Assignment Topic №4: Stochastic optimization

Dmytro Marushkevych

September 6, 2022

The last assignment topic is on stochastic optimization. If you draw the topic "Stochastic optimization" at the oral exam, you will have to present a solution of one of the two assignments below.

Remember the five points:

- How can you test that your implementation is correct?
- Can you implement alternative solutions?
- Can the code be restructured e.g. by modularization, abstraction or object oriented programming to improve readability?
- How does the implementation perform (benchmarking)?
- Where are the bottlenecks (profiling), and what can you do about them?

As for the other assignments, performance tests are important, but remember to make correct comparisons. Benchmarks are interesting for e.g. gradient evaluations, and convergence of the optimization algorithm should be investigated as a function of real time (not iterations).

With stochastic optimization algorithms it is of particular interest to investigate how convergence is affected by the various control parameters for the different algorithms such as size of minibatch and learning rate for SGD.

Assignment 1: Logistic regression smoothing

Consider the logistic regression model with $y_i \in \{0, 1\}$ and $x_i \in \mathbb{R}$ such that with $p_i(\beta) = P(Y_i = 1 \mid X_i = x_i)$

$$\log \frac{p_i(\beta)}{1 - p_i(\beta)} = f(x_i \mid \beta) = (\varphi_1(x_i), \ldots, \varphi_p(x_i))^T \beta$$

for some $\beta \in \mathbb{R}^p$ and fixed basis functions $\varphi_1, \ldots, \varphi_p : \mathbb{R} \to \mathbb{R}$. The assignment is on minimizing the *penalized* negative log-likelihood

$$H(\beta) = -\frac{1}{N} \sum_{i=1}^N \Big(y_i \log p_i(\beta) + (1 - y_i) \log(1 - p_i(\beta)) \Big) + \lambda \|f_\beta''\|_2^2$$

over the basis coefficients $\beta \in \mathbb{R}^p$. Implement stochastic optimization algorithms for estimating β for fixed λ. Use e.g. polynomial or B-spline basis functions, and test the regression model using the horse data and simulated data. (Data is available on Absalon. The binary variable indicates if the horse dies after hospital admission, and x is the temperature of the horse when admitted.)

Compare the stochastic gradient algorithms to other optimization algorithms and investigate how different choices of basis (e.g. the default B-spline basis or the Demmler-Reinsch basis), affect convergence.

Assignment 2: Log-logistic dose-response curves

The four-parameter log-logistic dose-response model is the nonlinear model of the mean of a real valued random variable Y given x defined as

$$f(x \mid \alpha, \beta, \gamma, \rho) = \gamma + \frac{\rho - \gamma}{1 + e^{\beta \log(x) - \alpha}}.$$

Here, $\alpha, \beta, \gamma, \rho \in \mathbb{R}$.

In this assignment you will consider the model with observations

$$Y_i = f(x_i \mid \alpha, \beta, \gamma, \rho) + \varepsilon_i$$

for $i = 1, \ldots, N$, with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ independent. The parameter vector $(\alpha, \beta, \gamma, \delta) \in \mathbb{R}^4$ will be estimated by nonlinear least squares, that is, by minimizing

$$\frac{1}{N} \sum_{i=1}^{N} \left(y_i - f(x_i \mid \alpha, \beta, \gamma, \rho) \right)^2.$$

Implement stochastic optimization algorithms for estimating the parameters of this model. Compare the resulting optimization algorithm(s) with non-stochastic algorithms e.g. gradient descent or the Newton algorithm. It is particularly interesting to investigate convergence of the algorithms for N large.

You can sample the x-s either from a grid, e.g. $e, e^2, \ldots, e^{15}$, or generate $\log(x)$ from a $\mathcal{N}(0, \omega^2)$-distribution. Consider how you could exploit if the x-s all fall in a small number of grid points in your implementation.