Regression Splines

Following the concept of Basis Functions, Regression Splines is a non-parametric method which an extension to the linear models falls under the non-linear regression functions similar to Polynomial Regression

But unlike Polynomial Regression that transform the design matrix X globally, **Splines** fit a Basis Functions **locally** within each region and fitted using Ordinary Least Squares :

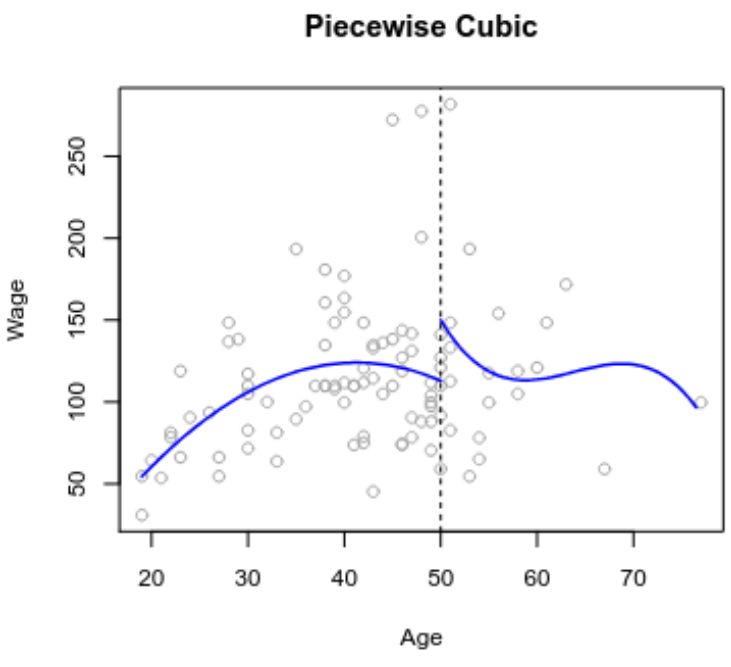
$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i$$

- This is a **Cubic Polynomial Regression** applied globally

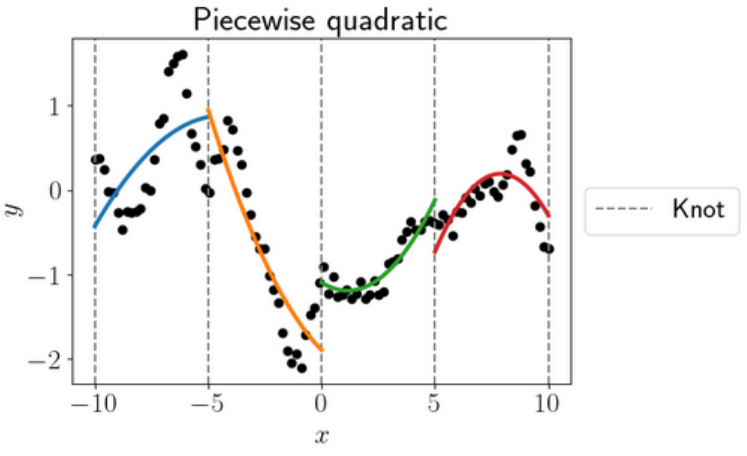
Piecewise Polynomials

$$y_i = \begin{cases} \beta_{01} + \beta_{11} x_i + \beta_{21} x_i^2 + \beta_{31} x_i^3 + \epsilon_i & \text{if } x_i < \kappa \\ \beta_{02} + \beta_{12} x_i + \beta_{22} x_i^2 + \beta_{32} x_i^3 + \epsilon_i & \text{if } x_i \geq \kappa \end{cases}$$

- Here we fitting the **Cubic Polynomial Regression** two times depending on the region κ which called **Knot**

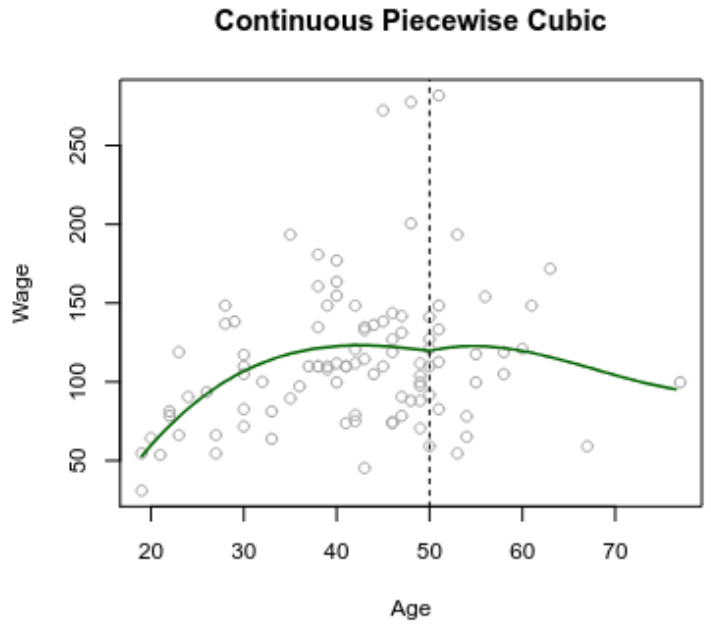


Using more knots results in more flexible piecewise polynomial, generally when fitting **piecewise polynomials splines** we will end up fitting $K + 1$



The thing to notice is that the fit is discontinuous at the **knots** to fix that we add a **constraint** to reinforce **continuity**

Let's take for example this **piecewise poly spline** and τ_K as the **knots** location :



Truncated Power Basis Function :

$$y_i = \beta_0 + \beta_1 h_1(x_i) + \cdots + \beta_{K+d} h_{K+d}(x_i) + \epsilon_i$$

- $h_K(x_i) = x^{K-1}$ represent the **polynomial basis function**
- $h_{K+d}(x_i)$ is a **basis function** to adjust the *intercept* to fix the **discontinuity**
  With :

  $$h_{K+d}(x) = (x - \tau_k)_+^{d-1} = \begin{cases} (x-\tau)^{d-1} & \text{if } x > \tau \\ 0 & \text{otherwise} \end{cases}$$

  for **Cubic Polynomial** :

  $$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 (x - \tau_1)^3 + \epsilon_i$$

- One **knot** τ
- The polynomial of an order $d = 4$

if $x < \tau_1$: **Region 1**

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3$$

if $x \geq \tau_1$: **Region 2**

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 (x - \tau_1)^3$$

By doing some algebra we get :

$$y_i = (\beta_0 + \beta_4 \tau_1^3) + x_i(\beta_1 + 3\beta_4 \tau^2) + x_i^2(\beta_2 - 3\beta_4 \tau) + x_i^3(\beta_3 + \beta_4) + \epsilon_i$$

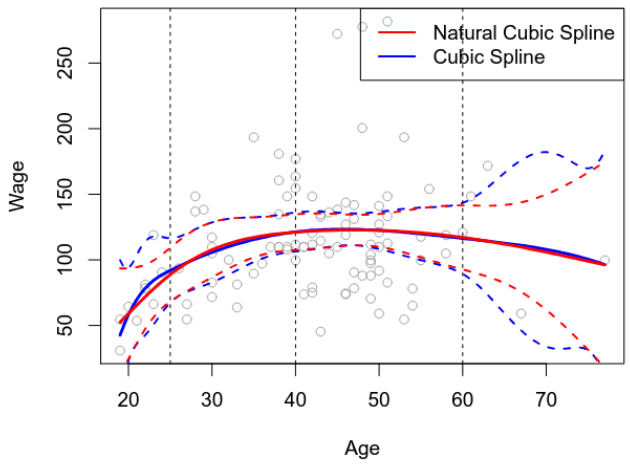
- The coefficient β_4 from the **truncated basis function** act as an **adjustment factor** across all local coefficients $\beta_0, \beta_1, \beta_2, \beta_3$

At $x = \tau_1$ both **Region 1** and **Region 2** share the same value which ensures **continuity**

Note : This mean we will always fit $K + d$ coefficients , with K being the number of **knots**

Limits of Piecewise Polynomials

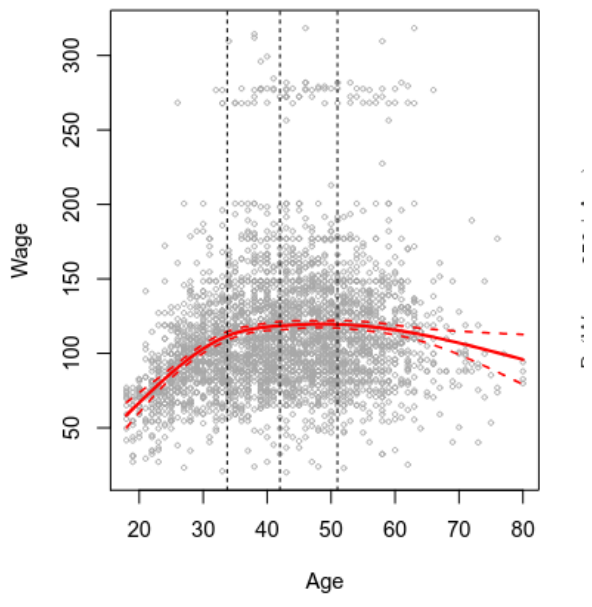
They can have very high variance at the outer range of the predictors , where X takes very large/small values



Choosing the Number and Locations of the Knots τ

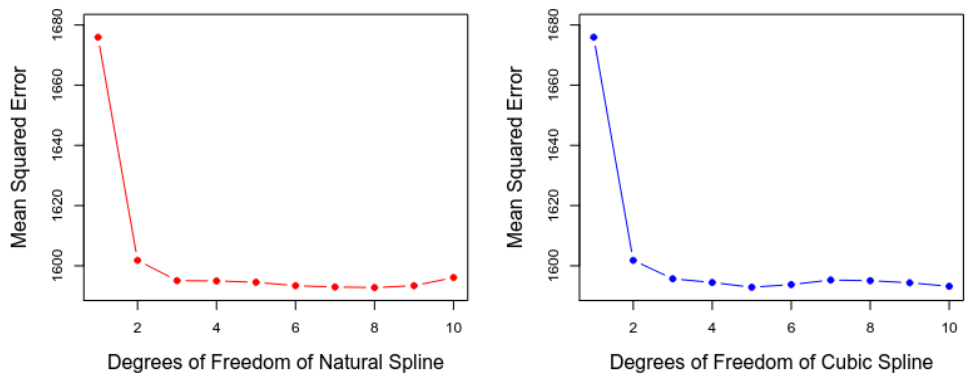
When fitting a **spline** the first question should be where we place the **Knots** τ, It might be easy to just place more **Knots** where the data vary the most and and less when it's stable a more, while this can work we would like more sophisticated way to select and locate the optimal number of **knots**

One way to do it is to specify the desired degree of freedom :



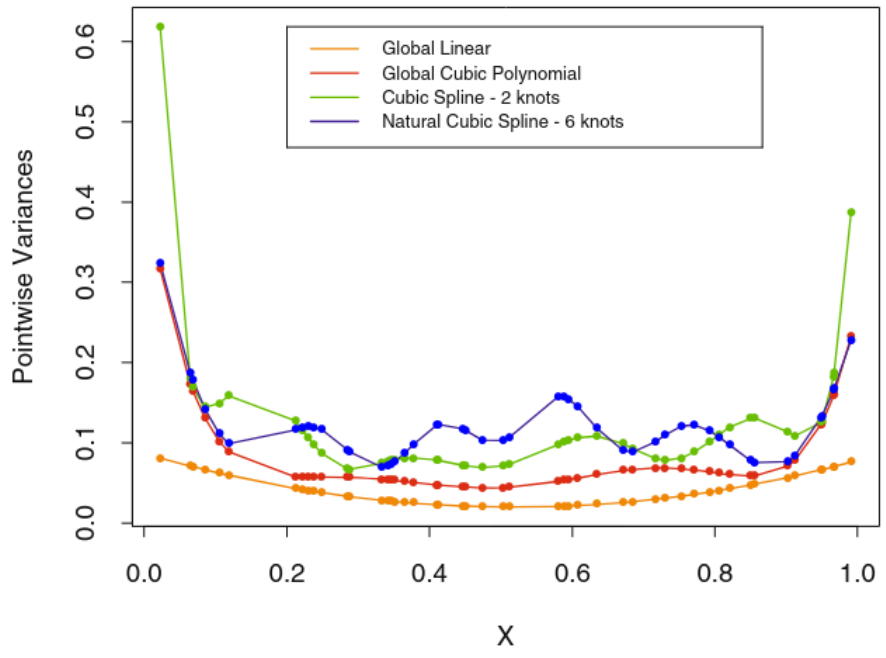
Next question will be how many degree freedoms should our spline contain, the most objective way to do it is using Cross-Validation:

- Removing a portion of the data 20%
- Use the spline to make predictions for the held-out portion
- Compute the **RSS** or any other evaluation metric Evaluation Metrics
- Repeat for different numbers of $knots$
- The value that results in the smallest RSS is chosen



Natural Cubic Splines

It was shown earlier that the behavior of polynomials to be **Wiggly** and unstable near the boundaries, and extrapolation can cause to wrong inference and predictions.



Natural Cubic Spline adds additional constraints on top of the ones existing for **Cubic Spline**

- The function is linear beyond the boundary **knots**

  This free up freedom degrees which can be used on additional knots in the middle region, it's a trade off variance for bias on the boundaries which is acceptable since there is less data and information.

Starting from the **Truncated Power Basis** :

$$h_{K+d}(x) = (x - \tau_k)_+^3 = \begin{cases} (x - \tau_k)^3 & \text{if } x > \tau_k \\ 0 & \text{otherwise} \end{cases}$$

- With τ_k being the location of the the kth knot

This Cubic spline doesn't enforce the **natural constraints** (Linearity on boundaries)

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 (x - \tau_1)^3 + \epsilon_i$$

Using this d helper function define as :

$$d_k(X) = \frac{(x - \tau_k)_+^3 - (x - \tau_K)_+^3}{\tau_K - \tau_k}$$

With :

- τ_K being the last knot on th right
- $(x - \tau_k)_+^3 - (x - \tau_K)_+^3$ is just centering the region τ_k by the last knot τ_K

Now are basis functions for **Natural Spline** are :

$$N_1(x) = 1, \;\; N_2(x) = X, \;\; N_{k+2} = d_k(x) - d_{K-1}(x)$$

This might be confusing by it's made clear when we write :

$$y_i = \beta_0 + \beta_1 N_1(x) + \beta_2 N_2(x) + \beta_3 N_3(x) + \beta_4 N_4(x) + \varepsilon$$

With 4 knots τ.

- When $x < \tau_1$ the first knot : which is the first boundary region

  $$y_i = \beta_0 + \beta_1 N_1(x) + \beta_2 N_2(x) + \varepsilon$$

  a Linear model
- When $x > \tau_4$ the last knot : the second boundary region

  $$y_i = \beta_0 + \beta_1 N_1(x) + \beta_2 N_2(x) + \beta_3 N_3(x) + \beta_4 N_4(x)$$

  Note : For any $d_k(x)$ when $x > \tau_4$, $d_k(x) = 3x^2 + 3x(\tau_k - \tau_4) + (\tau_4^2 + \tau_4 \tau_k + \tau_k^2)$ is a **quadratic**

So :

$$N_3(x) = d_1(x) - d_3(x) = \text{quadratic - quadratic} = \text{Linear}$$

and :

$$N_4(x) = d_2(x) - d_3(x) = \text{quadratic-quadratic=Linear}$$

With N_1, N_2 also being **Linear**
Results in a **Linear Model** when $x > \tau_4$

B-Splines

(random ideas not compete)

B-spline basis are piece-wise polynomial functions of order k, they overlap each other on knots which results in support and **continuity** , like the adjacent ones overlaps

$$B_{i,1}(x) = \begin{cases} 1 & \text{if } \tau_i \le x < \tau_{i+1} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, K + 2M - 1$. (By convention, $B_{i,1} = 0$ if $\tau_i = \tau_{i+1}$).

$$B_{i,m} = \frac{x - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1}(x) + \frac{\tau_{i+m} - x}{\tau_{i+m} - \tau_{i+1}} B_{i+1,m-1}(x)$$

every B-spline is built recursively(cox-de Boor) following the formula above it either take the left or the right support

Changing a coefficient β_j effects only the spline on the small region

$$y_i = \sum^{M} \beta_j B_j(x)$$

- local control
- low variance