

Optimisation Algorithms in Statistics I – Autumn 2020

Assignment 3

Frank Miller, frank.miller@stat.su.se,
Department of Statistics, Stockholm University

October 23, 2020

Perform the solutions individually and send your report **until November 5** to me. Try to keep this deadline. However, if you have problems with it, there will be a final deadline on January 25 for all assignments. Please include your name in the filename(s) of your solution file(s).

Problem 3.1

Let $g(x_1, x_2)$ be the bimodal normal mixture density function shown in Lecture 1. On the homepage, you can find file `bimodal.r` where the function g is defined and a contour-plot in $[-1, 3] \times [-1, 2.5]$ is drawn.

Start with following starting triangle: $(-0.5, 1), (0, 1), (-0.5, 0)$. Compute the first four iterations of the Nelder-Mead algorithm. For this, follow the algorithm's description in Givens and Hoeting (2013) and use the default parameters $(\alpha_r, \alpha_c, \alpha_e, \alpha_s) = (1, 2, \frac{1}{2}, \frac{1}{2})$. You can do this half-manually: you can evaluate g with R, but you need not to program the complete Nelder-Mead algorithm. Draw the computed triangles in the contour plot.

Problem 3.2

Let $g(\mathbf{x}), \mathbf{x} \in \mathbb{R}^3$ the multivariate normal mixture with

$$g(\mathbf{x}) = \sum_{i=1}^4 w_i f_i(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),$$

where $f_i(\cdot; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is the density of the multivariate normal distribution with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$.

We use here mixing weights $w_i = 1/4, i = 1, \dots, 4$, mean vectors

$$\boldsymbol{\mu}_1 = (0, 0, 0)^\top, \quad \boldsymbol{\mu}_2 = (2, 2, 0)^\top, \quad \boldsymbol{\mu}_3 = (2, 0, 2)^\top, \quad \boldsymbol{\mu}_4 = (0, 2, 2)^\top, *$$

and the covariance matrices $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \mathbf{I}$ are the identity matrix and $\boldsymbol{\Sigma}_4 = 0.98 \cdot \mathbf{I}$ (it is not difficult to program the density of the multivariate normal, especially here when covariance matrices are simple; alternatively a package like `mvtnorm` can be used).

*In an earlier version of this assignment, I wrote by mistake $(1, 1, 0), (1, 0, 1), (0, 1, 1)$. In this case, there is only one global maximum.

- a. Identify first the four local maxima by using different starting values for the Nelder-Mead or another algorithm of your choice. Which of the local maxima is the global maximum?
- b. Use now the Particle Swarm Optimization in the R-function `psoptim` in package `pso` with the aim to identify the global maximum. Choose different values for the swarm size and for the average percentage of informants for each particle. (If you like, you can also investigate other parameters in the same way; but this is optional). You might consider another maximum number of iterations to keep the running times reasonable. Do not forget to define an appropriate search space.

Run the PSO repeatedly (e.g. 100 times) for each chosen configuration of swarm size and average percentage of informants. Check if the result was close to the true global maximum. Based on this, Identify values for these parameters which work well for maximisation of this function g .

Problem 3.3

An experiment was conducted investigating how the growth of garden cress depends on a (potentially) toxic fertilizer. The investigated range of fertilizer was between 0 and 1.2% concentration in the water. Growth of garden cress was measured in yield in *mg* after 5.5 days of growth. $n = 81$ experiments were conducted and the data is on the homepage in the file `cressdata.txt` (columns: observation number, fertilizer concentration, yield).

A fifth-degree polynomial is supposed to be fit to the data using least squares with L^1 regularisation (Lasso). The objective function to be minimised is

$$g(\boldsymbol{\beta}) = \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1, ** \quad (1)$$

where \mathbf{X} is the design matrix with columns 1, fertilizer, ..., fertilizer⁵, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_5)^\top$ the parameter vector to be determined and \mathbf{y} is the yield-data. $\lambda \geq 0$ is a fixed regularisation constant.

Note that regularisation in the situation of high degree polynomial models is illustrated in Chapter 5.2 of Goodfellow, Bengio and Courville (2016). Note further that $\lambda = 0$ corresponds to the least squares estimation, where the solution of the optimisation problem is $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$.

- a. Program yourself the Lasso objective function (1). Choose several algorithms to solve the optimisation problem; you can use e.g. some methods in `optim` or `psoptim`. Motivate choice of parameters and starting values used in these algorithms. Determine the Lasso estimator for $\lambda = 0, 10$ and two or three other λ -values with the chosen algorithms. Report also the objective value $g(\hat{\boldsymbol{\beta}})$ in order to compare the algorithms. Which algorithm is best, which are good?
- b. Choose the estimated $\boldsymbol{\beta}$ of a good algorithm and plot the predicted regression functions for the chosen λ -values together with the data. Comment on the shape of the regression functions having the illustration in Chapter 5.2 in Goodfellow et al. (2016) in mind.

**In an earlier version of this assignment, I forgot by mistake the square at the LS-expression. The problem works with or without the square, but the choice $\lambda = 10$ in Part a. is most meaningful if the first norm is squared.