Stockholms universitet

Optimisation Algorithms in Statistics I – Autumn 2020
Assignment 5

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

November 13, 2020

Perform the solutions individually and send your report **until November 27** 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 5.1

We have independent data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from a bivariate Cauchy-distribution with unknown location parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ and known scale matrix $\Sigma = I$. The negative log likelihood function of the bivariate Cauchy-distribution is (after removing a constant) proportional to

$$g(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log(1 + \|\boldsymbol{\theta} - \mathbf{x}_i\|_2^2).$$

This function has to be *minimised* to find the Maximum Likelihood estimator. It's gradient with respect to the vector $\boldsymbol{\theta}$ is

$$\sum_{i=1}^{n} (\boldsymbol{\theta} - \mathbf{x}_i) \frac{2}{1 + \|\boldsymbol{\theta} - \mathbf{x}_i\|_2^2}.$$

$n = 5$ data points are given:

$$\mathbf{x}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} -4 \\ 11 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 19 \\ 8 \end{pmatrix}, \mathbf{x}_4 = \begin{pmatrix} -1 \\ 4 \end{pmatrix}, \mathbf{x}_5 = \begin{pmatrix} 1 \\ 28 \end{pmatrix}.$$

a. Plot the function $g(\boldsymbol{\theta})$ as contour plot for $\boldsymbol{\theta} \in S = [-20, 30] \times [-10, 40]$.

b. Program your own optimiser with the Stochastic steepest descent algorithm with momentum. The function should allow for linearly changing the step size α (see lecture notes and Chapter 8.3.1 of Goodfellow, Bengio and Courville, 2016). It should also have a parameter β for the momentum (see lecture notes and Chapter 8.3.2 of Goodfellow, Bengio and Courville, 2016, where the parameter is called α). The iteration number can be fixed, $T = 10000$.

c. Set first the momentum parameter $\beta = 0$. Determine good parameters for the linear decreasing step size (called $\alpha_0, \tau, \alpha_\tau$ in the lecture), such that the algorithm arrives close to the global maximum of g in most of the time, independent of which starting value in the set S is used. Plot a search path into the contour plot using your chosen parameters.

d. Use now the momentum as well. Determine parameters for the linear decreasing step size and the momentum parameter β, such that the algorithm arrives close to the global maximum of g in most of the time, independent of which starting value in the set S is used. Plot a search path into the contour plot using your chosen parameters.

Problem 5.2

Use the same function g as in Problem 5.1. Determine now the minimum using Simulated annealing with `method="SANN"` in `optim`.

a. Read the documentation of the SANN-method in `optim`. Especially check the parameters `gr, maxit, temp, tmax`. Explain how the temperature changes over the iterations according to the documentation.

b. Choose different values for the starting temperature for the cooling schedule `temp` to find a good value such that the algorithm converges in most of the times to the right global minimum. You can use here the starting vector $\boldsymbol{\theta}^{(0)} = (30, -10)^\top$. If you want, you can check the results for one or two other starting values as well.

Problem 5.3 (administrative)

a. Send my your Swedish personal id-number such that I can use it in your course-certificate.

b. You will receive a course evaluation survey in some weeks; the deadline will be communicated in the survey. Please fill in this anonymous survey then – it is important for future courses. Thank you for your feedback!