732A38 Computational statistics
Division of Statistics and Machine Learning
Department of Computer and Information Science

Computer lab 2

Instructions

- Create a report to the lab solutions in PDF.
- Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.
- **Include all your codes as an appendix into your report.**
- A typical lab report should 2-4 pages of text plus some amount of figures plus appendix with codes.
- The lab report should be submitted via LISAM before the deadline.

Assignment 1: Optimizing a model parameter

File mortality_rate.csv contains information about mortaily rates of the fruit flies during a certain period.

1. Import this file to R and add one more variable LMR to the data which is the natural logarithm of Rate. Afterwards, divide the data into training and test sets by using the following code:

```
n=dim(data)[1]
set.seed(123456)
id=sample(1:n, floor(n*0.5))
train=data[id,]
test=data[-id,]
```

2. Write your own function **myMSE** that for given parameters *lambda* and list *pars* containing vectors *X, Y, Xtest, Ytest* fits a LOESS model with response Y and predictor X using loess() function with penalty *lambda* (parameter enp.target in loess()) and then predicts the model for Xtest. The function should compute the predictive MSE, print it and return as a result.
3. Use a simple approach: use function **myMSE**, training and test sets and the following lambda values to estimate the predictive MSE values:
    - Lambda=0.1,0.2,…,40

   Create a plot of the MSE values versus lambda and comment on which lambda value is optimal. How many evaluations of myMSE were required to find this value?
4. Use optimize() function for the same purpose, specify range for search [0.1, 40] and the accuracy 0.01. Have the function managed to find the optimal MSE value? How many myMSE function evaluations were required? Compare to step 3.

5. Use optim() function and BFGS method with starting point lambda=35 to find the optimal lambda value. How many myMSE function evaluations were required? Compare the results you obtained with the results from step 4 and make conclusions.

Assignment 2: Maximizing likelihood

File *data.RData* contains a sample from normal distribution with some parameters μ, σ.

1. Load the data to R environment.
2. Write down the log-likelihood function for 100 observations and derive maximum likelihood estimators for μ, σ analytically by setting partial derivatives to zero. Use formulas derived to obtain parameter estimates for the data loaded.
3. Now you are assumed to derive maximum likelihood estimates numerically. Why it is a bad idea to maximize likelihood rather than maximizing log-likelihood?
4. Optimize the minus log-likelihood function with initial parameters μ=0, σ=1. Try both Conjugate Gradient method and BFGS algorithm with gradient specified and without.
5. Did algorithms converge in all cases? What were the optimal values of parameters and how many function and gradient evaluations it required for algorithms to converge? Which settings would you recommend?

Submission procedure

Assume that X is the current lab number.

If you are neither speaker nor opponent for this lab,

- Submit your report using *Lab X* item in the *Submissions* folder before the deadline.
- Make sure that you or some of your group members submits the group report using *Lab X group report* in the *Submissions* folder before the deadline

If you are a speaker for this lab,

- Submit your report using *Lab X* item in the *Submissions* folder before the deadline.
- Make sure that you or some of your group members does the following before the deadline:
    - submits the group report using *Lab X group report* in the *Submissions* folder before the deadline
    - Goes to Study room *Speakers X→Documents* and opens file *Password X.txt*. Then the student should put your group report into ZIP file *Lab X.zip* and protect it with a password you found in Password X.txt
    - Uploads the file to *Collaborative workspace* folder

If you are opponent for this lab,

- Submit your report using *Lab X* item in the *Submissions* folder before the deadline.

- Make sure that you or some of your group members submits the group report using *Lab X group report* in the *Submissions* folder before the deadline
- After the deadline for the lab has passed, go to Collaborative workspace folder and download *Lab X.zip*. Open the PDF in this ZIP file by using the password available in *Course Documents →Password X.txt,* read it carefully and **prepare at least two questions/comments/improvement suggestions** in order to put them at the seminar (i.e. at least two questions per opponent)