Biostatistics 778: Advanced Statistical Computing

Homework 1

Due date: 2007-11-15

Problems

1. *Newton's method*. Recall the mixture model problem discussed in class. We observe data $y_1, \ldots, y_n$ and assume that

$$y_i \sim \lambda \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \lambda)\mathcal{N}(\mu_2, \sigma_2^2)$$

   where $\lambda \in [0, 1]$. Write a program that calculates the maximum likelihood estimates of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and λ using Newton's method. In addition to the point estimates, your program should also produce asymptotic standard errors for the MLEs. (Note: do not use the `nlm` or `optim` functions).

2. *Penalized Regression*. Write a function that computes maximum penalized likelihood estimates for a regression model with a quadratic penalty on the regression parameters. If $\ell(\boldsymbol{\beta})$ is the usual log-likelihood for the parameter vector $\boldsymbol{\beta}$, then your function should maximize

$$\ell_p(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) - \lambda \boldsymbol{\beta}'\boldsymbol{\beta}$$

   where λ is a user-specified penalty parameter. Specifically, your function should compute estimates for a Poisson model with log link and a logistic regression model. Use Newton's method to maximize the penalized likelihood ℓ_p (and again, do not use `nlm` or `optim`). Your function should return the following elements:

   - a vector containing the estimated parameters
   - a matrix containing an asymptotic covariance matrix for the parameters
   - the value of ℓ_p at the maximum
   - an indicator of whether the Newton algorithm converged or not
   - the number of iterations of the Newton algorithm used

Build an R package which incorporates the functions you have written above for the mixture model and for the penalized regression. The name of the R package should be `Homework1`. Your R package should:

- have a name space
- have a help file corresponding to each exported function
- the help files should have examples of each function's usage
- pass `R CMD check` (using R 2.6.0)