Lecture 6: Smoothing Splines

MATH5824 Generalised Linear and Additive Models

Richard P Mann

MATH5824 Generalised Linear and Additive Models

Reading

Course notes: Chapter 4, Sections 4.1–4.4

www.richardpmann.com/MATH5824

From Interpolation to Smoothing

Interpolation assumes: observations are exact, $y_i = f(t_i)$.

Reality: observations contain noise,

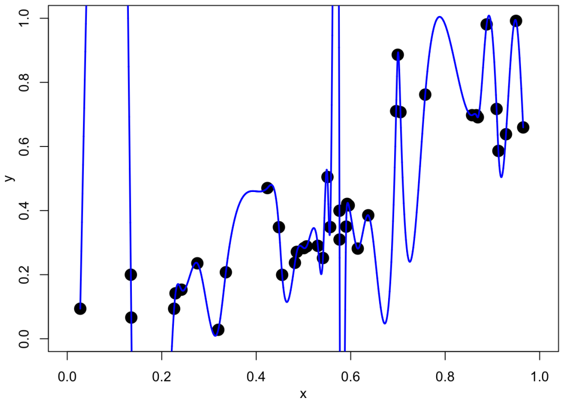
$$y_i = f(t_i) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Problem with interpolation:

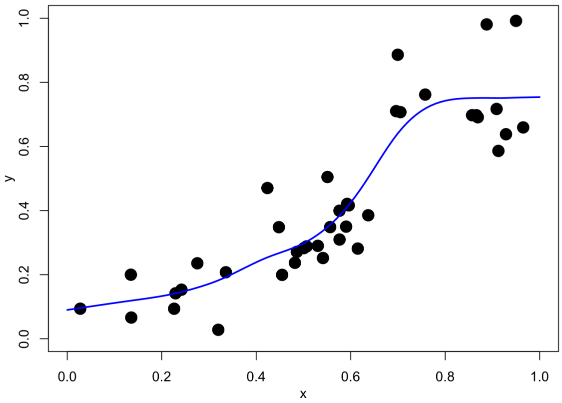
- Fitting every noisy data point exactly produces an overly wiggly curve
- The interpolant tracks noise rather than the underlying signal
- Poor predictions at new locations

$\Rightarrow$ We need to balance **fit to data** against **smoothness**.

Interpolating vs. Smoothing on Noisy Data



(a) Interpolating spline



(b) Smoothing spline

The interpolating spline passes through every point but is overly wiggly. The smoothing spline produces a more realistic estimate of the underlying function.

The Log-Likelihood Perspective

The log-likelihood for the model $y_i = f(t_i) + \epsilon_i$ is:

$$\ell(f; \mathbf{y}) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - f(t_i))^2 - n \log \sigma$$

Maximising ℓ without constraints $\Rightarrow$ interpolation ($f(t_i) = y_i$).

This is equivalent to minimising the residual sum of squares:

$$\mathrm{RSS}(f) = \sum_{i=1}^{n} (y_i - f(t_i))^2$$

Unconstrained minimisation always overfits.

The Penalised Least-Squares Criterion

Idea: Add a roughness penalty to the RSS.

Minimise the **penalised sum of squared residuals**:

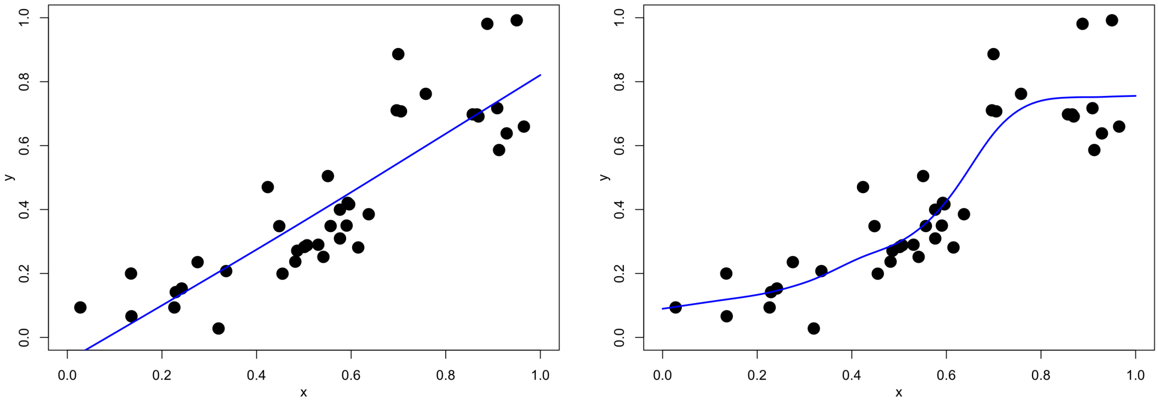
$$R_\nu(f,\lambda) = \underbrace{\sum_{i=1}^{n}(y_i - f(t_i))^2}_{\text{goodness of fit}} + \underbrace{\lambda\, J_\nu(f)}_{\text{roughness penalty}}$$

where $J_\nu(f) = \int [f^{(\nu)}(t)]^2\, dt$ and $\lambda \geq 0$ is the **smoothing parameter**.

Trade-off controlled by λ:

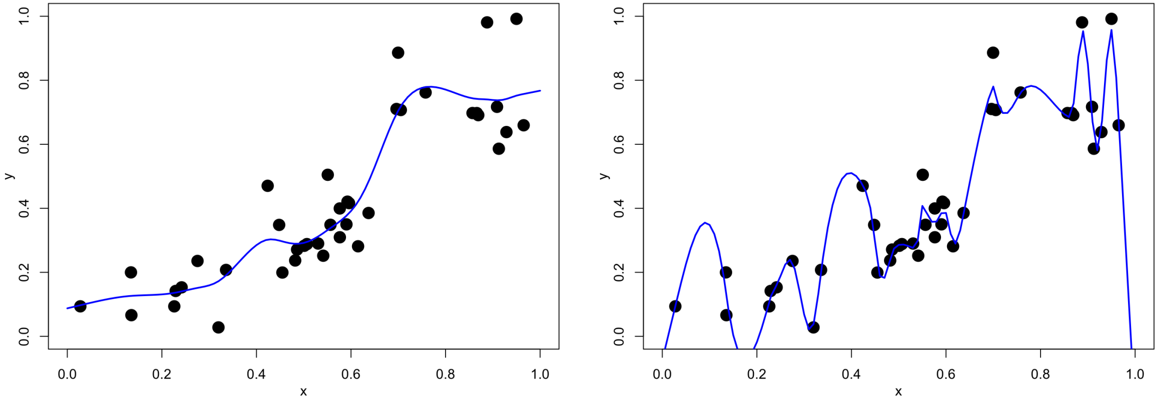
- $\lambda \to 0$: no penalty $\Rightarrow$ interpolating spline (overfit)
- $\lambda \to \infty$: maximal penalty $\Rightarrow$ polynomial regression (underfit)

Effect of λ: Strong and Moderate Smoothing



(a) Very large λ — essentially linear　　(b) Moderate λ — reasonable fit

Effect of λ: Weak and Very Weak Smoothing



(c) Smaller λ — more flexible

(d) Very small λ — near interpolation

Limiting cases ($\nu = 2$, cubic smoothing spline):

- $\lambda \to 0$: natural cubic interpolating spline
- $\lambda \to \infty$: ordinary least squares line ($f(t) = \hat{a}_0 + \hat{a}_1 t$)

The Solution is a Natural Spline

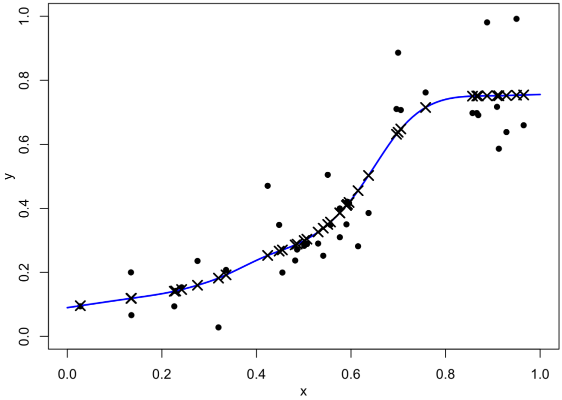
Proposition (4.1)

If $\hat{f}$ minimises $R_\nu(f, \lambda)$ with fitted values $\hat{y}_i = \hat{f}(t_i)$, then $\hat{f}$ is a pth-order natural spline with $p = 2\nu - 1$ and knots at the data locations.

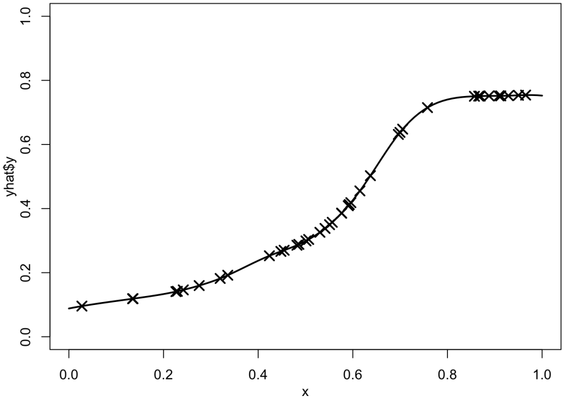
Proof idea: Suppose $\hat{f}^*$ interpolates the same fitted values $(\hat{y}_i)$ but has smaller roughness $J_\nu(\hat{f}^*) < J_\nu(\hat{f})$. Then $R_\nu(\hat{f}^*, \lambda) < R_\nu(\hat{f}, \lambda)$, contradicting optimality.

Consequence: We only need to search over natural splines, which have a finite-dimensional parameterisation.

Proposition 4.1 Illustrated



(a) Smoothing spline: data (•) and fitted values (×)



(b) Interpolating spline of the fitted values

The smoothing spline is identical to the interpolating spline of its own fitted values — confirming it is a natural spline with knots at the data locations.

Matrix Formulation

Using the natural spline representation (cubic case, $\nu = 2$):

$$\hat{f}(t) = a_0 + a_1 t + \sum_{i=1}^{n} b_i |t - t_i|^3, \qquad \sum b_i = 0, \quad \sum b_i t_i = 0$$

Fitted values in matrix form:

$$\hat{\mathbf{f}} = \mathbf{K}\mathbf{b} + \mathbf{L}\mathbf{a}$$

where:

- $\mathbf{K}$ is $n \times n$ with (i, k)th element $|t_i - t_k|^3$
- $\mathbf{L}$ is $n \times 2$ with columns $\mathbf{1}$ and $(t_1, \ldots, t_n)'$
- $\mathbf{b} = (b_1, \ldots, b_n)'$ and $\mathbf{a} = (a_0, a_1)'$

Proposition (4.2)

The roughness penalty can be written as a quadratic form:

Linear $(\nu = 1)$: $J_1(f) = -2\,\mathbf{b}'\mathbf{K}_1\mathbf{b}$

Cubic $(\nu = 2)$: $J_2(f) = 12\,\mathbf{b}'\mathbf{K}_2\mathbf{b}$

The penalised criterion becomes:

$$R_\nu(f, \lambda) = (\mathbf{y} - \mathbf{Kb} - \mathbf{La})'(\mathbf{y} - \mathbf{Kb} - \mathbf{La}) + \lambda^*\,\mathbf{b}'\mathbf{Kb}$$

subject to $\mathbf{L}'\mathbf{b} = \mathbf{0}$, where $\lambda^* = c_\nu \lambda$.

The Solution

Proposition (4.3)

The penalised least-squares solution is:

$$\begin{pmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{L}' \\ \mathbf{L} & \mathbf{K} + \lambda^* \mathbf{I}_n \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{0} \\ \mathbf{y} \end{pmatrix}$$

Observations:

- When $\lambda^* = 0$: reduces to interpolation ($\hat{f}(t_i) = y_i$)
- When $\lambda^* \to \infty$: $\hat{\mathbf{b}} \to \mathbf{0}$, leaving only $\hat{f}(t) = \hat{a}_0 + \hat{a}_1 t$
- The matrix equation simultaneously enforces the natural spline constraints and minimises the penalised criterion

Summary

Key points:

- Smoothing splines balance fit to data against roughness via λ
- The penalised criterion is $R_\nu(f, \lambda) = \mathrm{RSS} + \lambda\, J_\nu(f)$
- $\lambda = 0$ gives interpolation; $\lambda \to \infty$ gives polynomial regression
- The minimiser is always a natural spline with knots at data locations
- The solution is obtained by solving a linear system involving $\mathbf{K}$, $\mathbf{L}$, and λ

Next lecture: Fitting smoothing splines in R and choosing the smoothing parameter.