Linear Splines

Consider a simple linear regression model. Letting x be the explanatory variable, the fitted model approximates the function, $E(Y|x)$, with a linear function having the form

$$b_0 + b_1 x.$$

If a linear approximation is not adequate, one could use polynomial regression to arrive at an approximation having the form

$$b_0 + b_1 x + b_2 x^2 + \cdots + b_p x^p.$$

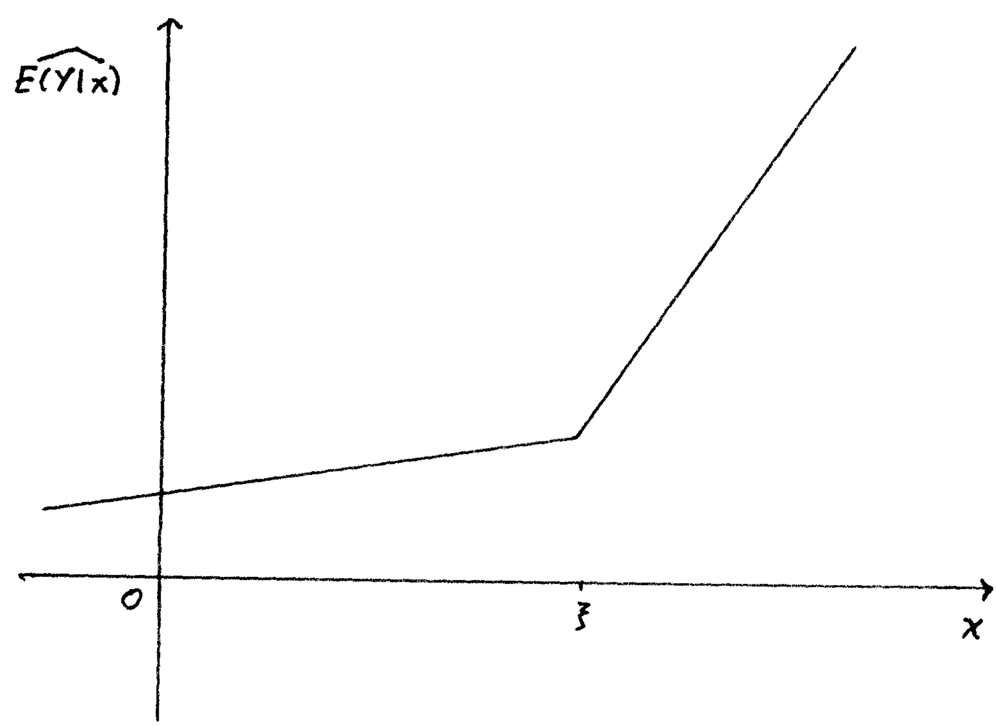
Alternatively, one could use a *continuous piecewise linear spline*. If we use a single preset *knot*, ξ, such a spline can be created by using OLS regression to fit a multiple regression model based on the basis functions

$$h_0(x) = 1, \quad h_1(x) = x, \quad \& \quad h_2(x) = (x-\xi)_+.$$

In practice, one would create a new variable h_2 (thinking of h_2 as a transformation of x), and arrive at a fit having the form

$$b_0 + b_1 x + b_2 h$$
$$= b_0 + b_1 x + b_2 (x-\xi)_+.$$

We have a piecewise linear function for which the slope is b_1 on $(-\infty, \xi)$ and the slope is $b_1 + b_2$ on (ξ, ∞). Furthermore, the approximating function is continuous at ξ, which is the location at which it can bend.



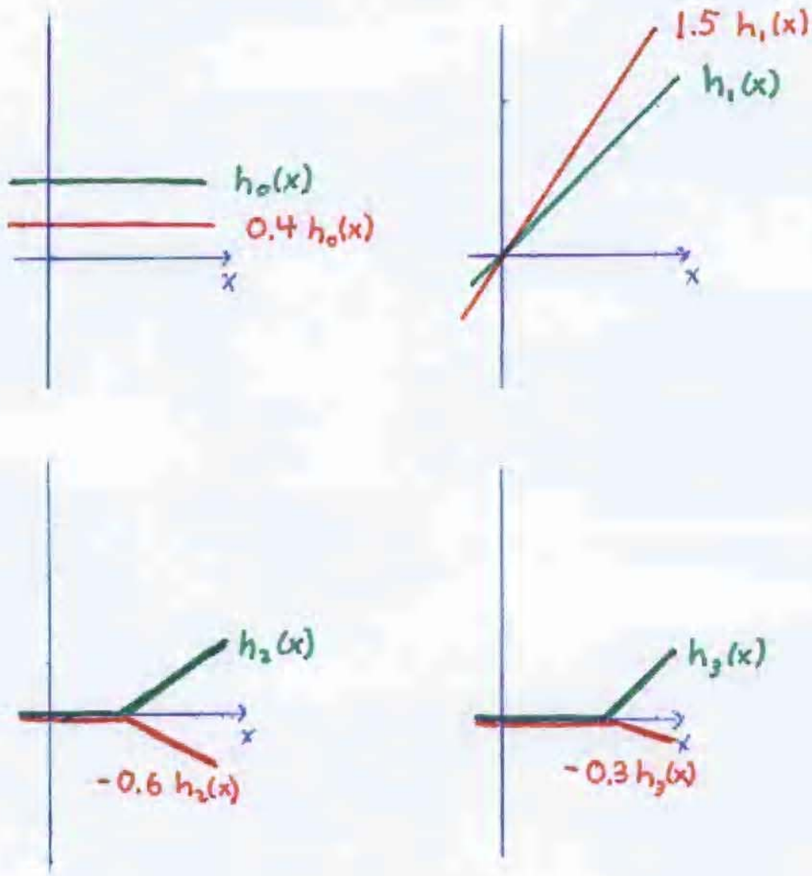
If we specify two knots, ξ_1 and ξ_2, and do a regression using the basis functions (predictors)

$$h_0(x) = 1,$$
$$h_1(x) = x,$$
$$h_2(x) = (x - \xi_1)_+,$$
&
$$h_3(x) = (x - \xi_2)_+,$$

we will obtain an approximating function having the form

$$b_0 + b_1 x + b_2 (x - \xi_1)_+ + b_3 (x - \xi_2)_+ .$$

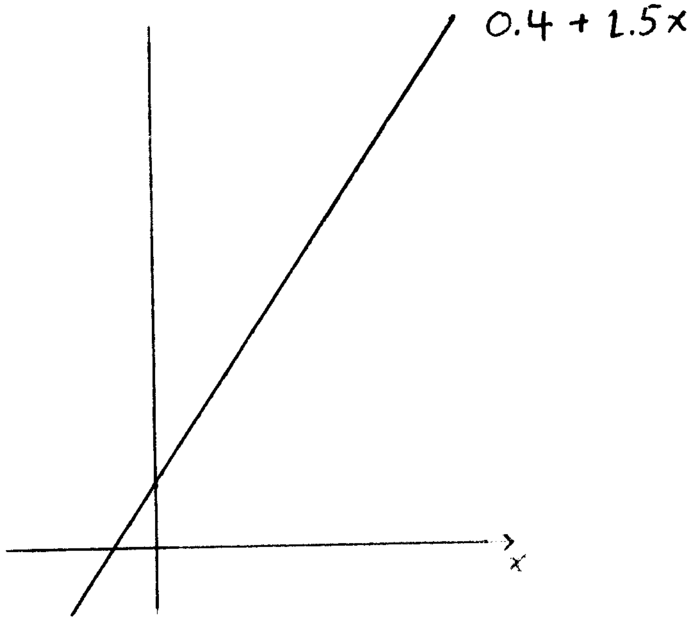
To better illustrate how the final approximating function is a linear combination of the basis functions, I will show $h_0(x)$, $h_1(x)$, $h_2(x)$, and $h_3(x)$ in green, and $b_0 h_0(x)$, $b_1 h_1(x)$, $b_2 h_2(x)$, and $b_3 h_3(x)$ in red, using $b_0 = 0.4$, $b_1 = 1.5$, $b_2 = -0.6$, and $b_3 = -0.3$.

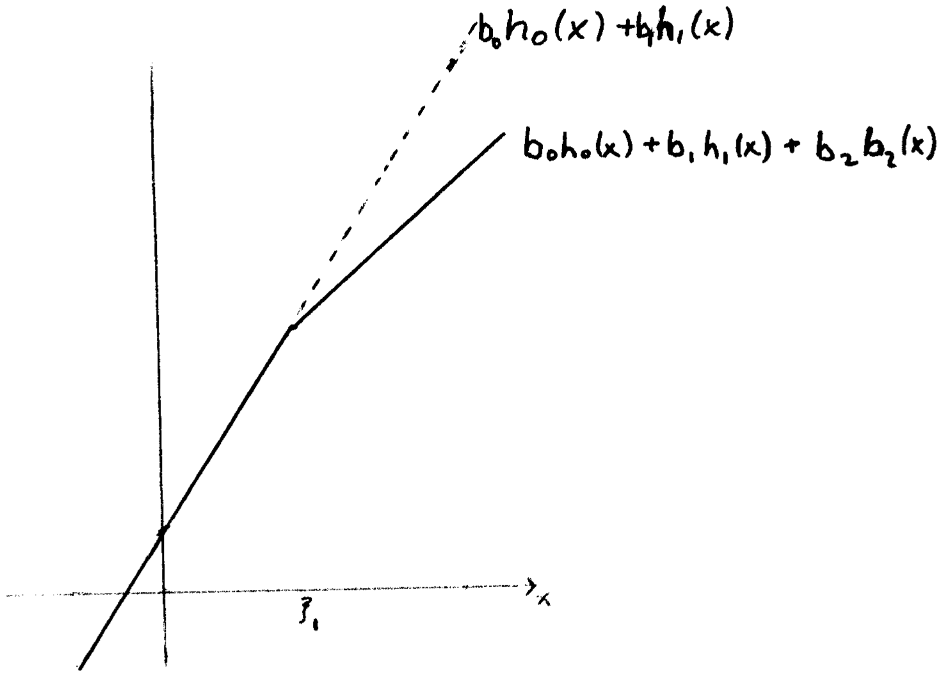


(The knots are $\xi_1 = 0.9$ and $\xi_2 = 1.4$.)

For the contributions from the first two basis functions we have

$$b_0 + b_1 h_1(x) = 0.4 + 1.5x.$$





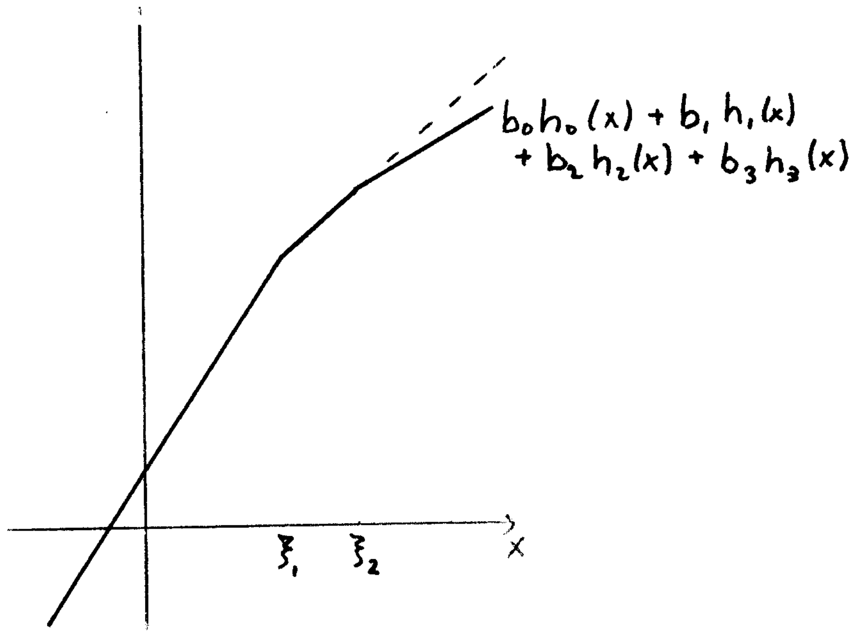
The above function results from adding $b_2 h_2(x)$.

To obtain the final function,

$$b_3 h_3(x) = -0.3\,(x-1.4)_+$$

$$= \begin{cases} 0.42 - 0.3x, & x > 1.4, \\ 0, & x \le 1.4, \end{cases}$$

is added.



Altogether, the final approximating function is

$$0.4 + 1.5x - 0.6(x-0.9)_+ - 0.3(x-1.4)_+$$

$$= \begin{cases} 1.36 + 0.6x, & x > 1.4, \\ 0.94 + 0.9x, & 0.9 < x \le 1.4, \\ 0.4 + 1.5x, & x \le 0.9. \end{cases}$$

To fit such a piecewise linear spline, one uses the x values to create values for h_2 and h_3, and then does multiple regression using x, h_2, and h_3. (It isn't very different from fitting a polynomial model using x, x^2, and x^3.)

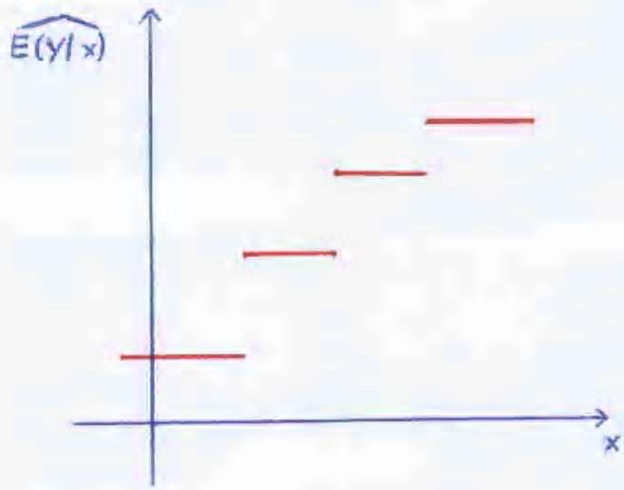
With the current example, to fit the spline model, four parameters have to be estimated using least squares. Four parameters also have to be estimated to fit a cubic polynomial model. For some underlying functions the spline model would be better, and for other functions it would be better to fit the cubic polynomial model. Other four parameter models, which are generally less desirable because the fitted models typically are not everywhere continuous, are a four piece piecewise constant function based on basis functions of the form

$$h_1(x) = I_{(-\infty, \xi_1]}(x),$$
$$h_2(x) = I_{(\xi_1, \xi_2]}(x),$$
$$h_3(x) = I_{(\xi_2, \xi_3]}(x),$$
&
$$h_4(x) = I_{(\xi_3, \infty)}(x),$$

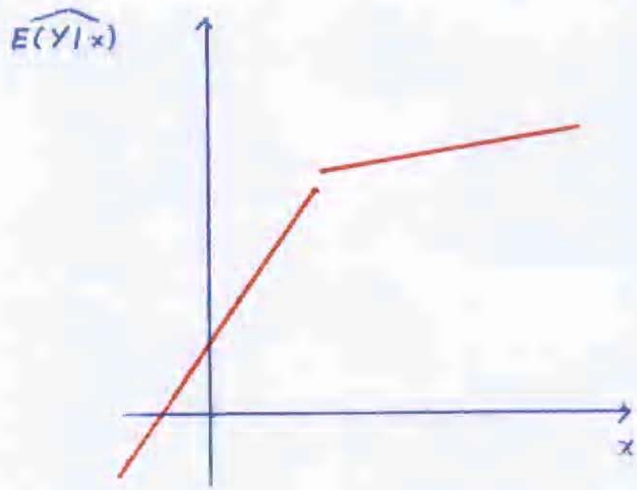
and a two piece piecewise linear function, where cases with x values in $(-\infty, \xi]$ are used to fit a linear function on $(-\infty, \xi]$, and cases with x values in (ξ, ∞) are used to fit a linear function on (ξ, ∞), resulting in a fitted approximation having the form

$$\widehat{E(Y \mid x)} = \begin{cases} b_3 + b_4 x, & x > \xi, \\ b_1 + b_2 x, & x \le \xi, \end{cases}$$

which need not be continuous.



a fit based on the method from top half of p. 10



a fit based on the method from bottom half of p. 10