Assignment 2 | 140.778 (Adv Stat Comp)
Due: 20 Dec | 2nd term, 2000–2001

EM algorithm

Part A

Consider the linear model $y^\star = X\beta + \varepsilon$ where $\varepsilon \sim \mathrm{N}(0, \sigma^2 I)$.

Suppose that X is observed, but that rather than observing $y_i^\star$, we observe $y_i = \min\{y_i^\star, c\}$ (where the censoring value c is known and is constant for all i).

Work out an EM algorithm to obtain the MLEs of β and σ. (You will no doubt find it very useful to look at problem 10.8 in K Lange, *Numerical analysis for statisticians*, pp 126–127.)

Write an R (or Splus, or even Matlab) function to implement your algorithm.

Part B

1. Download the comma-delimited file `data2.csv` from the course web page. The y's are censored at 10.0.
2. Fit the linear model $y = X\beta + \varepsilon$, ignoring the censoring.
3. Use your EM function to get the MLEs of β and σ, taking account of the censoring.
4. Consider playing with starting values. Does the likelihood surface have multiple modes?
5. If time permits, try using a parametric and/or nonparametric bootstrap to get SEs of β and σ.
6. If time permits, write a function to calculate the observed-data log likelihood, and apply the `nlm` (or `nlmin`) function to get MLEs. Using `hessian=TRUE` will allow you to get an estimated variance matrix.
7. Consider comparing the number of iterations, computer time, and sensitivity to starting values for the EM vs `nlm` approaches.

Useful R functions:
`read.table`, `nlm`, `lm`, `dnorm`, `pnorm`, `unix.time`