Nonlinear Regression Models

- Polynomial regression
- Spline regression
- Smoothing splines
- Local regression
- Generalized additive models

Polynomial Regression

Assume $x_i \in \mathbb{R}$ [a]

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 \cdots + \beta_d x_i^d + \text{err.}$$

Create the following new variables: $X_2 = X^2, \cdots, X_d = X^d$, then treat as a multiple linear regression model:

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}_{n \times 1} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & \cdots & x_2^d \\ & & & & \\ 1 & x_n & x_n^2 & \cdots & x_n^d \end{pmatrix}_{n \times (d+1)} \begin{pmatrix} \beta_0 \\ \cdots \\ \beta_d \end{pmatrix}_{(d+1) \times 1} + \text{err.}$$

[a] From now on, assume $x \in \mathbb{R}$ is one-dimensional. Extensions to the multi-dimensional case will be discussed later.

- Fit a polynomial model in R: the following two are equivalent

```
> lm(y ~ X + I(X^2) + I(X^3) )
> lm(y ~ poly(X, 3))
```

where `poly(X,3)` generates a design matrix with columns being the orthogonal polynomials that form a basis for polynomials of degree 3.[a]

- We are more interested in the fitted curve, as well as the prediction at some new location x^*, and less interested in the estimated coefficients because coefficients depend on the choice of the basis function. For example, the two sets of coefficients from the above R commands have totally different interpretation while the two fitted curves are the same.

---

[a]For the orthogonal polynomials, the j-th basis function involves all the X^k terms for $k = 0, \ldots, j$.

How to draw the fitted curve in `R`

1. create a grid of values for X (from small to large);
2. obtain the prediction of Y on those X points;
3. connect those points using command "`lines`".

How to choose d?

- Forward approach: keep adding terms until the coefficient for the newly added term is not significant.
- Backward approach: start with a large d, keep eliminating the highest order term if it's not significant until the highest order term is significant.
- Instead of adding/eliminating by hypothesis testing, you can also run forward/backward variable selection with AIC/BIC, or Cross-validation.
- Why not search over all possible sub-models?

- Question: Suppose we've picked d, then should we test whether the other terms, x^j's with $j = 1, \ldots, d-1$, are significant or not?

  Usually, we don't test the significance of the lower-order terms. When we decide to use a polynomial with degree d, by default, we include all the lower-order terms in our model.

- Why? For example, suppose the true curve is equal to X^2, i.e., if we fit a polynomial model with bases $1, X, X^2$, then the optimal model is of size 1.

  However we fit the data using "`poly(X, 2)`", then the optimal model would be a full model with coefficients for all basis functions including the intercept.

- Of course, if you have a particular polynomial function in mind, e.g., the data are collected to test a particular physics formula $Y \approx X^2 +$ constant, then you should test whether you can drop the intercept and the linear term.

- Or if experts believe the relationship between Y and X should be $Y \approx (X-2)^2$, then you should check the R output for

  ```
  lm(Y ~ X + I((X-2)^2))
  ```

  to test whether you can drop the linear term and the intercept.

- Otherwise, all we care is whether the data can be fitted by a polynomial of order d where d is the highest order.

Global vs Local

- When the data are too wiggly, we have to use a high-order polynomial. But high-order polynomials are not recommended in practice: results are not stable, tail-behavior is bad, and difficult to interpret.

- Using polynomial functions, we make a global assumption on the true mean function $\mathbb{E}(Y \mid X = x) = f(x)$. But global model is less flexible when data are wiggly. (Recall the discussion on linear regression and KNN.)

- Instead, we can try some local polynomial regression methods: estimate the function locally using piece-wise polynomials (splines) or fitting it locally (local regression).