas with sophisticated visualization and machine learning capabilities.

Location and Time

MARUM, University of Bremen, Leobener Str. 8, 28359 Bremen, Germany
MARUM I (main) building, room 2070

09.00 – 17.00 hrs.

Software

Practical exercises during the course will be done with Python. WE will provide laptops for each participant with the software installed on them.

Participants who would like to bring their own laptops may do so but are responsible to make sure that the software is running properly by the beginning of the course.
Please note that during the course, support can only be given for laptops provided by MARUM.

Python can be downloaded from https://www.anaconda.com/download/

Registration

To register for this course, please visit the course webpage.

Please note that your registration will be binding.

The registration deadline for this course is **30 May 2018**.

Any enquiries regarding the course should be addressed to early-career@marum.de.