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

Computer lab 1

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: Be careful with '=='

A pupil of a school is bad in arithmetic but good in programming. He writes a program to check if 1/3-1/4==1/12:

```
x1<-1/3;
x2<-1/4;
if (x1-x2==1/12){
print("Teacher said true")
} else{
print("Teacher lied")}
```

1. Check the result of this program. Comment why this happened.
2. Specify how the program can be modified to give a correct result

Assignment 2: Derivative

A widely known way to compute the derivative of function *f(x)* in point *x* is to use

$$f'(x) = \frac{f(x+\varepsilon) - f(x)}{\varepsilon}$$

1. Write your own function computing the derivative of function *f(x)=x* in this way. Take $\varepsilon=10^{-15}$
2. Compute your derivative function at point x=100000.

3. What is the value you obtained? What is the real value of the derivative? Explain the reason behind the discovered difference

Assignment 3: Variance

A known formula for estimating variance is

$$Var(x) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right)$$

1. Write your own function *myvar* estimating variance in this way
2. Generate vector x=($x_1 \ldots x_{10000}$) with 10000 random numbers, normally distributed with mean 10^8 and variance 1
3. For each subset X_i= {$x_1 \ldots x_i$}, i=1…10000 compute difference Y_i= myvar(X_i)-var(X_i), where var(X_i) is a standard variance estimation function in R. Plot the dependence Y_i on i. Draw necessary conclusions. How well does your function work? What is the reason behind such behavior?

Assignment 4: Linear Algebra

The Excel file "**tecator.xls**" contains the results of a study aimed to investigate whether a near-infrared absorbance spectrum and the levels of moisture and fat can be used to predict the protein content of samples of meat. For each meat sample the data consists of a 100 channel spectrum of absorbance records and the levels of moisture (water), fat and protein. The absorbance is -log10 of the transmittance measured by the spectrometer. The moisture, fat and protein are determined by analytic chemistry. The worksheet you need to use is "data". It contains data from 215 samples of finely chopped meat. The aim is to fit a linear regression model that could predict protein content

1. Import the data set to R
2. Optimal regression coefficients can be found by solving a system of the type **Aβ=b** where **A=X^{T}X** and **b=X^{T}Y.** Compute **A** and **b** for the given data set.
3. Try to solve **Aβ=b** with default solver *solve()*. What kind of result did you get? How can this result be explained?
4. Check the condition number of the matrix A and consider how it is related to your conclusion in step 3.
5. Scale the data set and repeat steps 2-4. How the result has changed and why?

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

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)