Suggestions, PS 10

(a) Fitting a smoothing spline (as in section 5.4) can be done with various computer packages. The R package smooth.spline is a good choice.

To get confidence bands, one option is to use formula (8.4) directly. Alternatively, you can estimate variance of the fitted function $\widehat{f}$ at the data points x_i themselves using $\widehat{\mathbf{f}} = S_\lambda \mathbf{y}$, with $\widehat{\mathbf{f}} = \widehat{\mathbf{y}}$ the vector of fitted values $\widehat{f}(x_i)$. Show $V(\widehat{\mathbf{f}}) = S_\lambda V(\mathbf{y}) S_\lambda^T$, with $V(\cdot)$ the covariance matrix. This gives variance of $f(x_i)$; you can interpolate this (even with splines) to get the variance of $f(x)$. Note you can get a column of S_λ by letting $\mathbf{y}$ above be the vector with all zeroes and a single 1, using a package like smooth.spline. (After you choose λ, check that your spline program uses and returns the value of λ whenever you request it).

(b) Here the matrix $\mathbf{H}$ will be $N \times p$, where in this case p is the number of basis functions $h_j(x)$, which for smoothing splines is the set of *natural splines* with knots at all of the data points x_i (any choice of spline basis functions will do; you can use the basis in Section 5.2.1). Again you can try a medium ($\tau = 1$) and a large value of τ. Using $\Sigma = I$ as the matrix part of the prior should be fine.

Note: there are 485 data points and as many basis functions for the natural splines. Thus you have a 485×485 matrix H. Operations on this large basis set may go well, but in some cases numerical instabilities may arise, depending on the algorithms you use. If this is the case it may be worthwhile to try for part (b) a basis of natural splines with a smaller number of *fixed* knots to do (b) a second time.

(c) Here the smoothing splines are computed on bootstrap samples; overlapping sample points should not cause any problems.

(d) The cross-validation error is just the error given by equation (7.49).

If you aggregate (average) the bootstrap estimates in part (c), you will get the bagging version of the estimator. You can then estimate the cross-validation error of this alternative estimate.