Jülich Supercomputing Centre
Forschungszentrum Jülich GmbH

Training course with hands-on sessions

Topic: **Parallel and Scalable Machine Learning**
(Training Course No. 1222019)

Speakers: Prof. Morris Riedel, Dr. Gabriele Cavallaro, JSC

Contents: The course offers basics of analyzing data with machine learning and data mining algorithms in order to understand foundations of learning from large quantities of data. This course is especially oriented towards beginners that have no previous knowledge of machine learning techniques. The course consists of general methods for data analysis in order to understand clustering, classification, and regression. This includes a thorough discussion of test datasets, training datasets, and validation datasets required to learn from data with a high accuracy. Easy application examples will foster the theoretical course elements that also will illustrate problems like overfitting followed by mechanisms such as validation and regularization that prevent such problems.

The tutorial will start from a very simple application example in order to teach foundations like the role of features in data, linear separability, or decision boundaries for machine learning models. In particular this course will point to key challenges in analyzing large quantities of data sets (aka big data) in order to motivate the use of parallel and scalable machine learning algorithms that will be used in the course. The course targets specific challenges in analyzing large quantities of datasets that cannot be analyzed with traditional serial methods provided by tools such as R, SAS, or Matlab. This includes several challenges as part of the machine learning algorithms, the distribution of data, or the process of performing validation. The course will introduce selected solutions to overcome these challenges using parallel and scalable computing techniques based on the Message Passing Interface (MPI) and OpenMP that run on massively parallel High Performance Computing (HPC) platforms. The course ends with a more recent machine learning method known as deep learning that emerged as a promising disruptive approach, allowing knowledge discovery from large datasets in an unprecedented effectiveness and efficiency.

The course will be given in English. This course is a PRACE training course.

Time: 25-27 February 2019, 9:00-16:30

Venue: Jülich Supercomputing Centre, Rotunda (building 16.4, room 301)

Registration: Please register via the form at the PRACE web site until 22 January 2019:
https://events.prace-ri.eu/event/817/registration/register#/register

Anyone interested is cordially invited to participate in this event.

sgd Dr. Sabine Höfler-Thierfeldt