ESL 5.4

Ex. 5.4 Consider the truncated power series representation for cubic splines with K interior knots. Let

$$f(X) = \sum_{j=0}^{3} \beta_j X^j + \sum_{k=1}^{K} \theta_k (X - \xi_k)_+^3. \tag{5.70}$$

Prove that the natural boundary conditions for natural cubic splines (Section 5.2.1) imply the following linear constraints on the coefficients:

$$\begin{aligned} \beta_2 &= 0, & \textstyle\sum_{k=1}^{K} \theta_k &= 0, \\ \beta_3 &= 0, & \textstyle\sum_{k=1}^{K} \xi_k \theta_k &= 0. \end{aligned} \tag{5.71}$$

Hence derive the basis (5.4) and (5.5).

$$f(x) = \sum_{j=0}^{3} \beta_j X^j + \sum_{k=1}^{K} \theta_k (X - \xi_k)_+^3$$

The natural boundary condition is $f''(x) = 0$.

when $x < \xi_1$

$$f(x) = \sum_{j=0}^{3} \beta_j X^j = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$$

$$f''(X) = 2\beta_2 + 6\beta_3 X = 0$$

$$\Rightarrow \beta_2 = 0, \beta_3 = 0$$

when $x \geq \xi_i$,

$$f(X) = \beta_0 + \beta_1 X + \sum_{k=1}^{K} \theta_k (X - \xi_k)_+^3 \quad (*)$$

$$\Rightarrow f''(x) = 6 \sum_{k=1}^{K} \theta_k (X - \xi_k)$$

when $f''(x) = 0$, we get $\sum_{k=1}^{K} \theta X = \sum_{k=1}^{K} \theta_k \xi_k$.

Thus $\sum_{k=1}^{K} \theta_k = 0$ and $\sum_{k=1}^{K} \theta_k \xi_k = 0$. $(**)$

From the above, we have proved the linear constraints on the coefficients of the natural cubic spline. Next, let's derive the basis of 5.4 and 5.5.

$$N_1(X) = 1, \quad N_2(X) = X, \quad N_{k+2}(X) = d_k(X) - d_{K-1}(X), \qquad (5.4)$$

where

$$d_k(X) = \frac{(X-\xi_k)^3_+ - (X-\xi_K)^3_+}{\xi_K - \xi_k}. \qquad (5.5)$$

Each of these basis functions can be seen to have zero second and third derivative for $X \geq \xi_K$.

Given $f(X) = \sum_{j=0}^{3} \beta_j X^j + \sum_{k=1}^{K} \theta_k (X-\xi_k)^3_+$, $\beta_2 = \beta_3 = 0$,

we have $0 = \sum_{k=1}^{K} \theta_k - \sum_{k=1}^{K} \theta_k \xi_k$

$= \sum_{k=1}^{K}$

$= \sum_{k=1}^{K-1} \theta_k (\xi_K - \xi_k)$

From $\sum_{k=1}^{K} \theta_k = 0$, we have

$$\theta_K = \sum_{k=1}^{K-1} \theta_k$$

Ex. 5.13 You have fitted a smoothing spline $\hat{f}_\lambda$ to a sample of N pairs (x_i, y_i). Suppose you augment your original sample with the pair $x_0, \hat{f}_\lambda(x_0)$, and refit; describe the result. Use this to derive the N-fold cross-validation formula (5.26).

$$\hat{f}_\lambda = \operatorname{argmin} RSS(f_\lambda)$$

$$= \operatorname{argmin} \sum_{i=1}^{N} (y_i - f(x_i))^2 + \lambda \int (f''(t))^2 dt$$

$$= \sum_{i=1}^{N} (y_i - f(x_i))^2 + (\hat{f}_\lambda(x_0) - f(x_0))^2 + \lambda \int (f''(t))^2 dt$$

$$\Rightarrow \sum_{i=0}^{N} (y_i - f(x_i))^2 + \lambda \int (f''(t))^2 dt$$

When using CV, (x_i, y_i) are separate for the validation purpose. To augment the training set with $(x_0, \hat{f}_\lambda(x_0))$, we rwrite $\hat{f}_\lambda(x_i)$:

$$\hat{f}_\lambda(x_i) = \sum_{j=1}^{N} S_\lambda(i,j) y_j$$

$$= \sum_{j \neq i} S_\lambda(i,j) y_i + S_\lambda(i,i) y_i$$

$\hat{f}_\lambda^{(-i)}(x_i)$ is predicted value for the i-th case for $\{x_i, y_i\}$.

$$\hat{f}_\lambda^{(-i)}(x_i) = \sum_{j \neq i} S_\lambda(i,j) y_i + S_\lambda(i,i) \hat{f}_\lambda^{(-i)}(x_i)$$

$$\hat{f}_\lambda^{(-i)}(x_i) = \hat{f}_\lambda(x_i) - S_\lambda(i,i) y_i + S_\lambda(i,i) \hat{f}_\lambda^{(-i)}(x_i)$$

$$\hat{f}_\lambda^{(-i)}(x_i) - S_\lambda(i,i) \hat{f}_\lambda^{(-i)}(x_i) = \hat{f}_\lambda(x_i) - S_\lambda(i,i) y_i$$

$$\hat{f}_\lambda^{(-i)}(x_i) = \frac{\hat{f}_\lambda(x_i) - S_\lambda(i,i) y_i}{1 - S_\lambda(i,i)}$$

Thus, $$y_i - \hat{f}_\lambda^{(-i)}(x_i) = y_i - \frac{\hat{f}_\lambda(x_i) - S_\lambda(i,i)\,y_i}{1 - S_\lambda(i,i)}$$

$$= \frac{y_i - \cancel{y_i S_\lambda(i,i)} - \hat{f}_\lambda(x_i) + \cancel{S_\lambda(i,i)\,y_i}}{1 - S_\lambda(i,i)}$$

$$= \frac{y_i - \hat{f}_\lambda(x_i)}{1 - S_\lambda(i,i)}$$