Stockholms universitet

Optimisation Algorithms in Statistics II – Spring 2021
Assignment 3

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

April 27, 2021

Perform the solutions individually and send your report **until May 11** to me. Try to keep this deadline. However, if you have problems with it, there will be a final deadline on August 31 for all assignments.
Please send me one **pdf-file** with your report (alternatively, Word is ok, too), and additionally, please send me your code in one separate **plain-text file** (an R-markdown, .rmd, is possible but not required).

Problem 3.1

Consider planning of an experiment when a cubic regression model

$$y_i = \beta_0 + \beta_1 w_i + \beta_2 w_i^2 + \beta_3 w_i^3 + \varepsilon_i, \quad i = 1, \ldots, n,$$

is assumed. The predictor variable w_i might be chosen in the interval $[-1, 1]$. However, due to practical circumstances, a distance of 0.05 between design points is required and at most one observation can be made at each point. Therefore, we require that observations can only be made using $w \in \{-1, -0.95, -0.9, \ldots, 1\}$. Further, the sample size n might be chosen by the experimenter, but each observation has a cost. To balance between the higher information from more observations and their higher cost, a penalized D-optimality criterion should be optimised here:

$$\text{Minimise } n/5 - \log\{\det(X^T X)\}$$

where X is the design matrix having rows $(1, w_i, w_i^2, w_i^3)$ and log is the natural logarithm. An example function is provided in the file `crit_HA3.r` on the course homepage which computes this criterion.

a. Write your own simulated annealing algorithm for the problem described here. You need to choose a reasonable proposal distribution (possibly on some neighbourhood of a design) and a cooling schedule.

b. What is the optimal design in this case based on your optimisation?

c. Consider at least three different cooling schedules and run them repeated times. Report the proportion of successes for these cooling schedules. Is there a clear winning schedule?

Problem 3.2

Consider the same optimal design problem as in Problem 3.1.

a. Plan a genetic algorithm specifically for this case by describing a selection mechanism, a fitness function, a crossover rule and a mutation rule. For this part, you need not to write code.

b. **(optional)** Write the genetic algorithm in `R` and run it. Do you obtain the same optimal design as with your simulated annealing algorithm?