Homework 1

Large-Scale Data Analysis

(team submission)

Fabian Gieseke & Christian Igel

Department of Computer Science
University of Copenhagen

23 April 2019

Submission Deadline:
11 June, 23:59

This assignment is a **team assignment**, i.e., you are supposed to work jointly on the task within your team! Each team has to hand in one report and has to jointly present the overall approach. If you cannot find a team, please get in touch with us until Thursday, April 25 (e.g., during the break of the lecture)

1 Task & Instructions (100 pts)

The goal is to approach a real-world problem and to come up with a competitive machine learning approach. Per team, please proceed as follows:

1. Find a team (3-4 students per team). Enter one of the available group slots on Absalon ("People → Competition (HW1)")

2. Decide on a public Kaggle team name with "$LSDA\_$" as prefix (e.g., "LSDA_CoolTeam").

3. Create Kaggle accounts and a Kaggle team. If no member is willing to join Kaggle, please get in touch with us as soon as possible!

4. Start working on the competition! You are allowed to use any programming language and any tools/computers to approach the competition. Keep in mind the competition rules (`https://www.kaggle.com/c/bigdata2019-flare-prediction/rules`)

5. You can make a few "submissions" on Kaggle per day. We highly recommend that you make your first submission as soon as possible. Note that your final Kaggle score will be taken into account for the evaluation (comparison among all LSDA teams)!

Deliverables:

- Absalon: You have to deliver *two separate files* via Absalon: A pdf file `report.pdf` containing your report (one report per team, max. 4 pages) and a `code.zip` file containing the associated code as zip file (do *not* include the competition dataset). Please point out the individual contributions per team member in the report. Make use of this Latex template: `https://www.ieee.org/content/dam/ieee-org/ieee/web/org/conferences/Conference-LaTeX-template_7-9-18.zip`.

- Presentation on June 13, 9:15–11:00: Each team also has to present its approach in class via a "flash talk" (2 minutes). Note that each team member has to actively participate in the flash talk.

More details regarding the report and the flash talks will be announced on Absalon.