Stockholms universitet

Optimisation Algorithms in Statistics I – Autumn 2020
Assignment 2

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

October 13, 2020

Perform the solutions individually and send your report **until October 22** 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 2.1

We consider the quadratic two-dimensional function

$$g(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^\top A\mathbf{x}, \quad \mathbf{x} \in I\!R^2.$$

with a symmetric and positive definite 2×2-matrix A. The function g has a maximum at $(0,0)^\top$ and the gradient is $\mathbf{g}'(\mathbf{x}) = -A\mathbf{x}$. For A, we consider three different function with

$$A_1 = \begin{pmatrix} 8 & 1 \\ 1 & 8 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & -2 \\ -2 & 12 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 0 \\ 0 & 100 \end{pmatrix}.$$

a. Program your own function for steepest ascent with Polyak's momentum with fixed step size α (without halving alpha or doing line search) and fixed momentum parameter β. These two parameters should be options in your function such that you can test different options. Choose a stopping criterion such that you have correct results up to around 6 digits. Your function should report the number of iterations used.

b. Plot a contour plot for the three cases and compute with `R` the eigenvalues of the matrices A_i and the condition number κ. For both the steepest ascent and the steepest ascent with momentum, compute the optimal parameters and the best convergence rate ρ in the three cases.

c. Run the steepest ascent method (i.e. use $\beta = 0$ in your function) using a starting value $\mathbf{x}^{(0)} = (-4,-2)^\top$. Use the optimal value and several other α-values in a grid of 0.01 or 0.02 around the optimal value. Report if the algorithm successfully found the maximum and how many iterations were needed for each parameter value. Is the performance best for the theoretically best α-value?

d. Run the steepest ascent method with momentum. Use the optimal α, β and some values around them and report convergence and number of iterations. How much does the momentum method improve performance.

e. **(optional)** Add the search paths for the optimal parameter values to each of the contour plots.

Problem 2.2

Use your steepest ascent program from Problem 1.3. Assume that the x-variable is measured in μg instead of mg. This means that the data is in Table 1.

x_i	0	0	0	100	100	300	300	900	900	900
y_i	0	0	1	0	1	1	1	0	1	1

Table 1: Data for Problem 2.2

The ML-estimate is unchanged for β_0 and is 1/1000 of the previous value for β_1. Use a pair of starting values which you used for Problem 1.3 and divide the β_1-value with 1000. If you run the steepest ascent program again, is the performance the same as when you analysed the mg data? Why or why not?

Problem 2.3

As in Problem 1.3 and 2.2, we consider ML estimation for simple logistic regression

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

a. Program a Stochastic gradient ascent algorithm with a fixed step size α and a predefined total number of iterations T for simple logistic regression. Your program should also plot the computed $(\beta_0^{(t)}, \beta_1^{(t)})$ in each iteration t ($t = 0, 1, \dots, T$) such that you can monitor the search path.

b. Analyze the dataset `logist.txt` (homepage; first column is x, second column y) with your algorithm using the starting value $(\beta_0^{(0)}, \beta_1^{(0)}) = (0.2, 0.5)$. Choose the total number of iterations T and the step size α. You might need to test different options first to come to a good choice. Explain why you have chosen these values T and α.