Stat 504: Handout 1

Example: Spline regression. In almost all real world problems a linear model provides at best an approximation of the relationship between the dependent variable and one or more independent variables. Such an approximation of the true relationship can be very helpful in understanding features of the problem being studied. For example, we may be able to get an idea how changes in one or more of the independent variables will affect the dependent variable.

As a simple example consider a polynomial regression model in one variable,

$$Y_i = \beta_1 + \beta_2 x_i + \ldots + \beta_k x_i^{k-1} + \epsilon_i, \quad i = 1, \ldots, n.$$

It is unlikely that the true relationship between Y and x is exactly a polynomial relationship. But polynomials can be used to approximate just about any smooth function. For some problems a good approximation would however require a large degree (i.e., value of k) of the polynomial. One way to avoid polynomials of high degree, and to make the model possibly more parsimonious, is by using spline functions.

For selected points $w_1 < \ldots < w_h$ of possible x-values, to be called *knots*, a *spline function* $s(x)$ is defined as a function of the form

$$s(x) = \begin{cases} s_0(x) & \text{if } x < w_1 \\ s_i(x) & \text{if } w_i \le x < w_{i+1}, \quad i = 1, \ldots, h-1 \\ s_h(x) & \text{if } x \ge w_h, \end{cases}$$

where

$$s_i(x) = \beta_{1i} + \beta_{2i} x + \ldots + \beta_{ki} x^{k-1}, \quad i = 0, 1, \ldots, h.$$

Thus, $s(x)$ is obtained by patching together $h+1$ different polynomials defined on nonoverlapping intervals. Note that we are now dealing with $(h+1)k$ unknown $\underline{\beta}$'s. With a good choice of the knots, we may however be able to use a low degree for the $h+1$ polynomials that form the spline function and still obtain a better approximation of the true relationship between Y and x than with one polynomial of high degree.

We will assume that the knots are known; they may, for example, correspond to natural points where a change in the relationship can be expected. If x denotes time, for example, the beginning or ending of a war or the date of a presidential election could be natural knots. (If knots are not known, we could treat them – and possibly the number of knots h as well – as unknown model parameters. Such a model can however no longer be expressed as a linear model.)

Typically, we would like the spline function to be smooth at the knots. The smoothness conditions that are often used are

$$\left.\frac{d^j s_i(x)}{dx^j}\right|_{x=w_i} = \left.\frac{d^j s_{i-1}(x)}{dx^j}\right|_{x=w_i}$$

for $j = 0, 1, \ldots, k-2$, $i = 1, \ldots, h$. (The expression at the left hand side of this equation denotes the j-th derivative of $s_i(x)$ with respect to x evaluated at $x = w_i$.) Thus, for $j = 0$ the smoothness condition simply requires continuity at the knots. For $j = 1$ the requirement is differentiability at the knots; the conditions require higher order differentiability at the knots for $j \geq 2$.

If all s_i's are of degree 1 (i.e., $k = 2$), the spline function is known as a linear spline. Quadratic and cubic splines are spline functions with polynomials of degree 2 and 3, respectively. Cubic splines are especially popular: while still relatively simple, with a good choice of knots they can provide good approximations to relatively complicated functions.

If $\underline{\beta} = (\beta_{10}, \ldots, \beta_{k0}, \ldots, \beta_{1h}, \ldots, \beta_{kh})^T$ denotes the vector of parameters of the spline function, the model may be written as

$$\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\epsilon},$$

where a row of $\mathbf{X}$ looks like

$$(0, \ldots, 0, 1, x, \ldots, x^{k-1}, 0, \ldots, 0).$$

The nonzero elements in this row correspond to the parameters $\beta_{1i}, \ldots, \beta_{ki}$, where the value of i depends on the location of x relative to the knots $w_1, \ldots, w_h$. Note again that this formulation of the model as a linear model requires that we know the knots.

The smoothness conditions are linear in the parameters. For example, for $j = 0$, the continuity condition at knot w_1 requires that

$$\beta_{10} + \beta_{20}w_1 + \ldots + \beta_{k0}w_1^{k-1} = \beta_{11} + \beta_{21}w_1 + \ldots + \beta_{k1}w_1^{k-1},$$

or

$$\underline{\lambda}^T \underline{\beta} = 0,$$

where

$$\underline{\lambda} = (1, w_1, \ldots, w_1^{k-1}, -1, -w_1, \ldots, -w_1^{k-1}, 0, \ldots, 0)^T.$$

Thus, the entire collection of smoothness conditions may be written as

$$\mathbf{A}\underline{\beta} = \underline{0},$$

where $\mathbf{A}$ is a $h(k-1) \times (h+1)k$ matrix. In other words, the model is the restricted model

$$\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\epsilon}, \;\; \mathbf{A}\underline{\beta} = \underline{0}.$$

Spline functions provide therefore an example where a restricted linear model occurs quite naturally.