MAP estimation for curve modeling with free-knot splines.

Amate Laure, Laboratoire I3S, CNRS-UNSA
amate@i3s.unice.fr

Rendas Maria João, Laboratoire I3S, CNRS-UNSA
rendas@i3s.unice.fr

Abstract: In the context of curve modeling, splines have been widely used and studied ([1],[2],[3],[4]). Main arguments for using spline functions as function approximators are their ability to fit complex forms with arbitrary good accuracy, and the existence of basis functions, the B-splines, with attractive numerical properties. Our use of splines is motivated by the desire to find a parsimonious parametric description of a curve $s(t)$ of which we observe noisy samples:

$$Z_n = s(t_n) + \epsilon(t_n) \simeq f(t_n; \hat{\theta}(Z_1^N)) + \varepsilon(t_n), \qquad n = 1, \ldots, N, \qquad t_n \in [0, 1] \ .$$

In this equation, $f(t;\theta)$ is a spline function, and $\varepsilon(t)$ represents observation noise ($\epsilon(t)$) and modeling errors. The approximating spline is an element of

$$\mathcal{G} = \cup_{k \in \{k_{min}, k_{max}\}} \$_k^m \ ,$$

where, inspired by [1], $\$_k^m$ is the space of splines of degree m with k knots. Vector $\theta = (k, \xi, \beta)$ identifies $f(\cdot;\theta)$ as a member of $\mathcal{G}$: the number of knots, k, identifies $\$_k^m$, the knot vector ξ indicates that $f \in \$_{k,\xi}^m$, and β are the B-Spline.
We use a Bayesian approach and $\hat{\theta}(Z_1^N)$ are the MAP estimates in $\mathcal{G}$ for a postulated prior. While MAP estimation of $s(t)$ has been addressed by other authors, leading to a model $\hat{s}(t) \notin \mathcal{G}$, the problem of building a parametric model for s in $\mathcal{G}$ has received much less attention. An exception is [5], where in a slightly different context, MAP estimation in "union models" is also addressed. We depart in a significant way from the work of these authors, in the crucial issue of defining MAP estimates for spaces with this composite structure. A careful interpretation of the notation $p(\theta|Z_1^N)$ – that we prefer to decompose as $p(\xi, \beta|k, Z_1^N)P(k|Z_1^N)$ – is required: for different values of k, densities $p(\xi, \beta|k, Z)$ are defined with respect to distinct measures, and their direct comparison is meaningless, invalidating direct maximization of the numbers $p(\theta|Z_1^N)$.
We advocate that MAP estimation in models with the union structure of $\mathcal{G}$ must proceed in a stepwise manner:

$$\begin{tikzcd} & \hat{k}_{MAP} \arrow[dr] & \\ Z_1^N \arrow[ur] \arrow[rr] & & (\hat{\xi}, \hat{\beta})_{MAP} \end{tikzcd}$$

Our implementation is inspired of BARS [4], using MC methods to sample from the joint posterior on (k, ξ). By marginalizing over ξ, we approximate $P(k|Z_1^N)$. Optimization is based on a simulated annealing chain over ξ for k fixed at $\hat{k}_{MAP}$ ($\hat{\beta}_{MAP}$ is determined analytically). We note that we also estimate the noise variance. In the experiments presented, the approximation accuracy (MSE) of our MAP model $f(t; \hat{\theta}_{MAP})$ is comparable to that of BARS's estimates $\hat{s}_{BARS}$, with the advantage of providing a compact data description with $2\hat{k}_{MAP} + 1$ parameters.

References

[1] C. DeBoor, *A Practical Guide to Splines.* Springer, 1978.

[2] Z. Luo and G. Wahba, "Hybrid adaptive splines," *Journal of the American Statistical Association*, vol. 92, no. 437, pp. 107–116, 1997.

[3] M. J. Lindstrom, “Penalized estimation of free-knot splines,” *Journal of Computational and Graphical Statistics*, vol. 8, no. 2, pp. 333–352, 1999.

[4] I. DiMatteo, C. Genovese, and R. Kass, “Bayesian curve-fitting with free-knot splines,” *Biometrika*, vol. 88, pp. 1055–1071, 2001.

[5] C. Andrieu, N. D. Freitas, and A. Doucet, “Robust full bayesian learning for radial basis networks,” *Neural Comput.*, vol. 13, no. 10, pp. 2359–2407, 2001.