1. [20 pts] (3.3 in the textbook) Consider the trade-off of bias and variance of the smoothing spline by the simulation data obtained using eq(3.13) (page 110).
(a) Create an R object for drawing graphs similar to those in figure 3.2 (page 108) and figure 3.3 (page 109) for the smoothing spline.
(b) Run the object produced in (a) and observe the influence of the smoothing parameter on bias, variance and mean squared error. Consider the relationship between the appearance of data and these three values, and the behavior of three values near the two ends of the data region.

2. [30 pts] (3.6 in the textbook) Consider the trade-off of bias versus variance of the Nadaraya-Watson estimator of simulation data obtained using eq(3.13) (page 110).
(a) Create an R object for drawing graphs similar to those in figure 3.2 (page 108) and figure 3.3 (page 109) by utilizing the object presented in (A) of this chapter (chapter 3).
(b) Run the object produced in (a) and observe the influence of the bandwidth on bias, variance and mean squared error. Consider the relationship between the appearance of data and these three values, and the behavior of these three values near the two ends of the data region.
(c) Estimate the optimal bandwidth in terms of CV or GCV using the object presented in (B) in this chapter (Chapter 3). Construct graphs of bias, variance and mean squared error using the object created in (a), and verify that the optimal bandwidth in terms of CV or GCV yields satisfactory bias, variance and mean squared error values. If the optimal bandwidth does mot provide a sufficient result, discuss the reason and attempt the use of other statistics for choosing the optimum bandwidth.

3. [10 pts] (3.8 in the textbook) Construct an R object for drawing weight diagrams (i.e. equivalent kernels) of local linear regression on the basis of eq(3.127) (page 142) and eq(3.128) (page 143). Assume that $\{X_i\}$ are $\{1, 2, \cdots, 20\}$ and compare the values of the equivalent kernels with those of the Nadaraya-Watson estimator (in particular, the equivalent kernels for calculating estimates near two ends of the data region). Furthermore, verify eq(3.130) (page 143) numerically.

4. [20 pts] (3.10 in the textbook) Check numerically that the natural boundary conditions eq(3.172) (page 158) and eq(3.173) (page 158) minimize the value defined by eq(3.171) (page 158), which the cubic spline function gives.
(a) Create an R object for constructing a cubic spline function under diverse boundary conditions. When the natural boundary conditions are employed, the function is realized by solving eq(3.192) (page 165). When the natural boundary conditions are not employed, the right-hand side of eq(3.192) is modified.
(b) Produce an R object for calculating numerically the value defined by eq(3.171) using the cubic spline function.

5. [10 pts] (3.11 in the textbook) Calculate the values of the elements of $\boldsymbol{L}$ and $(\boldsymbol{I} + \lambda \boldsymbol{L})^{-1}$ defined in eq(3.231) (page 176) on the assumption that the predictor values are written as $\{X_i\} = \{1, 2, 3, \cdots, 20\}$. Eq(3.229) (page 176) can be used for this purpose. Attempt the use of previously reported procedure (Chapter 2) to derive them and compare the results.

6. [10 pts] (3.15 in the textbook) The cross-validation (for LOESS) implemented in the object presented in (I) of this chapter is based on the assumption that the estimate obtained by deleting

one data is represented by eq(3.41) (page 117); however, this is an approximation. Compare the results obtained using the object presented in (I) with those obtained by conducting the cross-validation based on its real definition.

7. [20 pts] (3.16 in the textbook) Consider the fact that the supersmoother sometimes does not function well when the number of data is large.

(a) Generate data sets (the number of data is n) based on the equation below.

$$Y_i = \sin(0.2\pi X_i) + \epsilon_i,$$

where $\{X_i\}$ is $\{0.1, 0.2, \cdots, 0.1n\}$ and ϵ_i is a realization of $N(0, 0.1^2)$. The number of data (n) is 100, 500 and 2000.

(b) Confirm that the estimates are not satisfactory when the number of data is large. Discuss the mechanism behind this phenomenon and determine countermeasures.