

Splines

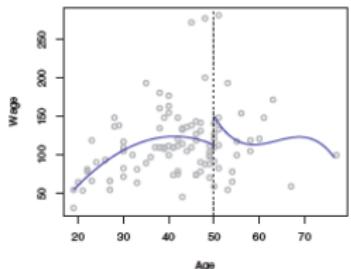
Interpolating splines

- ▶ ‘Spline’ is a mathematical tool useful in many contexts finalised to approximate functions or to **interpolate data**.
- ▶ we choose K points $\xi_1 < \xi_2 < \dots < \xi_K$, called **knots**, along the x -axis.

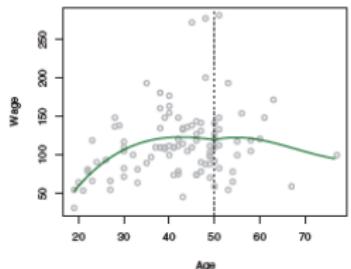
- ▶ a function $f(x)$ is constructed, so that it passes exactly through the knots and is free at the other points
- ▶ we look for “smooth” functions
- ▶ between two successive knots, in the interval (ξ_i, ξ_{i+1}) , curve $f(x)$ coincides with a **suitable polynomial**, of prefixed degree d ,
- ▶ these sections of polynomials meet at point ξ_i ($i = 2, \dots, K - 1$),
- ▶ in the sense that the resulting function $f(x)$ has a continuous derivative from degree 0 to degree $d - 1$ in each of the ξ_i .

Interpolating splines

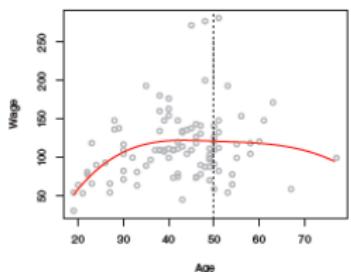
Piecewise Cubic



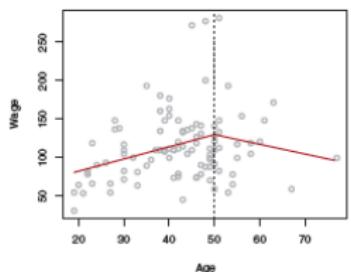
Continuous Piecewise Cubic



Cubic Spline



Linear Spline



Regression splines

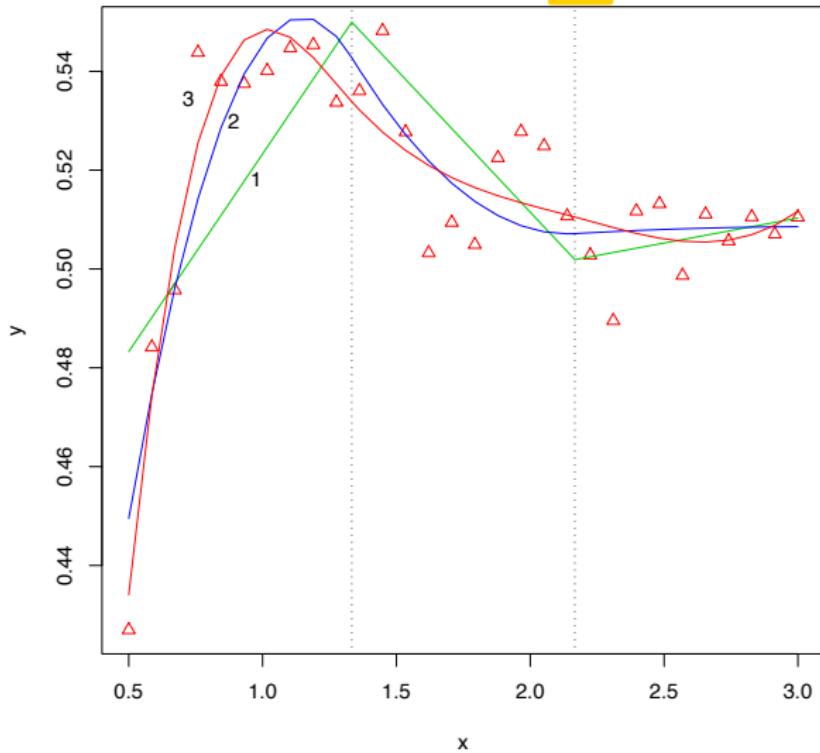
- ▶ We have n observed points (x_i, y_i) for $i = 1, \dots, n$ that we want to interpolate,
- ▶ we apply these ideas to parametric regression, by fitting a **cubic spline** ($d = 3$) to the n points
- ▶ we divide the x -axis into $K + 1$ intervals separated by K knots, ξ_1, \dots, ξ_K , and interpolate the n points with the **least squares criterion**
- ▶ the obtained function is called **regression spline**

Regression splines

- ▶ The number K of knots and their position along the x -axis need to be chosen
- ▶ Because K is a **tuning parameter** regulating the complexity of the model, we need to perform a model selection according to bias-variance trade-off
- ▶ Once the number K has been set, a reasonable choice for knots position is uniformly along the x_i range.

Regression splines

Interpolated functions for $d = 1, 2, 3$



Smoothing splines

- ▶ Let us consider the **penalized least squares** criterion

$$D(f, \lambda) = \sum_{i=1}^n [y_i - f(x_i)]^2 + \lambda \int_{-\infty}^{\infty} \{f''(t)\}^2 dt$$



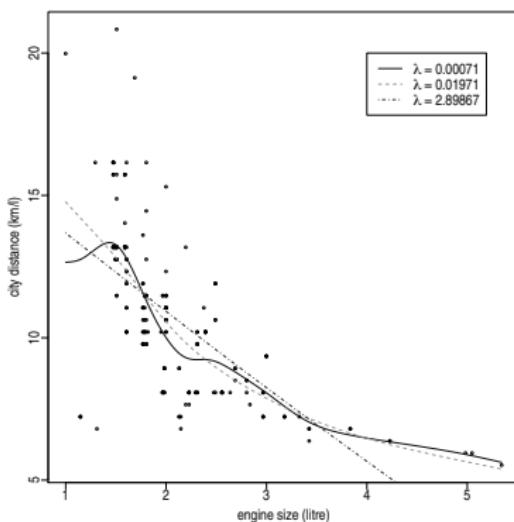
where λ is a positive **penalisation parameter** of the roughness degree of curve f (quantified by the integral of $f''(x)^2$), and therefore acts as a **smoothing parameter**.

- ▶ A noteworthy mathematical result shows that the solution to that minimization problem is represented by a **natural cubic spline** whose knots are distinct points x_i .



Smoothing splines

Estimate of city distance according to engine size by a smoothing spline, for three choices of λ



We can also use the criteria discussed earlier for the choice of smoothing parameter λ

Summarizing...

We have relaxed the linearity assumption while still attempting to maintain as much interpretability as possible. To this end, we consider approaches such as splines and local regression.

- ▶ Regression splines involve dividing the range of X into K distinct regions. Within each region, a polynomial function is fit to the data. However, these polynomials are constrained so that they join smoothly at the region boundaries, or knots. Provided that the interval is divided into enough regions, this can produce an extremely flexible fit.
- ▶ Smoothing splines are similar to regression splines, but arise in a slightly different situation. Smoothing splines result from minimizing a residual sum of squares criterion subject to a smoothness penalty.
- ▶ Local regression is similar to splines, but differs in an important way. The regions are allowed to overlap, and indeed they do so in a very smooth way.
- ▶ Generalized additive models allow us to extend the methods above to deal with multiple predictors.

Generalized Additive Models

Generalized Additive Models

- ▶ So far we have seen a number of approaches for flexibly predicting a response y on the basis of a single predictor x . These approaches may be seen as **extensions of simple linear regression**.
- ▶ Here we explore the problem of flexibly predicting y on the basis of several predictors, x_1, \dots, x_p .
- ▶ Generalized additive models (GAMs) provide a general framework for extending a standard linear model by allowing **non-linear functions of each of the variables**, while maintaining additivity.
- ▶ The beauty of GAMs is that we can use splines and local regression as **building blocks** for fitting an additive model

Additive models

- ▶ A natural way of extending the multiple linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

in order to allow for nonlinear relationships between each feature and the response is to replace each linear component $\beta_j x_{ij}$ with a smooth nonlinear function $f_j(x_{ij})$.

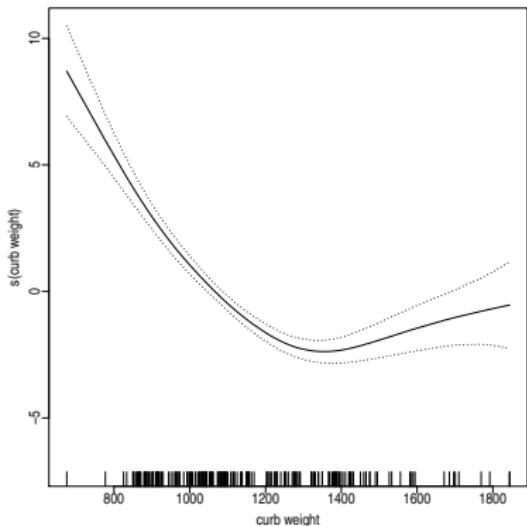
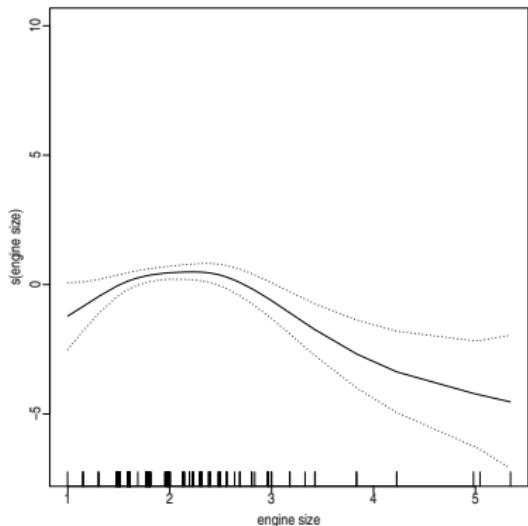
- ▶ We can then write

$$y_i = \beta_0 + \sum_{j=1}^p f_j(x_{ij}) + \varepsilon_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip}) + \varepsilon_i$$

- ▶ this is a Generalized Additive Model (GAM).
- ▶ It is called additive because we calculate a separate f_j for each X_j and then add together all of their contributions.

Additive models: Example

Estimate of city distance according to engine size and curb weight by an additive model with a spline smoother



Generalized Additive Models

GAM important properties:

- ▶ GAMs allow us to fit a non-linear f_j to each x_j , so that we can automatically model **non-linear relationships** that standard linear regression will miss. This means that we do not need to manually try out many different transformations on each variable individually.
- ▶ The non-linear fits can potentially make more accurate predictions for the response y .
- ▶ Because the model is additive, we can still examine the effect of each x_j on y individually while holding all of the other variables fixed.
- ▶ If we are interested in **inference**, GAMs provide a useful representation.

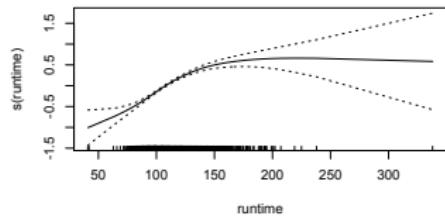
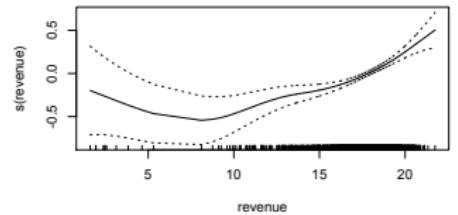
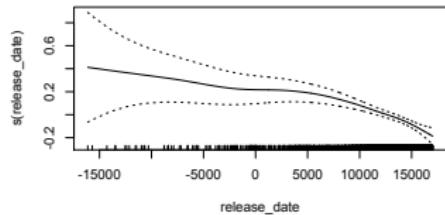
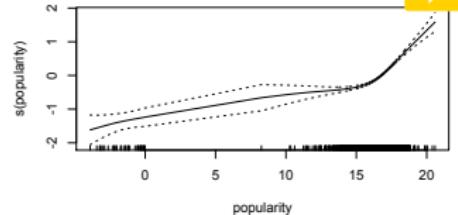
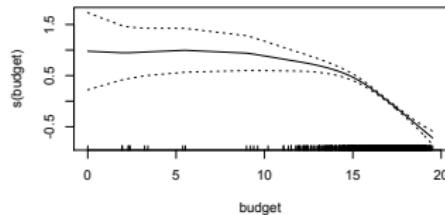
Generalized Additive Models

Case study

- ▶ We consider the case of a dataset about movies
- ▶ We are interested in understanding the variable 'average vote' obtained by movies
- ▶ We want to study the relationship with other variables such as 'budget', 'popularity', 'revenues', 'runtime'

Generalized Additive Models

Case study: we may appreciate some results obtained with GAM



Flexibility vs Interpretability of models

There is a trade-off between flexibility and interpretability

