7.1: Generalized Additive Models (GAM)

Dr. Bean - Stat 5100

1 Why GAMs?

Up to this point we have assumed models of the form

$$E\left(Y|X_1, X_2, \ldots, X_{p-1}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1}$$

which quickly fail if the effects cannot be modeled linearly (with possible variable transformations).

As with LOESS regression, GAMs do not assume a particular model form. GAMs only assume that the effects of each variable are additive (no interactions) i.e.

$$E\left(Y|X_1, X_2, \ldots, X_{p-1}\right) = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_{p-1}(X_{p-1})$$

where f_k is some unknown function that relates X_k to Y.

In practice, f_k are approximated using **cubic smoothing splines**, though any smoothing strategy will do. Splines are preferred because they are computationally much more efficient than weighted regression (as is used in LOESS). The key is that the smoothing occurs *over one dimension at a time* after accounting for the effects of the other dimensions.

2 Fitting GAMs

2.1 Splines

Splines are a series of piecewise polynomials that is continuous and differentiable at the break points, called **knots** (see Figure 1).

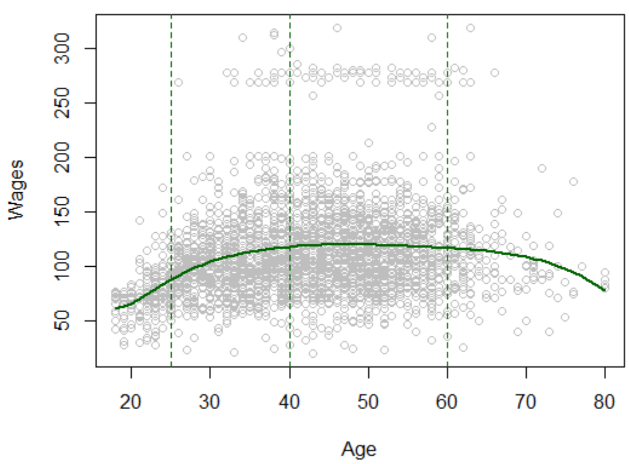

Figure 1: Example of cubic spline with three knots. Code to create image taken from `https://datascienceplus.com/cubic-and-smoothing-splines-in-r/`.

The more knots, the more "wiggly" your data has the chance to become.

Smoothing Splines place a knot at every unique value of X. This will most certainly lead to overfitting, so the splines are constrained by an *effective* degrees of freedom to prevent this. The effective degrees of freedom is often selected via **cross validation**.

The best part about smoothing splines is that it eliminates the need to select the placement of the knots.

2.2 Fitting GAMs

Remember that we have a unique smoothing spline that relates each X to Y after accounting for the effects of all other X's.
Hastie et al. (2002) outlines the algorithm for fitting GAMs

1. Set $b_0 = \bar{y}$ and all $\hat{f}_j \equiv 0$.
2. Cycle through all possible values for j each time updating predictions as:
   - Fit a smoothing spline S_j to the points (x_{ij}, y_j) after accounting for the effect of all other explanatory variables i.e.
   $$\hat{f}_j \leftarrow S_j \left[\left\{ y_i - b_0 - \sum_{k \neq j} \hat{f}_k(x_{ik}) \right\}_1^N \right]$$
   - To account for machine imprecision, recenter the splines around 0 i.e.
   $$\hat{f}_j \leftarrow \hat{f}_j - \frac{1}{N} \sum_{i=1}^{N} \hat{f}_j(x_{ij})$$
3. Repeat Step 2 until the predicted values of $\hat{f}_j$ stop changing (or change very little).

3 Comparisons to Other Methods

3.1 LOESS

- GAMs using splines are much more computationally efficient than LOESS.
  - Splines fit models to "chunks" of data, while LOESS uses weighted least squares on moving neighborhoods of data.
  - The sliding window requires continually recalculating distances between points in LOESS, which does not scale well to large datasets.
- GAMs fit smoothing curves to each variable individually, while LOESS fits one weighted regression surface to all explanatory variables at the same time.
- LOESS is great for one to two dimensional smooths on moderate to small datasets. GAMs are better suited for high dimensional data on large datasets.

3.2 OLS

- Both GAMs and OLS are easy to explain to other people (GAM effects are easy to visualize).
- GAMs do not require linearity like OLS.
- GAMs more computationally expensive to fit compared to OLS.
- GAMs can suffer from instability if the data is sparse at the endpoints.

4 Good Resources

- Hastie, Trevor, Tibshirani, Robert, and Friedman, Jerome (2001) “The Elements of Statistical Learning” (Chapter 9) `https://web.stanford.edu/~hastie/Papers/ESLII.pdf`