Splines

MATH-412 - Statistical Machine Learning

Polynomial regression is linear w.r.t. parameters

In 1 dimension :

$$f(x) = w_0 + w_1 x + w_2 x^2 + \ldots + w_d x^d$$

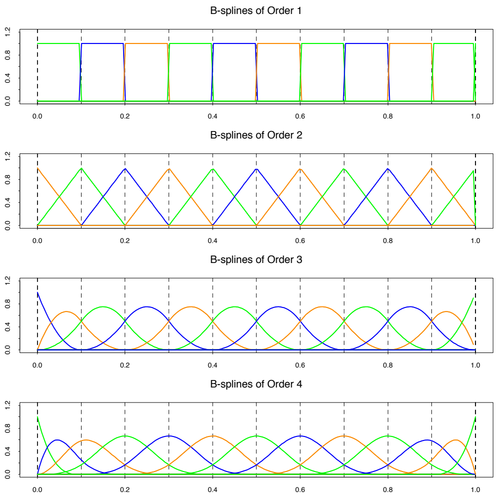
In p dimensions :

$$f(\mathbf{x}) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \sum_{i \leq j} \beta_{ij} x_i x_j + \sum_{i \leq j \leq k} \beta_{ijk} x_i x_j x_k \ldots$$

But... global polynomial bases are not always adapted, because

- Changes affect the function globally
- Higher variability of high degree terms

B-splines : a nice basis for piecewise polynomial functions



Splines of order $d+1$ are piecewise polynomial of degree d.

B-splines recursion

$\xi_1 < \ldots < \xi_K$: “inner” knots

Let

τ_0	$\ldots$	τ_d	τ_{d+1}	$\ldots$	τ_{K+d}	τ_{K+d+1}	$\ldots$	τ_{K+2d+1}
ξ_0	$\ldots$	ξ_0	ξ_1	$\ldots$	ξ_K	ξ_{K+1}	$\ldots$	ξ_{K+1}

The j^{th} spline of degree d, $B_{j,d}$: is defined by

$$B_{j,0}(x) = 1_{\{\tau_j \leq x \leq \tau_{j+1}\}}$$

$$B_{j,d+1} = \omega_{j,d}\, B_{j,d} + (1 - \omega_{j+1,d})\, B_{j+1,d}$$

with $\omega_{j,d}(x) = \frac{x - \tau_j}{\tau_{j+d+1} - \tau_j}$ if $\tau_j < \tau_{j+d+1}$ and 0 else.

Cardinal B-splines

If all knots are equidistant (and also outside of the interval) then $B_{1,d} = B_{1,d-1} * B_{1,0}$

and $$B_{j+1,d}(x) = B_{1,d}(x - jh), \quad j \in \mathbb{Z}.$$

where h is the inter-knot distance.

Splines with an increasing number of continuous derivatives

We start with $x \in \mathbb{R}$, and consider predictors of the form

$$f(x) = \sum_{j=0}^{K+d} w_j \, B_{j,d}(x)$$

with the $B_{j,d}$ forming a basis of functions, and K the number of “inner knots”.

Splines of different orders :

- degree $d = 0$ (order 1)= piecewise constant functions
- degree $d = 1$ (order 2)= piecewise linear functions that are $\mathcal{C}^0$
- degree $d = 2$ (order 3)= piecewise quadratic functions that are $\mathcal{C}^1$
- degree $d = 3$ (order 4)= piecewise cubic functions that are $\mathcal{C}^2$

Natural cubic splines

It is often desirable to impose that the part of the predictor extrapolating outside of the data range is **linear**.

The set of cubic spline functions, with K "inner" knots, that are linear outside of the range of the data form a subspace of dimension K : the natural cubic splines.
For K "inner" knots, one possible basis for the subspace takes the form :

$$\nu_1(x) = 1, \quad \nu_2(x) = x, \quad \nu_{k+2}(x) = d_k(x) - d_{K-1}(x),$$

with

$$d_k(x) = \frac{(x - \xi_k)_+^3 - (x - \xi_K)_+^3}{\xi_K - \xi_k}, \quad 1 \leq k \leq K - 2.$$

In practice, it is equivalent to use B-splines and impose that the function is linear outside of the interval of interest, via constraints on the coefficients.

In practice : solving a regression problem with splines

Say with natural cubic splines.

$$f(x) = \sum_{j=1}^{K} w_j \nu_j(x)$$

Given a training set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$

- Build a corresponding design matrix

$$\mathbf{N} = \begin{bmatrix} \nu_1(x_1) & \ldots & \nu_K(x_1) \\ \nu_1(x_2) & \ldots & \nu_K(x_2) \\ \vdots & \vdots & \vdots \\ \nu_1(x_n) & \ldots & \nu_K(x_n) \end{bmatrix}$$

- Solve the linear regression problem

$$\min_{\boldsymbol{w}} \sum_{i=1}^{n} (y_i - \boldsymbol{\nu}(x_i)^\top \boldsymbol{w})^2$$

In practice : solving a regression problem with splines

The previous equation can be written vectorially

$$\min_{\boldsymbol{w}} \|\boldsymbol{y} - \mathbf{N}\boldsymbol{w}\|_2^2$$

It leads to the *normal equations*

$$\boxed{\mathbf{N}^\top\mathbf{N}\boldsymbol{w} - \mathbf{N}^\top\boldsymbol{y} = 0}$$

so that provided $\mathbf{N}^\top\mathbf{N}$ is invertible, we have

$$\hat{\boldsymbol{w}} = (\mathbf{N}^\top\mathbf{N})^{-1}\mathbf{N}^\top\boldsymbol{y}$$

Smoothing splines

Theorem : The solution of the optimization problem

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

is a *natural cubic spline function with knots at all the* x_i*s*. It is thus of the form

$$f(x) = \sum_{j=1}^{K} \hat{w}_j \, \nu_j(x) \qquad \text{with} \quad K = n,$$

with $\hat{\boldsymbol{w}}$ the solution of

$$\min_{\boldsymbol{w}} \|\boldsymbol{y} - \mathbf{N}\boldsymbol{w}\|_2^2 + \lambda \boldsymbol{w}^\top \boldsymbol{\Omega} \, \boldsymbol{w}, \qquad \text{with} \quad \Omega_{jk} = \int \nu_j''(t) \nu_k''(t) dt,$$

namely

$$\hat{\boldsymbol{w}} = (\mathbf{N}^\top \mathbf{N} + \lambda \boldsymbol{\Omega})^{-1} \mathbf{N}^\top \boldsymbol{y}.$$

Working with splines

Need to choose

- The degree d
  - Most of the time : $d = 0, 1$ or 3 (constant, linear or cubic splines).
- The number of "inner" knots K
  - Most of the time : K is chosen by cross-validation
  - Possible to choose $K = n$ with regularization $\Rightarrow$ *smoothing splines*
- The position of the knots
  - Most of the time : quantiles (=percentiles) of the data
  - or on a regular grid
- Whether to use *regression splines* vs *smoothing splines*

Computation efficiency of splines

- Since B-splines have small support, if $\mathbf{B}$ is the B-spline design matrix, $\mathbf{B}^\top\mathbf{B}$ is a banded matrix, with inversion linear in K.
- For smoothing splines, it is possible to solve the problem using the B-spline basis and $\hat{\boldsymbol{y}} = \mathbf{N}\hat{\boldsymbol{w}} = \mathbf{N}(\mathbf{N}^\top\mathbf{N} + \lambda\boldsymbol{\Omega})^{-1}\mathbf{N}^\top\boldsymbol{y}$ can thus be computed in $O(n)$ time.

Multivariate splines

Idea : Use *tensor product basis* :

For a function of (x_1, x_2) and given two spline bases :

$$\big(h_j(x_1)\big)_j \qquad \text{and} \qquad \big(\tilde{h}_k(x_2)\big)_k$$

use basis elements of the form

$$g_{jk}(x_1, x_2) = h_j(x_1)\tilde{h}_k(x_2).$$

Then find functions of the form :

$$f(x_1, x_2) = \sum_{j,k} w_{jk}\, g_{jk}(x_1, x_2).$$

Extensions

- The theorem characterizing the form of the solution of

$$\min_f \widehat{\mathcal{R}}_{n,2}(f) + \lambda \int f''(\mathbf{x})^2 d\mathbf{x},$$

  for $\widehat{\mathcal{R}}_{n,2}$ the empirical quadratic risk, can be extended to multiple dimensions with *thin-plate splines*, which are different than the tensor product splines.

- Possible to use splines as feature vectors for the input data for other problems than least square regression : the splines provide a particular form of *feature map*.