Smoothing Splines

- In Regression Splines (let's use NCS), we need to choose the number and the location of knots.

- What's a Smoothing Spline? Start with a "naive" solution: put knots at all the observed data points $(x_1, \ldots, x_n)$:

$$\mathbf{y}_{n \times 1} = \mathbf{F}_{n \times n} \boldsymbol{\beta}_{n \times 1}.$$

  Instead of selecting knots, let's carry out the following ridge regression (Ω will be defined later):

$$\min_{\boldsymbol{\beta}} \left[\|\mathbf{y} - \mathbf{F}\boldsymbol{\beta}\|^2 + \lambda \boldsymbol{\beta}^t \Omega \boldsymbol{\beta} \right],$$

  where the tuning parameter λ is often chosen by CV.

- Next we'll see how smoothing splines are derived from a different aspect.

Roughness Penalty Approach

- Let $S[a,b]$ be the space of all "smooth" functions defined on $[a,b]$.
- Among all the functions in $S[a,b]$, look for the minimizer of the following penalized residual sum of squares

$$\mathsf{RSS}_\lambda(g) = \sum_{i=1}^{n}[y_i - g(x_i)]^2 + \lambda \int_a^b [g''(x)]^2 dx,$$

  where λ is a smoothing parameter.

- Theorem. $\min_g \mathsf{RSS}_\lambda(g) = \min_{\tilde{g}} \mathsf{RSS}_\lambda(\tilde{g})$ where $\tilde{g}$ is a NCS with knots at the n data points $x_1, \ldots, x_n$ (WLOG, $x_i \neq x_j$ and $x_1 < x_2 < \cdots < x_n$)

(WLOG, assume $n \geq 2$.) Let g be a smooth function on $[a, b]$ and $\tilde{g}$ be a NCS with knots at $\{x_i\}_{i=1}^n$ satisfying

$$g(x_i) = \tilde{g}(x_i), \quad i = 1 : n. \tag{1}$$

First, such $\tilde{g}$ exists since NCS with n knots has n dfs, so we can pick the n coefficients properly such that (1) is satisfied.

Next we want to show that

$$\int_a^b [g''(x)]^2 dx \geq \int_a^b [\tilde{g}''(x)]^2 dx \quad (*)$$

with equality holds if and only if $\tilde{g} \equiv g$. Recall that

$$\mathsf{RSS}_\lambda(g) = \sum_{i=1}^n [y_i - g(x_i)]^2 + \lambda \int_a^b [g''(x)]^2 dx.$$

So it is easy to conclude that if $(*)$ holds,

$$\mathsf{RSS}_\lambda(\tilde{g}) \leq \mathsf{RSS}_\lambda(g).$$

That is, for any smooth function g, we can find a NCS $\tilde{g}$ which matches $g(x_i)$ on the n samples and whose penalized residual sum of squares is not worse than the one of g. So Theorem follows.

PROOF : We will use integration by parts and the fact that $\tilde{g}$ is a NCS.

$$h(x) = g(x) - \tilde{g}(x). \quad \text{Note } h(x_i) = 0 \text{ for } i = 1, \ldots, n.$$

$$\begin{aligned}
\int_a^b [g''(x)]^2 dx &= \int_a^b \left[\tilde{g}''(x) + h''(x)\right]^2 dx \\
&= \int_a^b [\tilde{g}''(x)]^2 dx + \int_a^b [h''(x)]^2 dx + 2 \underbrace{\int_a^b \tilde{g}''(x) h''(x) dx}_{=0}
\end{aligned}$$

$$\begin{aligned}
\int_a^b \tilde{g}''(x) h''(x) dx &= \int_a^b \tilde{g}''(x) dh'(x) = \underbrace{h'(x)\tilde{g}''(x)\Big|_a^b}_{=0} - \int_a^b h'(x)\tilde{g}^{(3)}(x) dx \\
&= -\sum_{i=1}^{n-1} \tilde{g}^{(3)}\left(\frac{x_i + x_{i+1}}{2}\right) \int_{x_i}^{x_{i+1}} h'(x) dx \\
&= -\sum_{i=1}^{n-1} \tilde{g}^{(3)}\left(\frac{x_i + x_{i+1}}{2}\right) h(x)\Big|_{x_i}^{x_{i+1}} = 0
\end{aligned}$$