

KEN4258: Computational Statistics

Homework Assignment 1 (graded)

Christof Seiler

Posted on February 22 at 20:00 / Due on February 29 at 20:00

Logistics

The final submission in the form of a report should be one single pdf that you can upload to Canvas. One submission per group. The group can have a maximum of 5 members. Keep the report short. The main text should not exceed three pages. You can add an appendix to mention details that you find important. Otherwise, the structure of the report is up to you. You might want to add a link to your GitHub repository with your code (in R or Python).

Task: R^2 [100 points]

R^2 has a problem with datasets that have a large number of predictors with respect to the number of observations. The purpose of this graded homework assignment is to learn about this problem and provide a solution for it. To do that, follow these steps:

- 1) Create a Monte Carlo simulation to illustrate the problem. [25 points]
- 2) Provide a mathematical proof showing that the problem really exists. [25 points]
- 3) Propose a solution to address the problem. [25 points]
- 4) Find a real dataset to illustrate the problem and your fix. [25 points]