

Let $X = [1_n \ x]$ be the $n \times 2$ design matrix for the linear regression problem

$$y_i = \beta_0 + \hat{\beta}_1 x_i + r_i$$

Such that

$$\begin{aligned}\hat{y} &= X \hat{\beta} \\ &= X(X^T X)^{-1} X^T y \\ &\equiv H y\end{aligned}$$

Note the similarity with the smoothing spline fit, that are also linear in y :

$$\hat{f}^{ss} = H(N^T N + \lambda \Omega_N) N^T y \equiv S_N y \quad (\text{eq. 5.14})$$

and local linear regression

$$\hat{f}^{lr} = X(X^T W X)^{-1} X^T W y \equiv R_W y \quad (\text{eq. 6.18})$$

All three smoothers can be written on the form

$$\hat{f}(x_i) = \sum_{j=1}^N w_{ij} y_j$$

with weights w_{ij} equal to the (i,j) element in the relevant smoothing matrix (H , S_N or R_W).

We want to show that these linear smoothers preserve the linear part of the fit, that is

$$S_N y = S(\hat{y} + r) = \hat{y} + S_N r \quad (\Rightarrow) \quad S_N \hat{y} = \hat{y}$$

$$R_W y = R_W(\hat{y} + r) = \hat{y} + R_W r \quad (\Rightarrow) \quad R_W \hat{y} = \hat{y}$$

Smoothing splines

From chapter 5.4.1, we have the eigen decomposition of the smoother matrix

$$S_\lambda = U D_\lambda U^T = \sum_{j=1}^N p_j(\lambda) u_j u_j^T$$

where $p_j(\lambda)$ is the j 'th eigenvalue of S and u_j the associated eigenvector.

Then

$$S_\lambda = \sum_{j=1}^N p_j(\lambda) u_j u_j^T$$

Use $p_1(\lambda) = p_2(\lambda) = 1$ for all λ , then

$$S_\lambda = H + \sum_{j=3}^N p_j(\lambda) u_j u_j^T$$

where

$$H = u_1 u_1^T + u_2 u_2^T$$

seen using either the SVD of X in ch. 3.4.1 or that $S_\lambda \rightarrow H$ as $\lambda \rightarrow \infty$ (and $p_j(\lambda) \rightarrow 0, j > 2$).

$\hookrightarrow$ u_1 and u_2 span the column space of X and represent the linear part of the spline fit. The linear part is not shrunk / penalized. As $\lambda \rightarrow \infty$, we increasingly penalize curvature in the smoothing spline fit. The fit then approaches the least square solution.

Hence

$$S_\lambda y = H y + \sum_{j=3}^N p_j(\lambda) u_j u_j^T y$$

$$\begin{aligned} S_\lambda \hat{y} &= S_\lambda H y \\ &= \underbrace{H H}_{H} y + \sum_{j=3}^N \sum_{k=1}^2 p_j(\lambda) u_j u_j^T \underbrace{u_k u_k^T}_{=0} y \\ &= \hat{y} \end{aligned}$$

and

$$\begin{aligned} S r &= S (1 - H) y = S y - H y \\ &= \sum_{j=3}^N p_j(\lambda) u_j u_j^T y \end{aligned}$$

Local linear regression:

Apply the smoother R_W to $\hat{\gamma}$, we get directly

$$\begin{aligned}\hat{R_W \gamma} &= R_W \hat{\gamma} \\ &= X (X^T W X)^{-1} X^T W X (X^T X)^{-1} X^T \gamma \\ &= H \gamma\end{aligned}$$

Hence

$$R_W \gamma = \hat{\gamma} + R_W r$$

Backfitting implications:

An implication of the linearity presentation is:
that the mean of the linear
smoothing fit to a response is preferred.

Take the partial residuals in the backfitting algo (algo 9.1)

$$\hat{r}_{-j} = y - \alpha - \sum_{k \neq j} \hat{f}_k$$

These are mean zero.

When updating the j 'th function $f_j(x)$ estimate,
we apply a linear smoother S_j to $\hat{r}_{-j}$:

$$\hat{f}_j = S_j \hat{r}_{-j}$$

The mean value of this fit is zero.
The adjustment

$$\hat{f}_j = \bar{f}_j - \frac{1}{n} \sum_i \hat{f}_{ij}$$

does therefore not change the fit and is not necessary, formally. It is still recommended to include such adjustment step, since computational errors might cause the mean values to not be zero exactly.

