Chapter 7 Moving Beyond Linearity

1. Polynomial regression: Extends the linear model by adding extra predictors, obtained by raising each of the original predictors to a power.
2. Step functions: Cut the range of a variable into k distinct regions, in order to produce a qualitative variables.(Has the effect of fitting a piecewise constant function)
3. Regression splines: Dividing the range of X into K distinct regions. Within each region, a polynomial function is fit to the data. These polynomials are constrained so that they join smoothy at the boundaries(knots)🡪Can produce and extremely flexible fit
4. Smoothing splines: Similar to regression splines, Resulting from minimizing a residual sum of squares subject to smoothness penalty
5. Local regression: Similar splines, but allows regions to overlap
6. Generalized additive models: Extend the methods above to deal with multiple predictors

7.1 Polynomial Regression

Polynomial Function:



1. The coefficients can still be estimated using least squares linear regression🡪 The coefficients are not of particular interest
2. Unusual to use d greater than 3 or 4 because the shape would be overly flexible
3. (2X) standard error curves: The standard error for each reference point times 2

7.2 Step Functions

Drawbacks of polynomial regression: Imposes a global structure on the non-linear function of X.

Step function:

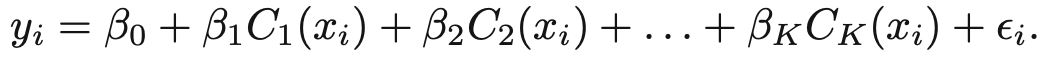
1. Break the range of X into bins
2. Fit a different constant in each bin🡪Amounts to converting a continuous variable into an ordered categorical variable
   1. Create cutpoints in the range of X and then construct K+1 new variables

A screenshot of a cell phone

Description automatically generated

Where I(.) is an indicator function that returns 1 if the condition is true, and returns 0 otherwise

1. Use least squares to fit a model using predictors



* 1. can interpreted as the mean value of Y for
  2. represents the average increase in the response for X in

1. Drawbacks of Step Functions: Unless there are natural breakpoints in the predictors, piecewise-constant functions can miss the action

7.3 Basis Functions

Have at hand a family of functions or transformations that can be applied to a variable X: , and fit the model:



Wavelets and Fourier series can be constructed

7.4 Regression Splines

7.4.1 Piecewise Polynomials

Fitting a cubic regression model of the form



Where coefficients differ in different parts of the range of X (The points where the coefficients changes are called knots)

1. Using more knots leads to a more flexible piecewise polynomial🡪If we place K different knots throughout the range of X, we will end up fitting K+1 different polynomials
2. The function can be discontinuous at the knots

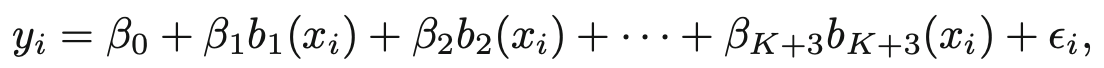
7.4.2 Constraints and Splines

Constraints:

1. Fitted curve must be continuous
2. Both the first and second derivatives of the piecewise polynomials are continuous
   1. We are requiring that the piecewise polynomial be not only continuous but also smooth
   2. Each constraint frees up one degree of freedom by reducing the complexity of the resulting polynomial fit🡪
      1. A degree-d spline: A piecewise degree-d polynomial, with continuity in derivates up to degree d-1 at each knot

7.4.3 The Spline Basis Representation

A cubic spline with K knots can be modeled as



For an appropriate chose of basis functions

Truncated basis power basis function:

A close up of a clock

Description automatically generated

Where is the knot

Representing a cubic spline: Start off with a basis for a cubic polynomials a and then add one truncated basis function per knot

1. In order to fit a cubic spline to a data set with K knots, we need to estimate K+4 (uses k+4 degrees of freedom) regression coefficients🡪
2. Splines can have high variance at the outer range of the predictor

Natural spline: A regression spline with additional boundary constraints🡪The function is required to be linear at the boundary

7.4.4 Choosing the Number and Locations of the Knots

Place the knots in a uniform fashion:

1. Specify the desired degrees of freedom
2. Have software automatically place the corresponding number of knots at the uniform quantiles of the data

Try out a number of knots and see which produces the best looking curve🡪Cross Validation

1. Remove a portion of the data
2. fit a spline with a certain number of knots to the remaining data
3. Use the spline to make predictions for the held-out portion
4. Procedure can be repeated for different of knots K
5. The value of K giving the smallest RSS is chosen

7.4.5 Comparison to Polynomial Regression

Regression splines often give superior results to polynomial regression.

1. Splines allow us to place more knots, and hence flexibility, over regions where the function f seems to be changing rapidly, and fewer knots where f appears more stable.

7.5 Smoothing Splines

7.5.1 An Overview of Smoothing Splines

Find a function g that makes RSS small, but also ensures its smoothness:

A picture containing object, table

Description automatically generated Loss+Penalty

Where is a nonnegative parameter. The function g that minimizes the equation is known as a smoothing spline

1. Loss function: Encourages g to fit the data well
2. Penalty term: Penalizes the variability in g🡪 measure of the total change in function over the entire range🡪 The larger the , the smoother g will be🡪 controls the bias variance trade-off
3. Function g(x) that minimizes the equation is a natural cubic spline with knots at

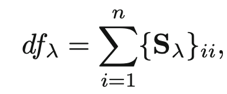
7.5.2 Choosing the Smoothing Parameter

A smoothing spline is simply a natural cubic spline with knots at every unique value of

1. As increases from 0 to , the effective degree of freedom, , decreases from n to 2🡪 is a measure of the flexibility of the smoothing spline



1. is a n-vector containing the fitted values of the smoothing spline at the training points
2. is defined to be



The sum of diagonal elements of the matrix

Find the appropriate smoothing parameter,

LOOCV can be computed vert efficiently for smoothing splines, with essentially the same cost as computing a single fit

A picture containing clock

Description automatically generated

\*We can compute each of the LOOCV fit using the original fit

7.6 Local Regression

Involves computing the fit a target point using only the nearby training observations

A screenshot of a cell phone

Description automatically generated

1. In order to obtain the local regression fit at a new point, we need to fit a new weighted least squares model by minimizing for a new set of weights
2. Choices to be made in Local Regression
   1. Define the weighting function
   2. The type of function to fit
   3. Span s(Step 1): Controls the flexibility🡪Smaller s, local and wiggly git; Larger S, global fit that uses most of the training data

Varying coefficient models: Fitting a multiple linear regression that is global in some variables, but local in another

Local regression can perform poorly if p is much larger than about 3 or 4.

7.7 Generalized Additive Models

Provides a general framework for extending a standard linear model by allowing non-linear functions of the variables, while maintaining additivity.

7.7.1 GAMs for Regression Problems

Additive models:

A close up of a logo

Description automatically generated

\*Calculate a separate for each , and then add together all of their contributions

\*Fitting a GAM with smoothing spline

a. Not quiet as simple as fitting a GAM with a natural spline

b. Backfitting: Fit a model involving multiple predictors by repeatedly updating the for each predictor in turn, holding the others fixed

\*WE can use local regression, polynomial regression, and other splines as the building blocks for GAMs

Pros and Cons of GAMs

Pros:

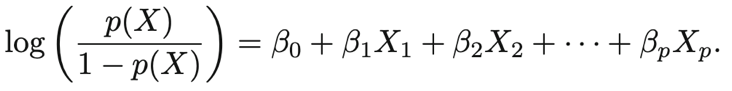
1. Allows fitting a non-linear to each , so that we can automatically model non-linear relationships that standard linear regression will miss
2. Non-linear fits can potentially make more accurate predictions for the response Y.
3. Because the model is additive, we can still examine the effect of each on individually while holding all of the other variables fixed.
4. Smoothness of the function for the variable can be summarized via degrees of freedom

Cons:

1. Model is restricted to be additive🡪May miss important interactions

7.7.2 GAMs for Classification Problems

Logistic Regression(Logit Model):



1. Logit is the log of odds of:

Logistics regression GAM:

A close up of a logo

Description automatically generated

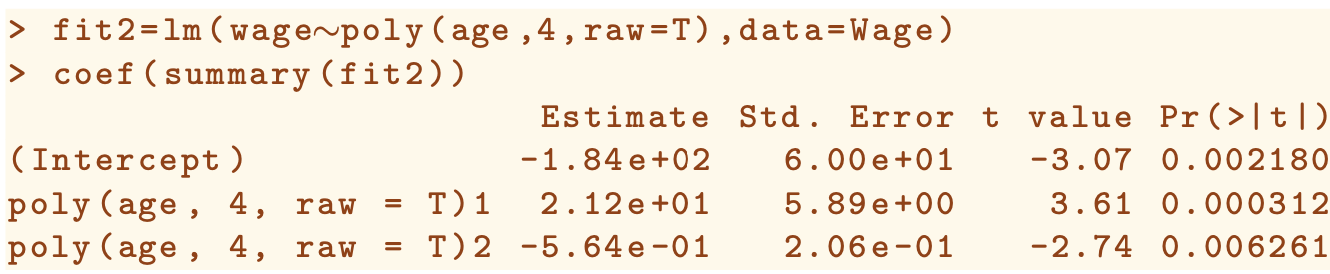
7.8 Lab: Non-linear Modeling

7.8.1 Polynomial Regression and Step Functions

A screenshot of a cell phone

Description automatically generated

1. Poly(): returns a matrix whose columns are a basis of orthogonal polynomials<-Each column is a linear combination of the variables age,age^2, age^3, age^4
   1. Arguments raw=TRUE will obtain age,age^2, age^3, age^4 directly



1. I(): use wrapper functions to create polynomial basis functions on the fly

A picture containing knife, table

Description automatically generated

1. Cbind(): building a matrix from a collection of vector; any function call such as cbind() inside a formula serves as a wrapper



Predict(): create a grid value of the ages that we want predictions, call the predict() function

A close up of a logo

Description automatically generated

Plot the data and add the fit from the degree-4 polynomial

A screenshot of a cell phone

Description automatically generated

A close up of a logo

Description automatically generated

1. Mar and oma arguments to the par() allow us to the control the margins of the plot
2. Title() creates a figure that spans both subplots

Anova() function: Performs an analysis of variance (ANOVA, using an F-test) in order to test the null hypothesis that a model is sufficient to explain the data against the alternative hypothesis that a more complex model is required

1. and must be nested models: the predictors in must be a subset of the predictors in
2. P values can be obtained by exploiting that poly() creates orthogonal polynomials

A screenshot of a cell phone

Description automatically generated

Glm() function:

1. Family=”binomial” fits a polynomial logistic regression model



1. Wrapper I() creates a binary response on the fly
   1. Wage>350 evaluates to a logical variable containing true or false
   2. Glm() coerces the wage>250 to binary by setting true to 1 and False to 0
2. Predict(): make predictions
   1. The default prediction type is type=”link”. Our fitted model is in the form of logit:

A close up of a logo

Description automatically generated

Predictions are in the form of , standard errors are also of this form

* 1. Confidence Interval for , we need to use the transformation:

A picture containing object, clock

Description automatically generated

The transformation should be applied to the result as well as the SE

A picture containing knife, table

Description automatically generated

Plot the fit and the SE band

A picture containing knife

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. Jitter() function: jitter the age values a bit so that observations with the sam age value do not cover each other up

Fit a step function:

Cut() function:

* + 1. Automatically picked the cut points, we only need to specify the number of cut points we want
    2. Returns a n ordered variable

Lm(): creates a set of dummy variables for use in the regression,

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

7.8.2 Splines

Splines library

Regression splines can be fit by constructing an appropriate matrix of basis functions

Bs():

1. Generates the entire matrix of basis functions for splines with the specified set of knots
2. By default, cubic splines are produced

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

1. Following creates a spline of 6 basis functions
2. Df option in bs() allows to produce a spline with knots at uniform quantiles of the data

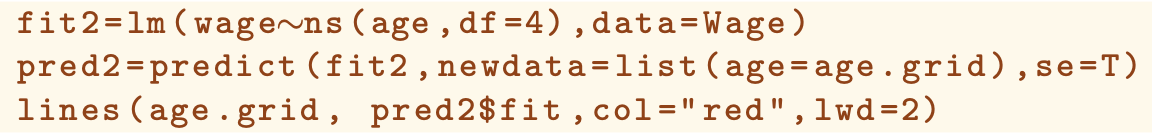
A screenshot of a cell phone

Description automatically generated

1. Degree argument allows fitting splines to any degrees

Fitting natural splines:

Ns() function: syntax the same as bs()



Fitting a smoothing spline

A picture containing table

Description automatically generatedA picture containing clock

Description automatically generated

Smooth.spline():

1. Df=16: the function then determines which lambda leands to 16 degrees of freedom
2. Cv=True, select the smoothness level by cross-validation

Local Regression:

Loess() function:

A picture containing table

Description automatically generatedA screenshot of a cell phone

Description automatically generated

1. The larger the span, the smoother the fit

7.8.3 GAMs

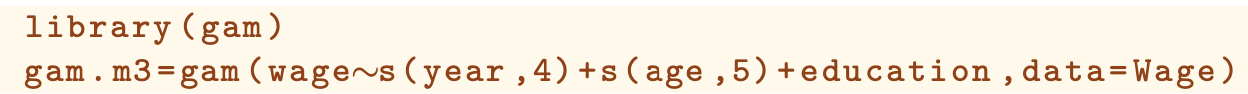
Since GAM is just a big linear regression model using appropriate choice of basis functions, we can simply use the lm() function:

Natural spline GAM:  


Fitting smoothing splines or other components that cannot be expressed in terms of basis functions:  
gam library in R

Gam():

1. S(): indicates that we would like to fit a smoothing spline, need to specify the degree of freedom
2. Qualitative variables will be converted into dummy variables



1. Call the plot() function to will produce the fitting plot🡪 Invokes the plot.gam() methods
   1. For other GAM models, we can call the plot.gam() to produce the fitting plot

A picture containing clock

Description automatically generated

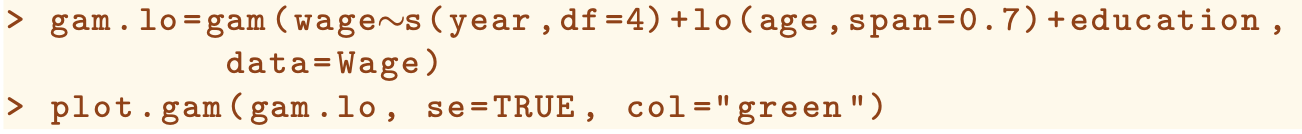
1. We can use anova() to compare different GAM models:  
   A screenshot of a cell phone

   Description automatically generated
   1. Gam.m1: Excludes year
   2. Gam.m2: linear relationship with year
   3. Gam.m3: Produces a smoothing splines function on year
2. Summary(): produces a summary of gam fit  
   A screenshot of text

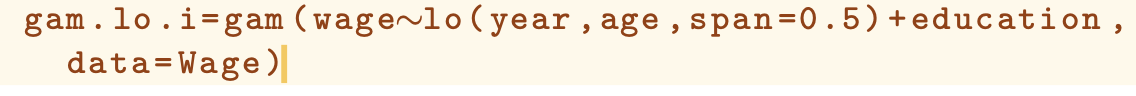
   Description automatically generated
   1. P-values for year and age correspond to a null hypothesis of linear relationship versus non-linear relationship
3. Predict() in gam works the same as lm:



1. Fit local regressions GAM: Using the lo() function



1. Use lo() to create interactions before calling the gam() function



* 1. Creates an interaction between year and age, fit by a local regression surface

1. Fit a logistic regression GA<, use the I() function in constructing the binary response, and set family=binomial

A picture containing knife, table

Description automatically generated