Computational Statistics 732A89 – Spring 2025
Computer Lab 2

Martin Andrae, Bayu Brahmantio, Frank Miller
Department of Computer and Information Science (IDA), Linköpings University

January 28, 2025

This computer laboratory is part of the examination for the Computational Statistics course. Create a group report (which is directly presentable, if you are a presenting group), on the solutions to the lab as a PDF file. Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.
All R code should be included as an appendix to your report.
A typical lab report should contain 2-4 pages of text plus some figures plus an appendix with codes. In the report, refer to all consulted sources and disclose all collaborations.
The report should be handed in via LISAM (or alternatively in case of problems by email) by **23:59 February 4, 2025** at the latest. Notice that there is a deadline for corrections 23:59 08 April 2025 and a final deadline of 23:59 29 April 2025 after which no submissions or corrections will be considered, and you will have to redo the missing labs next year. The seminar for this lab will take place **February 12, 2025**.
The report has to be written in English.

Question 1: Optimization of a two-dimensional function

Consider the two-dimensional function

$$f(x, y) = \sin(x + y) + (x - y)^2 - 1.5x + 2.5y + 1$$

for $x \in [-1.5, 4]$ and $y \in [-3, 4]$. The aim is to search for the position (x^*, y^*) of the global **minimum** of f in the given range for x and y.

a. Make a contour plot for this function.

b. Derive the gradient and Hessian matrix for the function and write `R`-code for them.

c. Write your own function applying the Newton algorithm that has the starting value (x_0, y_0) as a parameter.

d. Test several starting values such that you have examples which give at least three different results.

e. Investigate the candidate results which you have got. Is one of the candidates a global minimum? Which type of solutions are the other candidates?

Question 2: Maximum likelihood

Three doses (0.1, 0.3, and 0.9 g) of a drug and placebo (0 g) are tested in a study. Afterward, a dose-dependent event is recorded. The data of $n = 10$ subjects is shown in Table 1; x_i is the dose in gram; $y_i = 1$ if the event occurred, $y_i = 0$ otherwise.

x_i in g	0	0	0	0.1	0.1	0.3	0.3	0.9	0.9	0.9
y_i	0	0	1	0	1	1	1	0	1	1

Table 1: Data for Question 2

You should fit a simple logistic regression

$$p(x) = P(Y = 1|x) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x)}$$

to the data, i.e. estimate β_0 and β_1. One can show that the log likelihood is

$$g(\mathbf{b}) = \sum_{i=1}^{n} \left[y_i \log\{(1 + \exp(-\beta_0 - \beta_1 x_i))^{-1}\} + (1 - y_i) \log\{1 - (1 + \exp(-\beta_0 - \beta_1 x_i))^{-1}\} \right]$$

where $\mathbf{b} = (\beta_0, \beta_1)^T$ and the gradient is

$$\mathbf{g}'(\mathbf{b}) = \sum_{i=1}^{n} \left\{ y_i - \frac{1}{1 + \exp(-\beta_0 - \beta_1 x_i)} \right\} \begin{pmatrix} 1 \\ x_i \end{pmatrix}.$$

a. Write a function for an ML-estimator for (β_0, β_1) using the steepest ascent method with a step-size-reducing line search (back-tracking). For this, you can use and modify the code for the steepest ascent example of the lecture. The function should count the number of function and gradient evaluations.

b. Compute the ML-estimator with the function from a. for the data (x_i, y_i) above. Use a stopping criterion such that you can trust five digits of both parameter estimates for β_0 and β_1. Use the starting value $(\beta_0, \beta_1) = (-0.2, 1)$. The exact way to use backtracking can be varied. Try two variants and compare the number of function and gradient evaluations performed to convergence.

c. Use now the function `optim` with both the BFGS and the Nelder-Mead algorithm. Do you obtain the same results as for b.? Is there any difference in the precision of the result? Compare the number of function and gradient evaluations that are given in the standard output of `optim`.

d. Use the function `glm` in `R` to obtain an ML-solution and compare it with your results before.