Polynomial, Spline, and LOESS Regression

Dr Austin R Brown

Kennesaw State University

Introduction

Some of these materials have been adapted from materials created by Dr Taasoobshirazi as well as my former professor, Dr Khalil Shaife (thanks Drs S & T!)

Introduction

- Throughout this class, we've learned that one of the primary goals of regression analysis is to help us better understand why some quantitative, measurable phenomenon varies.
  - Why do people's LDL cholesterol levels vary?

- We can think of this both in a very pragmatic sense (e.g., diet and exercise are likely reasons why LDL levels vary) as well as a more technical sense (e.g., we want to identify those variables which give us ample model fit).

Introduction

- With respect to the latter purpose of regression analysis, we've looked at lots of various ways to reasonably verify model fit, including outlier diagnostics and assumption checking.
  - So far, we've learned that data transformation and ridge regression may ways to potentially alleviate such issues.

- As you recall, the reason why we performed transformations of the data was to solve any combination of issues involving the assumptions of linearity, normality, or homoskedacity.

- Interestingly, we have another tool available to us to help solve these problems within a linear regression framework called **polynomial regression**.
  - This is particular useful when we have non-linear relationships between our predictors and response.

Polynomial Models in One Variable

- So what is a polynomial regression model? Suppose we have a single predictor, X, for our response, Y. A polynomial regression model would involve squaring the predictor variable to create a new predictor variable, and then including both the squared and original predictors in a MLR:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \varepsilon_i$$

- This type of polynomial regression is referred to as a "second order polynomial in one variable."

Polynomial Models in One Variable

- Suppose we had two predictors now, x_1 and x_2. What would a second order polynomial in two variables look like?

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i}^2 + \beta_4 x_{2i}^2 + \beta_5 x_{1i} x_{2i} + \varepsilon_i$$

- We can generalize this out to k predictors, as well.

Polynomial Models in One Variable

- Okay, let's stick with the one variable model for now. We can have as high of an order model as we want, but in my experience, a second or third order model is about this highest I've seen it go.

- Even though we're manipulating one of our variables, nothing is really different at all in terms of the estimation or assumptions we have in a regular MLR model. It's just that now, the mean that we're fitting is:

$$E[Y|X] = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2$$

Polynomial Models in One Variable

- This can be a cool way around using a nastier type of transformation and is for sure easier to interpret (at least in lower order models). There are, however, some things we need to keep in mind:

1. We want to keep the order of the model as low as possible. Montgomery argues in favor of transformation as a first resort to keep the model as single order. I generally agree that simplicity is better than complexity (a concept called *parsimony*), but I think a transformation such as Box-Cox is far more complex than a second order model.
   - Note, while we didn't harp on this too much back in the introduction to regression concepts, if we have n observations, this means we can have a maximum of $n-1$ β's (the number of observations must be greater than the number of β's we're estimating). In terms of polynomial regression, if we have a single variable, that means our maximum order is $n-1$. Practically, we'd never do this for the same reason I recommend avoiding transformation whenever possible.

Polynomial Models in One Variable

2. When deciding on which order model is most appropriate, we have to general approaches: a forward selection method and a backward selection method.

- With a forward selection method, we start with a single order model and work our way up until we've arrived at an ample model. With backward selection, we start with the highest order model and work our way back.

- Note, use of either approach on the same data typically won't result in the same model. In general, I'd say the forward approach is better as it will typically lead to a more parsimonious model, and it's almost universally advised to stick with a second or possibly third order model.

Polynomial Models in One Variable

3. Extrapolation, while never advisable, is super inadvisable in polynomial regression. Why? Same reason as before: we have no idea if the same relationship holds for data outside of the range of the variables in our sample.

4. Ill-Conditioning part 1: As you can imagine, as the order of our model increases, so too does the linear dependence (i.e., multicollinearity) among our transformed predictors, which is problematic just as before.
   - Sometimes, this problem can be solved by centering the predictor (subtracting the sample mean from each value of the variable). However, this doesn't always work.

5. Ill-Conditioning part 2: If our x variable has a narrow range, then even a lower order polynomial model will begin to suffer from problems with multicollinearity.

Polynomial Models in One Variable

6. When we have multiple predictors in a polynomial model, some people argue that all terms should be included in the model, even if some are found to be statistically insignificant. Such a model is said to maintain **hierarchy**.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{1i}^3 + \varepsilon_i$$

If we decide to take an insignificant term out to improve model fit, then such a model is said to not maintain hierarchy.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_3 x_{1i}^3 + \varepsilon_i$$

Polynomial Models in One Variable

- Let's look at an example using the hardwood data in D2L.

Spline Regression

- Consider the scatterplot from the hardwood example. We could obviously see that the relationship resembled an inverted parabola and that indicated to us that a polynomial model might be useful here.

- If we think about it a little more contextually, we can see that at some point (around 10% or so), the relationship goes from a pretty strong positive relationship to a pretty strong negative relationship.

- Could we fit two different models? One for part of the data and one for another? This principle of building a piecewise regression model is commonly referred to as **spline** regression and can be implemented alongside polynomial regression, if needed.

Spline Regression

- That inflection point we noted before is more technically referred to as a "knot." We can technically have as many knots as we want, but again, parsimony should be a major consideration.

- Let's say we want to build a spline regression using a single knot. We'll refer to the value of the knot as $x^{(k)}$. We will now define a dummy variable called x_k:

$$x_k = \begin{cases} 0 & \text{if } x_1 \leq x^{(k)} \\ 1 & \text{if } x_1 > x^{(k)} \end{cases}$$

Spline Regression

- Then our model becomes:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 (x_{1i} - x^{(k)}) x_k + \varepsilon_i$$

Spline Regression

- If $x_1 \leq x^{(k)}$, then the model is:

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

- If $x_1 > x^{(k)}$, then the model is:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i} - \beta_2 x^{(k)} + \varepsilon_i$$
$$y_i = (\beta_0 - \beta_2 x^{(k)}) + (\beta_1 + \beta_2) x_{1i} + \varepsilon_i$$

Spline Regression

- Alright, seems cool, but how do we choose $x^{(k)}$? There probably should be some practical rationale behind our choice as we may be tempted to do this all the time and overfit our model (there's also a little bit of an ethical slant to this as well, specifically in academic research).

- How do we do this in R? Let's look at an example using the car data in D2L.

Spline Regression

- A couple of notes on spline regression:

1. When I talk about being cautious about overfitting a model, we can think of potential issues in a few respects. First, as our $\varepsilon_i \to 0$, this also implies $MSE \to 0$. If we don't have error, we can't run tests. It also introduces bias. Second, more practically, if we overfit a model to our specific data, then we lose generalizability.

Spline Regression

2. When performing academic research, we (should) start with some research question and determine our model prior to collecting data at all. Then once we've collected our data, we build our model and answer our research question. People have a tendency to get worried if they get non-significant results and may be tempted to do some wiggling with their data/modeling in order to get significant results. In my opinion (as well as my advisor's), this is unethical and frankly, bad science. It's a little different in industry, but is for sure a no-no in the academy.

LOcal regrESSion (LOESS)

- We can expand upon the idea of spline and polynomial regression with a slightly different and flexible approach.
- If we know we have a non-linear relationship, but perhaps not one that follows a particular pattern that can be easily modeled with a polynomial function on a global scale (i.e., using all of the points), it may be more beneficial to partition the data points into local neighborhoods to fit a piecewise model together in that way.
- This general concept of fitting local regression models is called LOESS.

LOESS

- Here's the big idea: we choose a span, denoted as s, which is the proportion of points used to compute a local regression model at a particular point, say x_0. We then choose the $n \times s$ points whose x values are closest to x_0.
  - Note, every x point will have an opportunity to serve as x_0.

- A weight is then assigned to each point in x_0's neighborhood, with those being nearer to x_0 having greater weights than those further away. A weighted least squares regression is then performed using the aforementioned weights by finding $\widehat{\beta}_0$ and $\widehat{\beta}_1$ that minimize:

$$\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)^2$$

LOESS

- Our fitted value, $\hat{y}_i$ for a specific x_0 is given by $\hat{\beta}_0 + \hat{\beta}_1 x_0$ as before.
- Let's try to use LOESS on the cars data to see how it compares to polynomial regression and splines using R.