Polynomial Regression

To fit a polynomial regression, with the order of the polynomial being d, we add $d-1$ columns to the design matrix $\mathbf{x}$, (e.g., $x_1^2, x_1^3, \ldots x_1^d$), and treat them just as we would any other predictor variables.

For example,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \ldots \beta_d X_1^d + \beta_{d+1} X_2 + \ldots \beta_{p+d-1} X_p + \epsilon$$

We are treating X_1 and X_1^2 (and X_1^3, etc.) as distinct predictor variables.

Smoothness

Polynomial functions vary continuously in all their arguments. In fact, they are "smooth" in the sense that all their derivatives exist and are continuous. This is desirable if you think the real regression function you are trying to model is smooth, but not if you think there are sharp thresholds or jumps. Polynomials can approximate thresholds arbitrarily closely, but you end up needing a very high order polynomial.

Interpretation

In a linear model, we were able to offer simple interpretations of the coefficients, in terms of slopes of the regression surface. But in polynomial regression there is not one answer to "what is the rate of change?".

The change associated with a one-unit change in X_1 is not equal to any of the coefficients.

Rather than trying to give one single rate of change (or slope or response-associated-to-a-one-unit-change) when none exists, make a plot of the polynomial itself.

Over-fitting and wiggliness

A polynomial of degree d can exactly fit any d points. Any two points lie on a line, any three on a parabola, etc. Using a high-order polynomial, or even summing a large number of low-order polynomials, can therefore lead to curves which come very close to the data used to estimate them, but predict very poorly. In particular, high-order polynomials can display very wild oscillations in between the data points. Plotting the function is a good way of detecting this.

Choosing the polynomial order

The best way to choose the polynomial order is on the basis of some actual scientific theory which says that the relationship should, indeed, be a polynomial of order d. Failing that, carefully examining the diagnostic plots is the next best bet.

Diagnostic plots

The appropriate diagnostic plot is of residuals against the predictor. There is no need to make separate plots of residuals against each power of the predictor.

Orthogonal polynomials

I have written out the polynomial regression above in its most readily comprehended form, but that is not always the best way to estimate it. We will get smaller standard errors when predictor variables are uncorrelated. While X_i and its higher powers are linearly independent, they are generally (for most distributions) somewhat correlated. An alternative to regressing on the powers of X_i is to regress on a linear function of X_i, a quadratic function of X_i, a cubic, etc., which are chosen so that they are *un*correlated. These functions, being uncorrelated, are called **orthogonal**.

Any polynomial could also be expressed as a linear combination of these **basis functions**, which are thus called **orthogonal polynomials**. The advantage, again, is that the coefficients estimates of these basis functions have less variance than using the powers of X_i.