Stockholms universitet

Optimisation Algorithms in Statistics II – Spring 2021
Assignment 1

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

March 23, 2021

Perform the solutions individually and send your report **until April 12** 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 1.1

Consider a simple logistic regression

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

and the estimation of $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ using maximum likelihood based on n observations.

The scaled negative log-likelihood $g(\mathbf{b}) = \frac{1}{n}\sum_{i=1}^n g_i(\mathbf{b})$ and its derivative $\mathbf{g}'(\mathbf{b}) = \frac{1}{n}\sum_{i=1}^n \mathbf{g}_i'(\mathbf{b})$ are described in a separate document (the link is among the reading on the course homepage). Let R be a discrete random variable, uniformly distributed on the set $\{1, \ldots, n\}$.

a. Show that the scaled negative log-likelihood is L-smooth and determine a constant L fulfilling the Lipschitz property of the derivative.

b. Show: When we apply SGD, we see an improvement in expectation from iteration t to iteration $t + 1$ if
$$\alpha < f(\mathbf{b}^{(t)}) := \frac{2\|\mathbf{g}'(\mathbf{b}^{(t)})\|_2^2}{LE\|\mathbf{g}_R'(\mathbf{b}^{(t)})\|_2^2}.$$

c. Compute $\|\mathbf{g}'(\mathbf{b})\|_2^2$ and $E\|\mathbf{g}_R'(\mathbf{b})\|_2^2$ and determine a constant s such that $E\|\mathbf{g}_R'(\mathbf{b})\|_2^2 \leq s$.

d. For the dataset in Table 1 (used in the first part of the course), compute L and s numerically. Produce a contour plot of function f in Part b. and explain the meaning of the contour lines in this plot.

x_i	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 Problem 1.1d and 1.3b and c

Problem 1.2

Consider least squares for a linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

with the function

$$g(\mathbf{b}) = \frac{1}{n}\|\mathbf{X}\mathbf{b} - \mathbf{y}\|_2^2$$

to be minimised. Show that g is L-smooth and m-strongly convex. Present expressions for L and m.

Problem 1.3

We consider again maximum likelihood estimation for simple logistic regression as in Problem 1.1 and use the same example dataset.

a. Program your own quasi-Newton algorithm using the BFGS method using a step-size halving line search.

b. Run the program for the dataset in Table 1 and report the number of iterations used.

c. **(optional)** Compare with the number of iterations used by the steepest ascent/descent method when the same starting value for $(\beta_0, \beta_1)^\top$ is used.