Statistics 322, Statistical Computing, Home Work 10

Handed Out: Nov. 19, 1996 Due: Dec. 3, 1996

This problem is about fitting B-splines (i.e. regression splines) to data. Use the same data set as used in the MATLAB script `s322eg14.m`.

The data set may be found in the file `s322eg14.dat` on the WWW. This file has mostly columns, which are the vectors of x_i, $m(x_i)$, Y_i respectively. However, being saved from MATLAB output, it also has some extraneous stuff, too, so you probably should edit that out first. A simple way to read this data set into your program is with the command:

```
loads 322eg14.dat ;
```

Note that this creates a new variable `s322eg14`. Look at this, and then modify it is needed. A good plan is to first make your program reads the data and gets it into the right form before starting to program the B-splines.

For higher tech types of file reading (e.g. "formatted reads", and reads involving strings), use commands such as `fopen` , `fread`, `fscanf`, `sscanf`, and `fclose`. When you save data in files, a much more efficient way (in terms of both disk space, and also organization) to do this is the command `save`, which saves a list of variables that you choose in a special `.mat` format.

Turn in a single page, with an overlay of several regression splines fit to the data, where the numbers of knots (including the endpoints, -1, and 1) are: 2, 3, 5, 9, 17. Take the knots to be equally spaced on the interval [-1,1]. The idea is to show how "number of knots works like a smoothing parameter".

Notes:

1. Note that the numbers of knots above have been chosen to allow "nesting". One possible algorithm might expand the design matrix at each step (i.e. just add new columns for the additional basis functions, this is the "updating" idea that makes the regression spline formulation so attractive) but this is not essential for this simple example. Indicate where the knots are on the graph.

2. Indicate your personal preference for the number of knots (i.e. the amount of smoothing).

3. Extra points will be given for especially attractive and easy to interpret pictures.