Assignments: Optimization

Niels Richard Hansen

September 30, 2019

The third assignment topic is optimization. If you draw the topic “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. That is, if you benchmark two implementations against each other, use exactly the same convergence criterion. Alternatively, investigate how the algorithms converge as a function of *real* time (not iterations).

Test of convergence and experiments with convergence criteria are important to investigate. Robustness towards the initial choice of parameters is important, and testing of the implementation on several (simulated) data sets is a good idea.

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

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

over the basis coefficients $\beta \in \mathbb{R}^p$. Implement the computation of H and its first and second derivative. Implement algorithms to minimize H for fixed λ and test them. Use 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.)

You can investigate the effect of using sparse matrices, how different optimization algorithms converge depending on the chosen basis (comparing e.g. the default B-spline basis to the Demmler-Reinsch basis), how convergence depend on n and p etc.

How can λ be chosen automatically?

Assignment 2: Random effects models

Consider the linear random effects model with observations

$$Y_{ij} = \beta_0 + \nu Z_i + \varepsilon_{ij}, \quad i = 1, \ldots, m; \quad j = 1, \ldots, n_i$$

where $Z_1, \ldots, Z_n$ are i.i.d. $\mathcal{N}(0,1)$-distributed and independent of $\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{1n_1}, \ldots, \varepsilon_{mn_m}$ that are themselves i.i.d. $\mathcal{N}(0, \sigma^2)$-distributed.

We do not observe Z_i (or ε_{ij} for that matter), and we are interested in estimating the parameters $\beta_0 \in \mathbb{R}$ and $\nu, \sigma > 0$. It is possible to compute the likelihood by observing that the vector (Y_{ij}) follows a joint Gaussian distribution with ν (and σ) entering into the covariance matrix.

In this assignment you are going to implement the EM-algorithm for maximum-likelihood estimation. Implement first a function that computes the maximum-likelihood estimator with complete observation of all y_{ij} and z_i. (Hint: this becomes a linear regression problem). Then implement the E-step and the full EM-algorithm.

Investigate performance and convergence. Investigate how sparsity can be exploited when m is large.

Implement a method for computing an estimate of the Fisher information.