Spline regression

Mark Andrews
Psychology Department, Nottingham Trent University

✉ mark.andrews@ntu.ac.uk

Normal basis function regression

- Polynomial regression can be seen as a type of *basis function* regression.
- In general, in basis function regression where we have one predictor variable x, we model $f(x)$, which is the nonlinear function of x, as a linear sum of K simple functions of x known as basis functions

$$y_i \sim N(\mu_i, \sigma^2), \quad \mu_i = f(x_i) = \beta_0 + \sum_{k=1}^{K} \beta_k \phi_k(x_i), \quad \text{for } i \in 1 \ldots n.$$

- Here, $\phi_1(x_i), \phi_2(x_i) \ldots \phi_k(x_i) \ldots \phi_K(x_i)$ are simple deterministic functions of x_i.

Polynomial basis functions

- In polynomial regression, our basis functions are defined simply as follows:
$$\phi_k(x_i) \triangleq x_i^k.$$

Spline basis functions

- There are many different types of basis functions that are possible to use, but one particularly widely used class of basis functions are *spline* functions.
- The term *spline* is widely used in mathematics, engineering, and computer science and may refer to many different types of related functions, but in the present context, we are defining splines as piecewise polynomial functions that are designed in such a way that each piece or segment of the function joins to the next one without a discontinuity.
- As such, splines are smooth functions composed of multiple pieces, each of which is a polynomial.

Cubic b-splines

- There are many types of spline functions that can be used, but one of the most commonly used types is *cubic b-splines*.
- The *b* refers to *basis* and the *cubic* is the order of the polynomials that make up the pieces.
- Each cubic b-spline basis function is defined by 4 curve segments that join together smoothly.
- The breakpoints between the intervals on which these curves are defined are known as *knots*.
- If these knots are equally spaced apart, then we say that the spline is *uniform*.
- For basis function k, its knots can be stated as

$$t_0^k < t_1^k < t_2^k < t_3^k < t_4^k,$$

so that the curve segments are defined on the intervals $(t_0^k, t_1^k]$, $(t_1^k, t_2^k]$, $(t_2^k, t_3^k]$, (t_0^3, t_4^k).

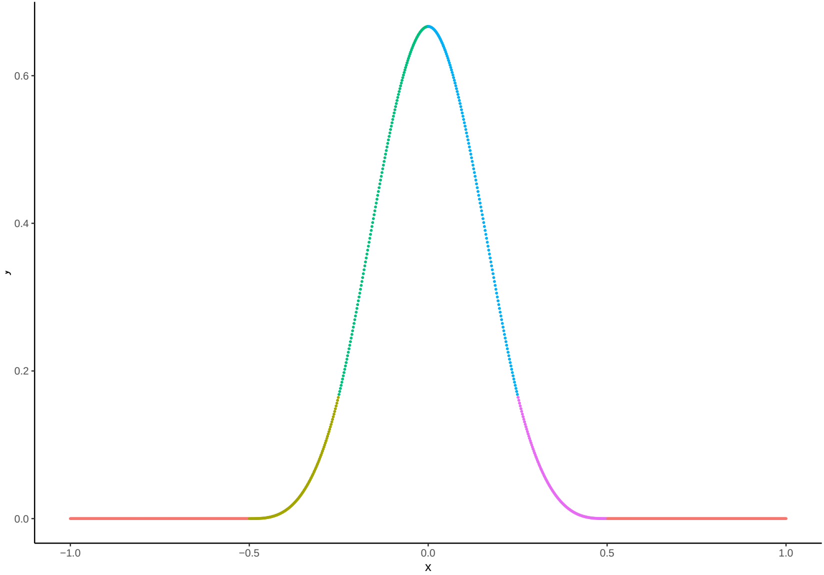
Cubic b-splines

The cubic b-spline is then defined as follows: {

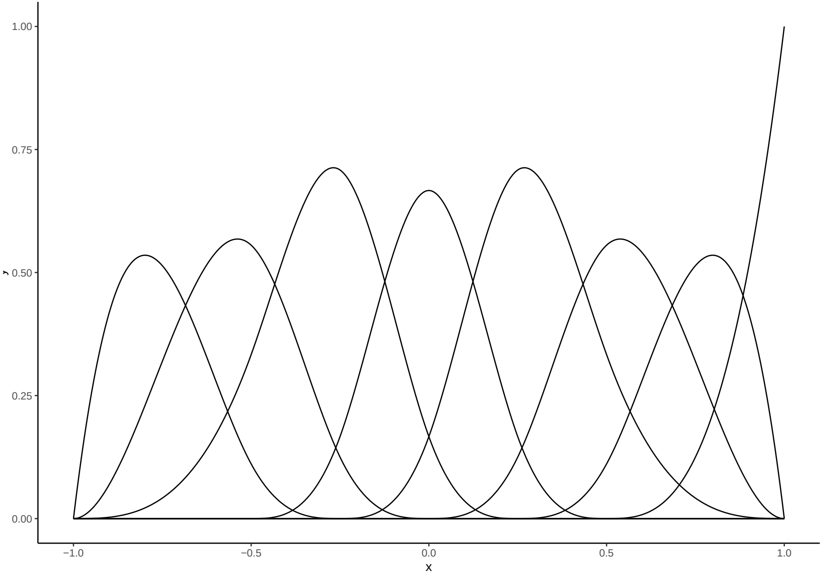
$$\phi_k(x_i) = \begin{cases} \frac{1}{6}u^3, & \text{if } x_i \in (t_0^k, t_1^k], \quad \text{with } u = (x_i - t_0^k)/(t_1^k - t_0^k) \\ \frac{1}{6}(1 + 3u + 3u^2 - 3u^3), & \text{if } x_i \in (t_1^k, t_2^k], \quad \text{with } u = (x_i - t_1^k)/(t_2^k - t_1^k) \\ \frac{1}{6}(4 - 6u^2 + 3u^3), & \text{if } x_i \in (t_2^k, t_3^k], \quad \text{with } u = (x_i - t_2^k)/(t_3^k - t_2^k) \\ \frac{1}{6}(1 - 3u + 3u^2 - u^3), & \text{if } x_i \in (t_3^k, t_4^k), \quad \text{with } u = (x_i - t_3^k)/(t_4^k - t_3^k) \\ 0 & \text{if } x_i < t_0^k \text{ or } x_i > t_4^k \end{cases}$$

}

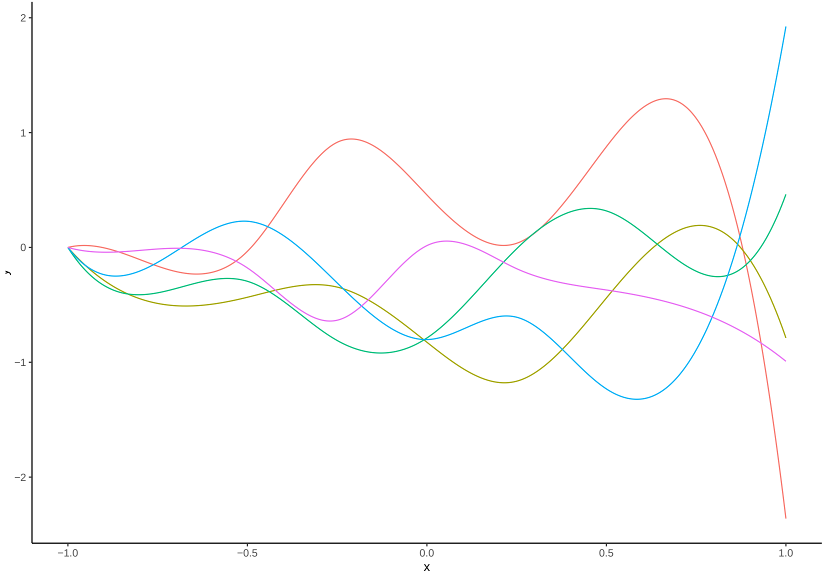
Cubic b-spline example



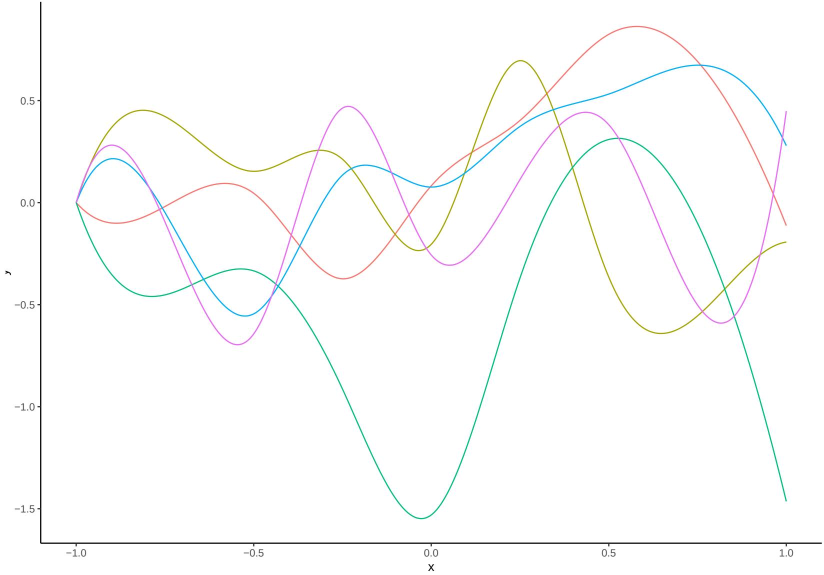
Uniformly spaced b-splines using `splines` package



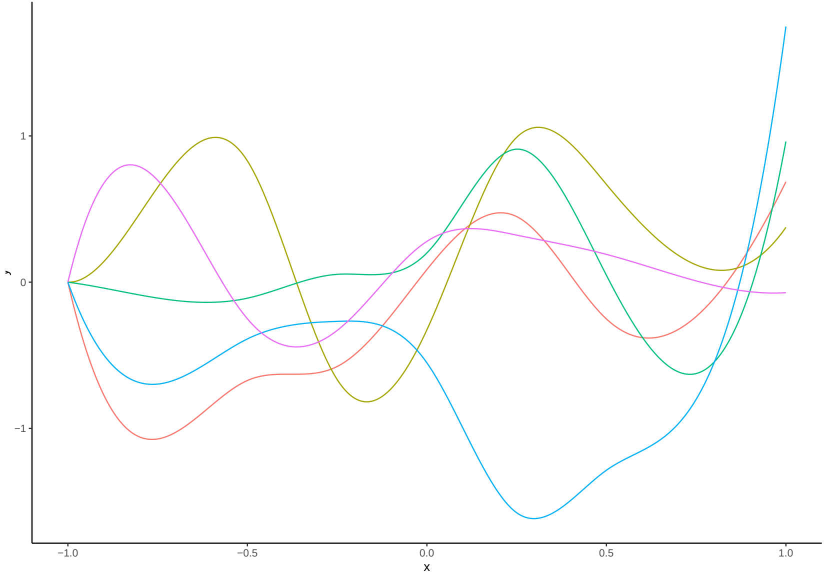
Weighted sums of b-splines: Example 1



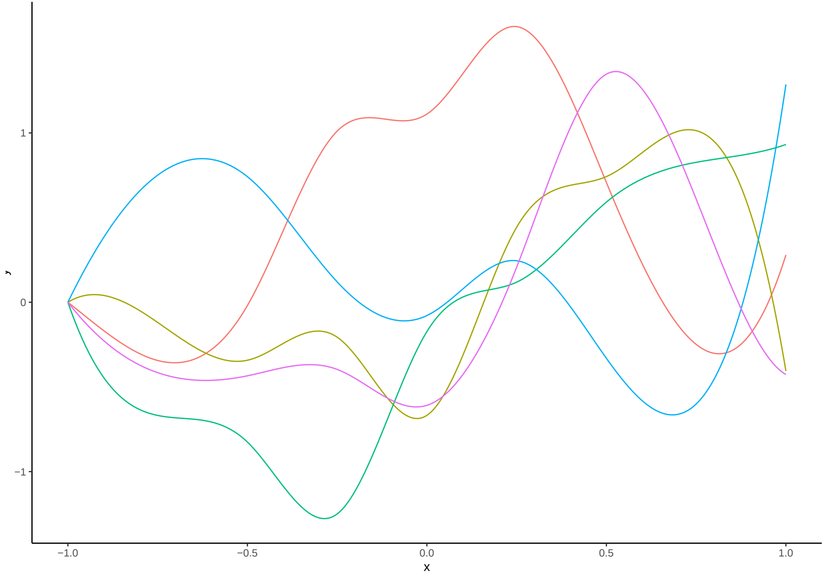
Weighted sums of b-splines: Example 2



Weighted sums of b-splines: Example 3



Weighted sums of b-splines: Example 4



An example

```r
library(splines)
knots <- seq(-500, 2500, by = 500)
M_bs <- lm(mean_fix ~ bs(Time, knots = knots)*Object,
           data=eyefix_df_avg)
```

Model predictions