Lecture 3: Splines and Model Comparison

Will Welch (adapted from materials by Matias Salibian-Barrera)

Department of Statistics, University of British Columbia

STAT 447B: Methods for Statistical Learning

September–December 2014

Today's Learning Goals

1. Review
2. Understanding Spline Basis Functions
3. Comparing Models

Linear Splines Review

Recall the linear splines model

$$E\,[\,Y|\,x] \;=\; \beta_0 + \beta_1\,x + \sum_{k=1}^{K} \beta_{k+1}\,f_k(x)$$

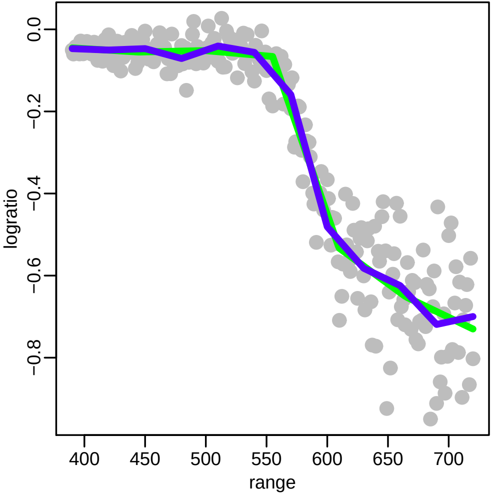
with basis functions $f_k(x)$ defined as

$$f_k(x) \;=\; (x - \kappa_k)_+ \;=\; \begin{cases} x - \kappa_k & \text{if } x - \kappa_k > 0 \\ 0 & \text{otherwise} \end{cases}$$

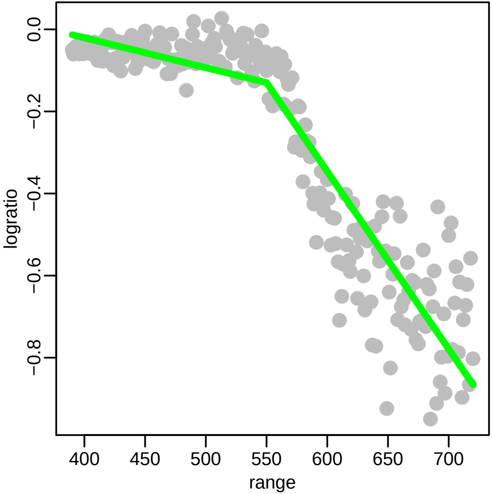
and the knots κ_k, $1 \leq k \leq K$ are typically chosen as

$$\kappa_k \;=\; \left(\frac{j}{K+1}\right) 100\% \text{ quantile of the observed covariates}$$

Lidar Data: Linear Splines, 5 Knots Versus 10 Knots



Lidar Data: Linear Splines, 1 Knot at 550



Lidar Data: Linear Splines, 1 Knot at 550

- 1 knot at $\kappa_1 = 550$
- `range` is the only explanatory variable, call it x
- To simplify the following activity, suppose x only takes the values

$$400, 450, 500, 550, 600, 650, 700$$

- Objective: Justify the form of the expanded **X** matrix.

Lidar Data: Models for Linear Splines, 1 Knot at 550

- We want to fit 2 lines, one for $x \leq 550$ and one for $x > 550$
- Suppose we fit

$$\begin{aligned} Y &= \beta_0 + \beta_1 x + \epsilon \quad \text{(for data with } x \leq 550\text{)} \\ Y &= \beta_2 + \beta_3 x + \epsilon \quad \text{(for data with } x > 550\text{)} \end{aligned}$$

- What's wrong with this? **Clicker question 1.**

Lidar Data: Models for Linear Splines, 1 Knot at 550

- The 2 lines have to join at $x = \kappa_1 = 550$
- Let's change the formulation a bit (the orignal formulation will be on Assignment 1)

$$\begin{aligned} Y &= \beta_0 + \beta_1 x + \epsilon \quad \text{(for data with } x \leq 550\text{)} \\ Y &= \beta_2 + (\beta_1 + \beta_3)x + \epsilon \quad \text{(for data with } x > 550\text{)} \end{aligned}$$

- Two lines join at $x = \kappa_1 = 550$, i.e.,
  $\beta_0 + \beta_1 \times 550 = \beta_2 + (\beta_1 + \beta_3) \times 550$
- Rewrite this linear dependency in the form $\beta_2 = \dots$
  Clicker question 2.

Lidar Data: Models for Linear Splines, 1 Knot at 550

- Hence the 2 lines become

$$Y = \begin{cases} \beta_0 + \beta_1 x + \epsilon & \text{(for data with } x \leq 550\text{)} \\ \beta_0 + \beta_1 x + \beta_3(x - 550) + \epsilon & \text{(for data with } x > 550\text{)} \end{cases}$$

- In the **X** matrix we need 3 columns: a column of 1's, a column of *x* values, and another column. What is the formula that gives the third column (corresponding to β_3)?
  Clicker question 3.

- (As β_2 has been eliminated, relabel β_3 as β_2 to agree with previous notation.)

Lidar Data: **X** for 1 Knot at 550

- Recall that to simplify the activity, we suppose x only takes the values

$$400, 450, 500, 550, 600, 650, 700$$

- Numerically, what is the last column of the **X** matrix?
Clicker question 4.

Lidar Data: Comparing Models

- Compute the linear spline fit to the `lidar` data set
  - using $K = 5$ knots; use it to compute predictions
  - using $K = 10$ knots; use it to compute predictions
- Which one is “better”?

Comparing Fits: Estimating Predictive Power

- In symbols, we're computing estimates

$$\widehat{E(Y|x)} = \hat{f}_K(x) \qquad \text{with } K = 5, 10$$

($\hat{f}_K(x)$ here is the prediction from the whole model with K knots, not just a basis function)

- Can we compare these fits using their sum of squared residuals?

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}_K(x_i)\right)^2$$

Lidar Data: Comparing Predictive Power

- Comparing sum of squared residuals

$$\text{SS(Residual)} = \frac{1}{n}\sum_{i=1}^{n}\Big(y_i - \hat{f}_K(x_i)\Big)^2$$

is equivalent to comparing

$$R^2 = 1 - \frac{\text{SS(Residual)}}{\text{SS(Total)}}$$

- $R^2 = 0.9242$ ($K = 5$)
- $R^2 = 0.9247$ ($K = 10$)

- What does this tell us? **Clicker question 5.**

Comparing Fits: Estimating Predictive Power

- Prediction error:

$$\left(Y - \hat{f}_K(x)\right)^2 \qquad \text{for future observations } (x, Y)$$

- This is a random quantity, it's customary to use the expected prediction error

$$E\left[\left(Y - \hat{f}_K(x)\right)^2\right]$$

  where the average is taken over observations (x, Y) that have not been used to compute $\hat{f}_K(x)$.

- We can't evaluate this quantity. Can we estimate it?

Comparing Fits: Estimating Predictive Power

- Cross-validation: remove some observations from the data set and pretend that they are “new” ones.
- More formally: split the data into a training set and a test set
- Fit the model using the training set, use it to predict the test set
- Compare the quality of the different fits using their average prediction error observed on the test set