**7.4 Fun Sheet**

**7.4.1 Piecewise Polynomials**

- Fitting separate low-degree polynomials over different regions of X (rather than fitting a high-degree polynomial over the entire region) is called piece-wise polynomial regression.

- Points where the coefficients change are called knots.

- A model of the form

where is an example of a piecewise cubic polynomial (d = 3) with 2 knots and uses 12 degrees of freedom.

- A model of the form

is an example of a piecewise quadratic polynomial (d = 2) with 0 knots and uses 3 degrees of freedom.

- Write an example of a piecewise quintic polynomial (d = 5) with 1 knot. How many degrees of freedom does it use?

Answer:

It uses 12 degrees of freedom.

- If we place K different knots throughout the range of X, the we end up fitting K + 1 different polynomials.

- Food for Thought: Why do we generally fit low-degree piecewise polynomials rather than high-degree ones?

Sample Answers: Because the flexibility comes from the knots. Using high degrees defeats its simplifying purpose. If the pieces are too flexible then you might as well use one high-degree polynomial over the entire region.

**7.4.2 Constraints and Splines**

- A degree-d spline is a piecewise degree-d polynomial with continuity in derivatives up to degree d-1 at each knot.

- No jumps 🡪 Constraint that the fitted curve must be continuous at each knot

- Smooth 🡪 Constraint that the 1st and 2nd derivatives (in the case of cubic) be continuous at each knot

- Each constraint frees up a degree of freedom.

- A piecewise cubic polynomial with 1 knot uses 8 degrees of freedom, but a cubic spline with 1 knot uses 5 degrees of freedom.

- In general, a cubic spline with K knots uses 4+K degrees of freedom.

- The only constraint on a linear spline is continuity at each knot, while a quadratic spline requires both that and continuity in the 1st derivative at each knot.

- Note that cubic splines are popular because most human eyes cannot detect the discontinuity at the knots.

**7.4.3 The Spline Basis Representation**

- To represent a cubic spline, add one truncated power basis function for each knot, defined as

where is the knot.

- To fit a cubic spline into a data set with K knots, we perform least squares regression with an intercept and 3+K predictors, of the form where are the knots.

- Write the model for a cubic spline with knots at x = 5 and x = 10.

Answer:

- A natural spline is a regression spline with additional boundary constraints: the function is required to be linear at the boundary (in the region where x is smaller than the smallest knot or larger than the largest knot).

- This leads to more stable estimates at the boundaries, with narrower confidence intervals.

**7.4.4 Choosing the Number and Locations of the Knots**

- The regression spline is most flexible in regions that contain a lot of knots, because in those regions polynomial coefficients can change rapidly.

- We can put more knots where we think the function might vary most rapidly.

- It’s common practice to place knots uniformly. Do this by specifying the degrees of freedom and then the number of knots and spacing is automatic. Note that the spacing is by data distribution rather than distance.

- To choose the number of knots, you can see what looks best or use Cross Validation. (Do this by removing a portion of the data, fitting a spline, predicting the held out portion, repeating until every observation has been left out once, and then computing the overall RSS. Repeat this procedure for different numbers of knots K. Then choose the K that lead to the smallest RSS. Alternatively, use glm() to construct the model instead of lm() and then use the cv.glm() function to produce the K-fold cross validation errors)

**7.4.5 Comparison to Polynomial Regression**

- Splines introduce flexibility by increasing the number of knots but keeping the degree of the polynomial fixed, which generally leads to more stable estimates than simply raising the polynomial degree.

Explanation: Say you want a flexible model. Using a high-degree polynomial could make a messy squiggly line vs using a regression spline 🡪 adds flexibility while keeping it stable.

-Name a reason to use splines instead of a high-degree polynomial.

- Splines allow you to place the knots/choose where it’ll be most flexible. 🡪 more in control

- Especially useful for boundaries where polynomials tend to go cray.

**7.8.2 Splines in R**

- What library is required for using splines? splines

- Which function generates the basis functions for splines? bs()

- What splines are produced by default? cubic

- Which function do you use to fit natural splines? ns()

- Why do we still use lm or glm? Because we’re still doing a linear model with least squares regression, just split into pieces.