STSCI 7170 - Homework 5 - Fall 2019

JGB

12/01/2019

DUE DATE: Tuesday, December 10, 2019

INSTRUCTIONS: Students should work independently of their classmates on these problems.

The textbook "Linear Models: A Mean Model Approach" by Barry Kurt Moser is available online from the Cornell library.

1a. Create a function in R which implements the EM algorithm for fitting a normal theory linear mixed variance components model.

1b. Use your EM fitting function to fit the model with four variance components to the blood serum data from homework 3. Your function should generate output showing the progress of the iterations for the fixed effect parameters, the variance components, and a convergence rule statistic.

1c. Use your EM function to fit a quadratic penalized spline model to the light detection and ranging (LIDAR) experiment available from the class canvas site. That is, fit the model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \sum_{j=1}^{k} u_j (x_i - \kappa_j)_+^2 + \epsilon_i \,,$$

where y is logratio and x is range, and $k = 9$ (i.e. 9 knots). [Note here that $(x_i - \kappa_j)_+^2$ means the square of $(x_i - \kappa_j)_+$.]

After fitting the model, overlay the fitted curve,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \sum_{j=1}^{k} \hat{u}_j (x - \kappa_j)_+^2 \,,$$

on a scatterplot of the raw data.

Upload your R code files on Canvas in a (compressed) folder called *hw5-yournetid*. The files should include:

yournetid-em.R - the EM fitting function

yournetid-serum.R - commands used create the input objects and to fit the variance components model to the blood serum data.

yournetid-lidar.R - commands used to fit the quadratic smoothing spline model to the lidar data, and to create the scatterplot and smooth.

The files should be such that I can run the code using the source command in R; i.e.

source("yournetid-em.R")

source("yournetid-serum.R")

source("yournetid-lidar.R")

assuming the data sets are in my working directory.

The outputs should include the parameter estimates and the marginal log-likelihood at each iteration of EM.

Do not require any packages that I may need to install.