1. Closed form of Smoothing Splines. Let $N_1(x), N_2(x), \ldots, N_n(x)$ be the n-basis for natural cubic splines

(a). Derive the closed form of smoothing spline solution

Given that the solution is in fact that of a natural spline, we use the equation $f(x) = \sum_{j=1}^{n} N_j(x)\theta_j$ where $N_j(x)$ is a set of N-dimensional basis functions used to represent natural cubic splines.

$$f''(t) = \sum_{i=1}^{n} N_i''(t)\theta_i \quad \text{for some } i \text{ and } t$$

Let us square this double derivative and take the integral

$$\int [F''(t)]^2 dt = \int \left[\sum_{i=1}^{n} N_i''(t)\theta_i\right] * \left[\sum_{j=1} N_j''(t)\theta_j\right] dt$$

Let us rearrange

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \int N_i''(t) N_j''(t)\, dt\, \theta_i \theta_j$$

$$= \theta^T \Omega \theta$$

where each $\Omega_{ij} = \int N_i''(t) N_j''(t) dt$

We can reduce the criterion to

$$RSS(\theta,\lambda) = (y-N\theta)^T(y-N\theta)+\lambda\theta^T\Omega\theta$$

Here, $N_{ij} = N_j(x_i)$

$\Omega_{ij} = \int N_j''(t)N_i''(t)dt$

as previously mentioned.

Let us take the derivative of the above w.r.t θ_1

$$-2N^T(Y-N\hat{\theta})+2\lambda\Omega\hat{\theta}=0$$

$$\cancel{2}\lambda\Omega\hat{\theta} = \cancel{2}N^T(Y-N\hat{\theta})$$

$$\lambda\Omega\hat{\theta} = N^TY-N^TN\hat{\theta}$$

$$\hat{\theta} = (N^TN+\lambda\Omega)^{-1}N^TY$$

The predictor $\hat{f}(x)=\sum_{j=1}^{\hat{}} N_j(x)\hat{\theta}_j$ is fitted as smoothing spline

(b). Please comment on the relationship between the smoothing spline estimator and the natural cubic spline estimator.

A smoothing spline estimator finds a function among the set of functions where both the first and second derivatives are continuous. The goal is to minimize the equation

$$RSS(f,\lambda) = \sum_{i=1}^{n} \{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt$$

smoothing parameter

The function that minimizes the above equation is in fact a natural cubic spline.

(c). Write down the definition of degree of freedom for smoothing splines

Let $\hat{f}$ denote the n-vector of fitted values $\hat{f}(x_i)$ at the training predictors x_i.

$$\hat{f} = N(N^T N + \lambda \Omega_N)^{-1} N^T y = S_\lambda y.$$

Suppose B_ξ is a matrix with n rows and m columns.
Each column represents a cubic spline basis function evaluated at training point x_i. The knot sequence is denoted by ξ.
In our situation, m is strictly smaller than n.
The vector of fitted spline values is governed by the equation $\hat{f} = B_\xi (B_\xi^T B_\xi)^{-1} B_\xi^T y$ which is equal to $H_\xi y$.

H_ξ is a projection matrix. Remarkably, M is equivalent to the trace of our matrix H_ξ. Hence, the effective degree of freedom df_λ is equal to the trace of S_λ.

(d) Comment on why smoothing splines do not suffer from overfitting

Since $f(x_i)$ is never equal to y_i for the smoothing spline

$\sum_{i=1}^{N} \{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2$, the curve is forced to omit many observations. The penalty of $\lambda \int \{f''(t)\}^2$ keeps the variance low.