ACTL3142 Week 8 - Moving Beyond Linearity

Tadhg Xu-Glassop

2025T2

Polynomial Regression

Polynomial Regression

Allow covariates to be non-linear wrt y by letting it be a polynomial of degree d in x. So,

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_d x_i^d,$$

where β_i are estimated as usual.

This is just linear regression, but adding polynomial terms for our data.

The degree of the polynomial d is a *hyperparameter* that controls the flexibility of the model and thus the bias-variance tradeoff.

Step Functions

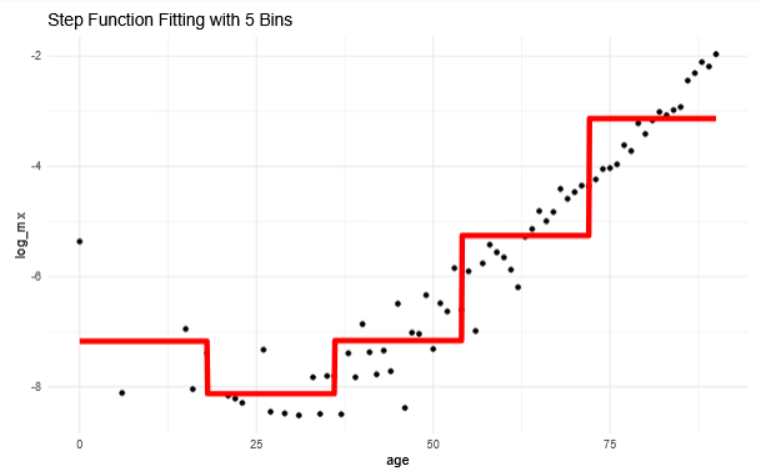
Step Functions

Use indicator functions that take a value of 1 if the covariate is in a specified range. So,

$$y_i = \beta_0 + \beta_1 \mathbb{1}_{\{c_1 \leq x < c_2\}} + \beta_2 \mathbb{1}_{\{c_2 \leq x < c_3\}} + \cdots + \beta_k \mathbb{1}_{\{x \geq c_k\}},$$

where β_i are the average value of x in the specified domain.

The number of breaks / steps is a *hyperparameter* that controls the flexibility of the model, and thus the bias-variance tradeoff.

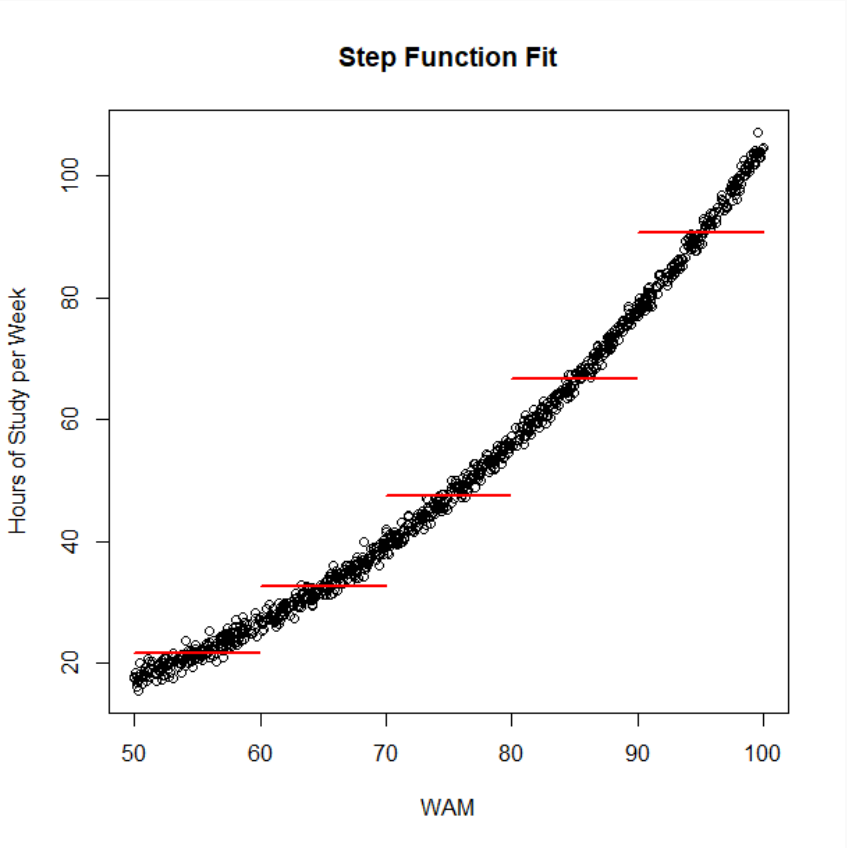


Step Functions demo

Step functions essentially convert numeric features to categorical, so interpretation becomes that of categorical.

$$\text{hours studied} = \beta_0 + \beta_1 \, \mathbb{1}_{\{60 \leq \text{wam} < 70\}} + \cdots + \beta_4 \, \mathbb{1}_{\{90 \leq \text{wam} \leq 100\}}$$

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	21.6987	0.3908	55.53	<2e-16	***
wam_cut(60,70]	10.9830	0.5473	20.07	<2e-16	***
wam_cut(70,80]	25.6675	0.5499	46.68	<2e-16	***
wam_cut(80,90]	45.0599	0.5541	81.33	<2e-16	***
wam_cut(90,100]	68.8462	0.5527	124.57	<2e-16	***

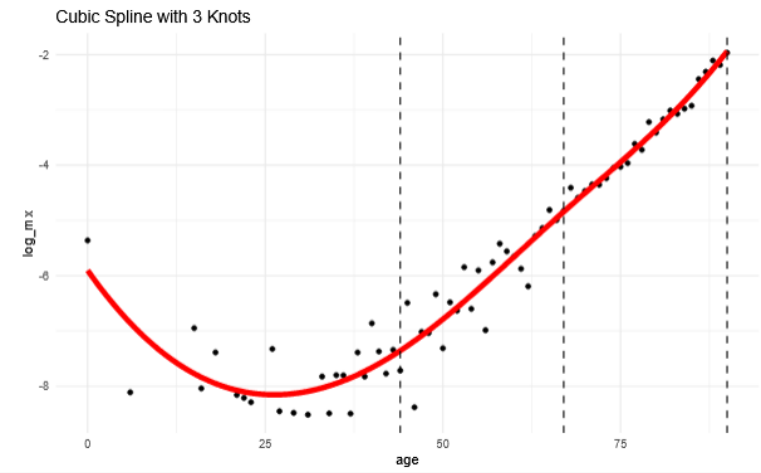


Splines

Regression Splines

1. Split the data into $c + 1$ sections with c 'knots.'
2. Between each knot, fit a d degree regression polynomial **under the restriction** that the entire fit is continuous up until the $d - 1$th derivative.

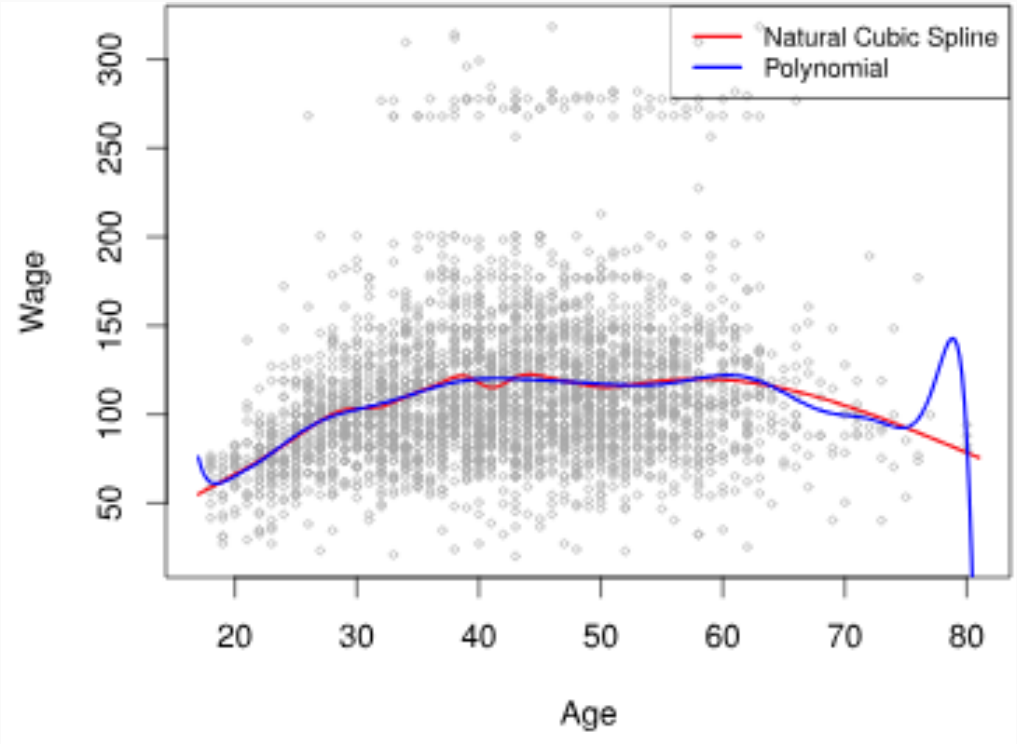
The number of knots c and the degree of the polynomials fitted d are *hyperparameters* which control the bias-variance tradeoff.



Splines (cont.)

Natural Spline

Very similar to a regression spline, but force the fit in the outer knots to be linear to prevent overfitting at extreme values.



Splines (cont.)

Smoothing Spline

A smoothing spline is the model g that minimises:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - g((x_i))^2 + \lambda \int \left(g''(x)\right)^2 \, dx,$$

where λ is a hyperparameter.

Think about $g''(x_i)$ as the concavity / curvature of the graph at x_i, and an integral as an infinitesimal sum. So, the extra penalty is an infitesimal sum of the 'bumpiness' over the graph.

Thus, λ is a *hyperparameter* that controls the flexibility of the model and thus the bias-variance tradeoff.

Local Regression

Local Regresion Method

Given a new observation x_0 which we want to predict y_0 for,

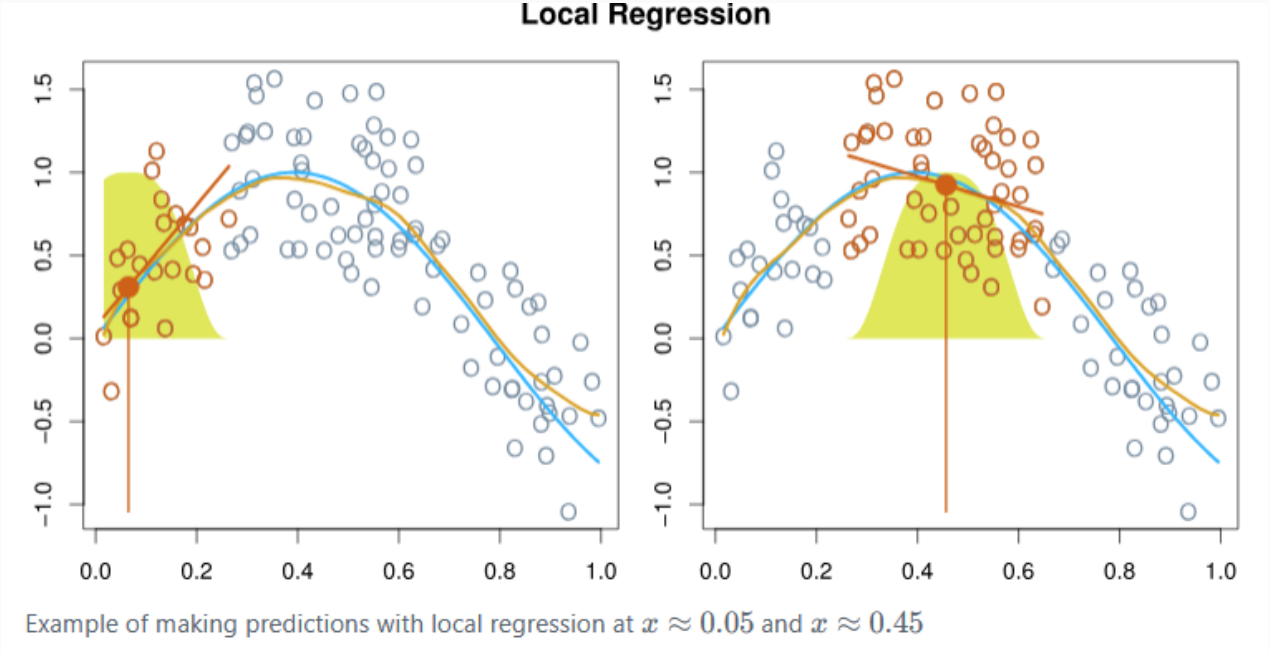
1. Gather the nearest k training points closest to x_0.
2. Assign each of these a weight K_i based upon some kernel K (points closer to x_0 have more weight).
3. Fit a *weighted* least squares with these observations and weights. That is, find β_i that minimises

$$\sum_{i=1}^{k} K_i(y_i - \beta_0 - \beta_1 x_i)^2.$$

4. The predicted value is thus $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

Local Regression (cont.)

The choice of the number of nearest points to consider k, and the kernel K used to assign weights, control the flexibility of the model and thus the bias-variance tradeoff.



Example of making predictions with local regression at $x \approx 0.05$ and $x \approx 0.45$

Generalised Additive Models

Generalised Additive Models

Suppose that

$$y_i = \beta_0 + f_1(x_{i,1}) + f_2(x_{i,2}) + \cdots + f_p(x_{i,p}) + \epsilon_i,$$

where all f_j are any function!

This just says we can combine all the models we have seen so far in this course into one big additive model :)

This works with GLM's as well, where our systematic component is a GAM.