An introduction to GAM(M)s

Stefano Coretta

12/07/2018

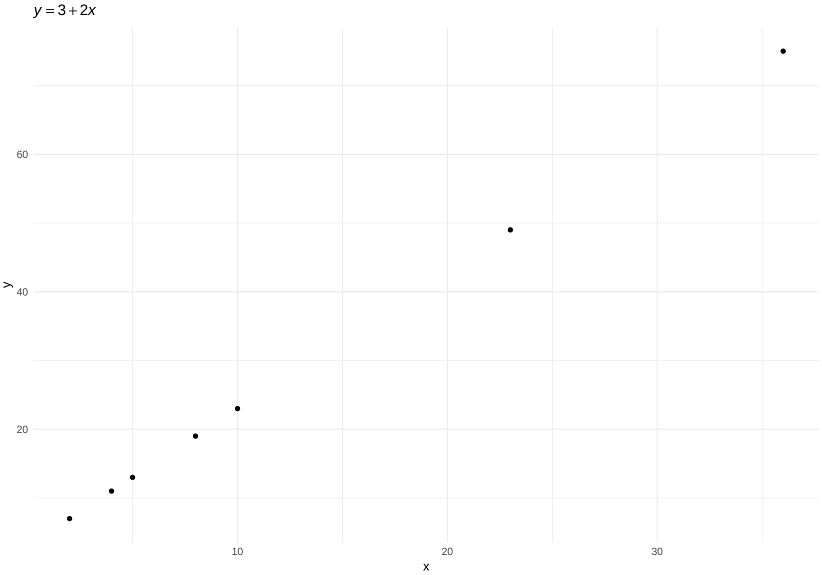
Time travel. . .

Linear models

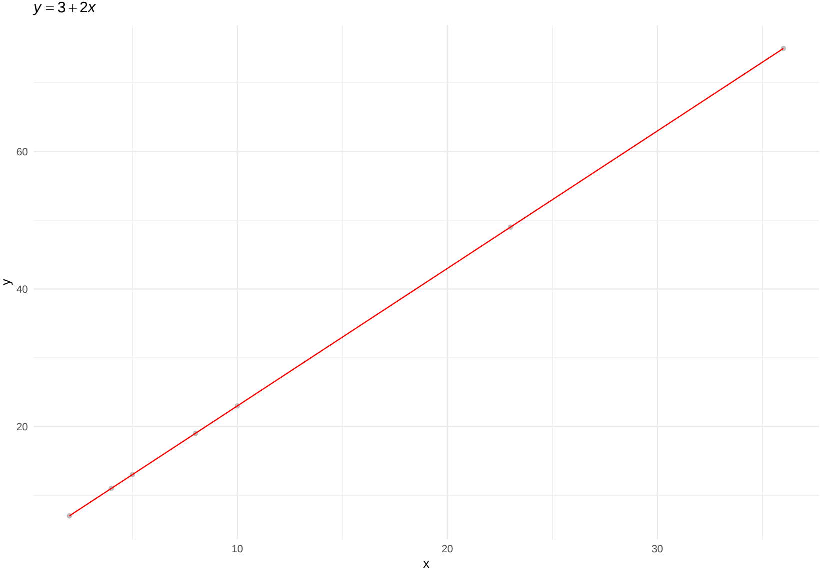
$$y = 3 + 2x$$

where $x = (2, 4, 5, 8, 10, 23, 36)$

Linear models



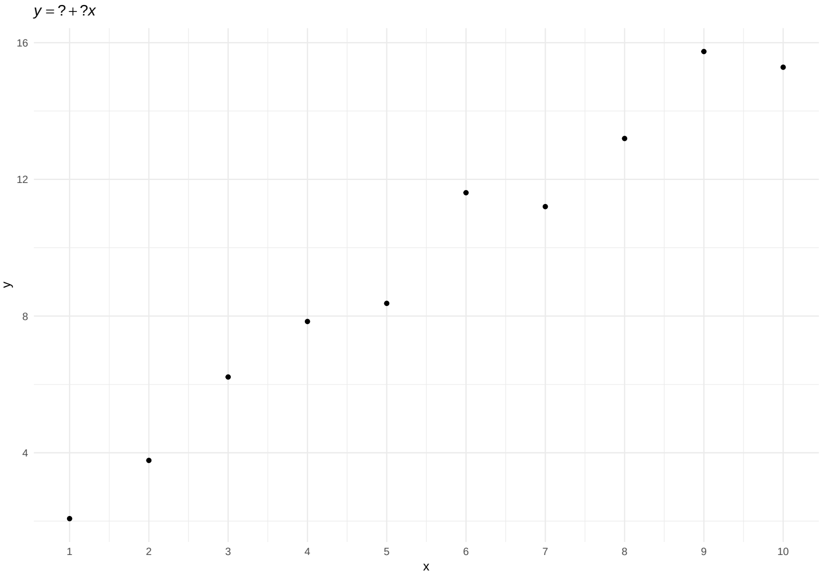
Linear models



Linear models

- In science, we have x and y...
- for example, vowel duration and VOT

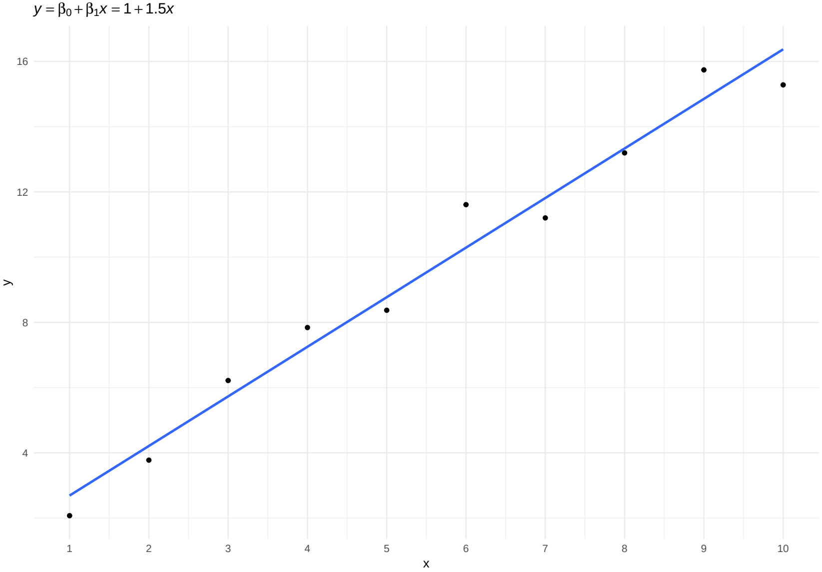
Linear models



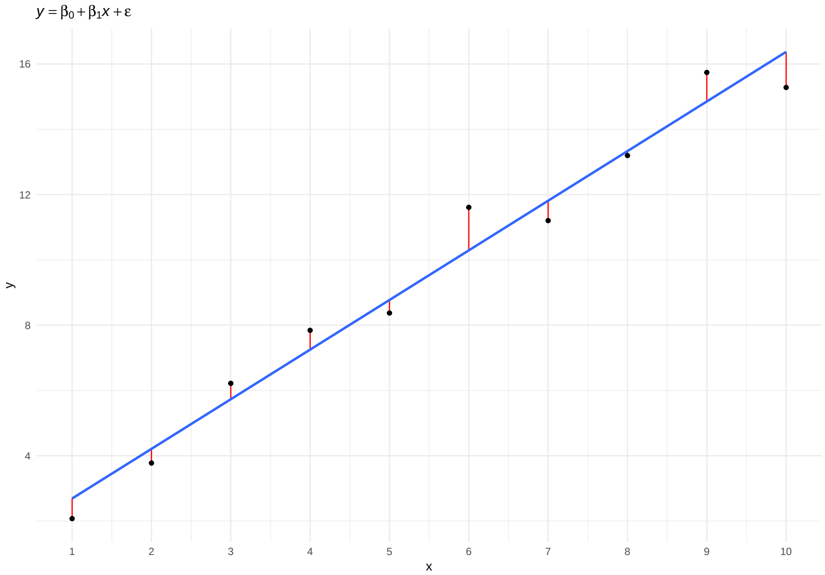
Linear models

- The formula: $y = \beta_0 + \beta_1 x$
  - β_0 is the **intercept**
  - β_1 is the **slope**
- We know x and y
  - we need to estimate β_0, $\beta_1 = \hat{\beta}_0, \hat{\beta}_1$
- We can add more predictors
  - $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$
- `lm(y ~ x, data)` ('y as a function of x')

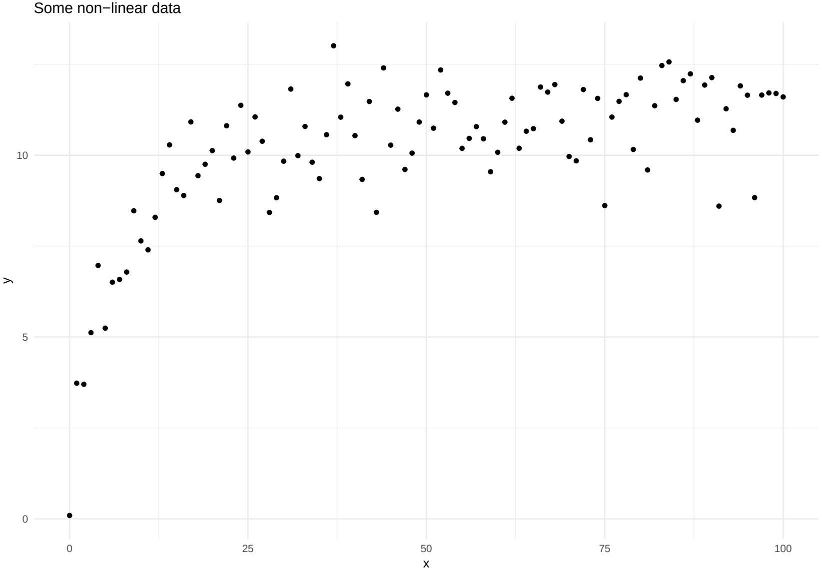
Linear models



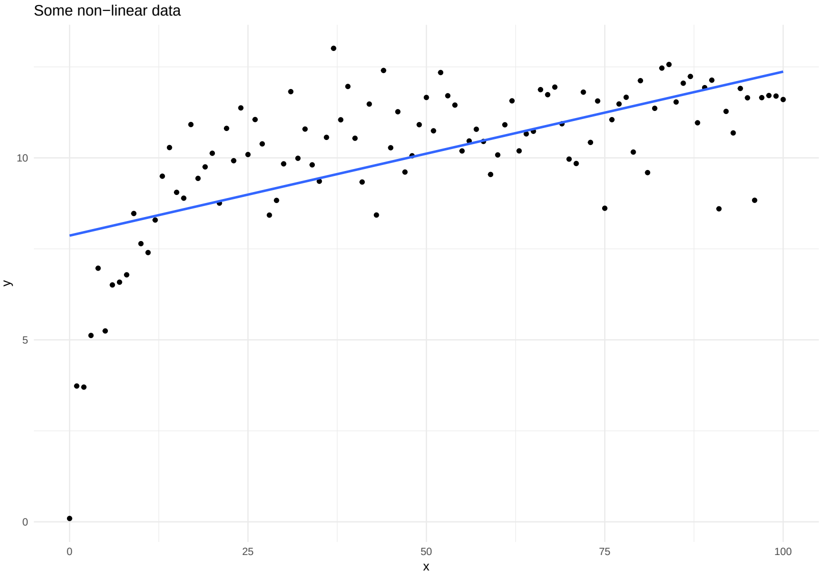
Linear models



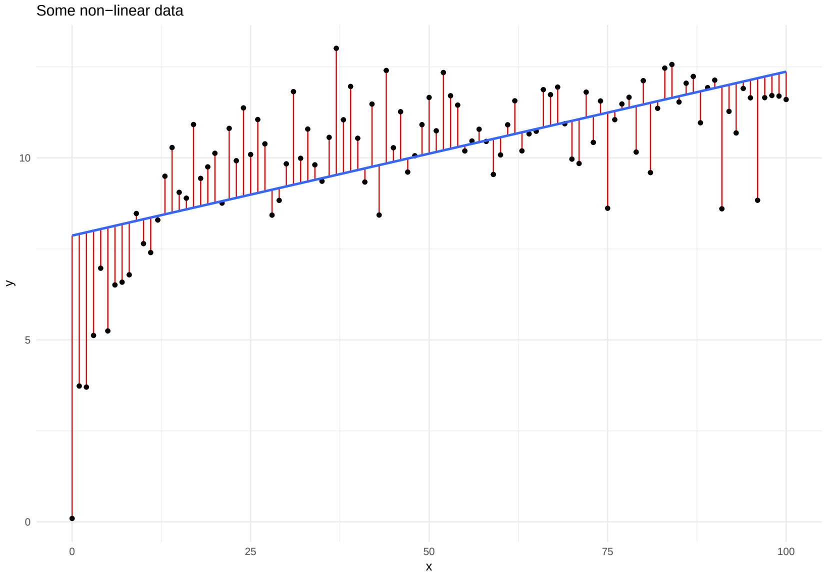
LM with non-linear data



LM with non-linear data



LM with non-linear data

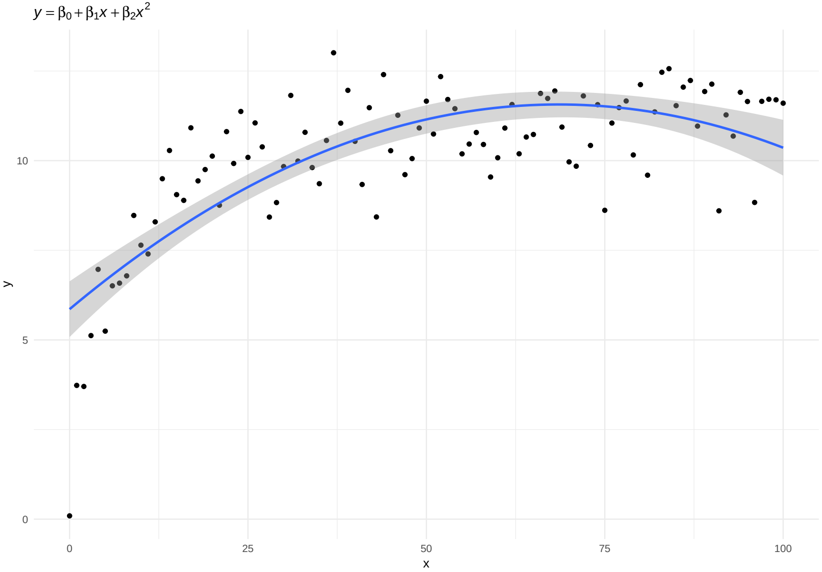


LM with non-linear data

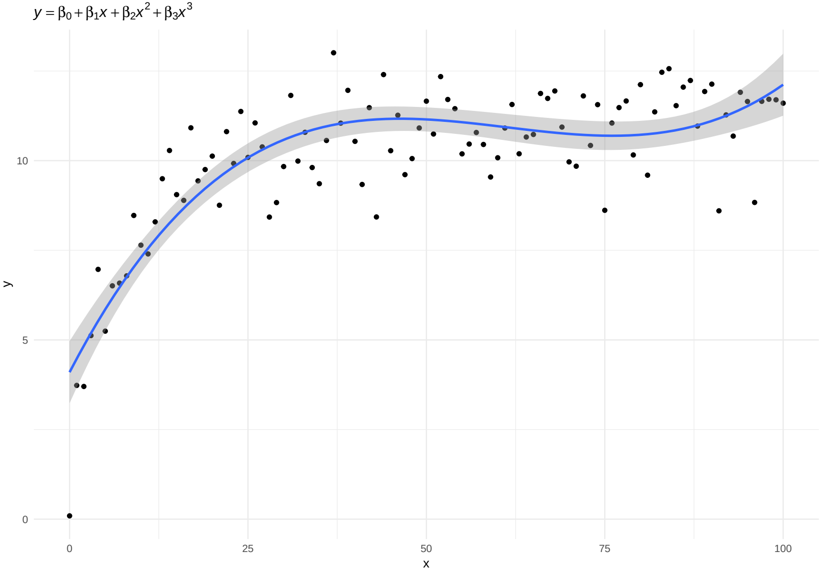
How to account for non-linearity in a linear model?

- Use **higher-degree polynomials**
  - quadratic: $y = \beta_0 + \beta_1 x + \beta_2 x^2$
  - cubic: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$
  - nth: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + ... + \beta_n x^n$

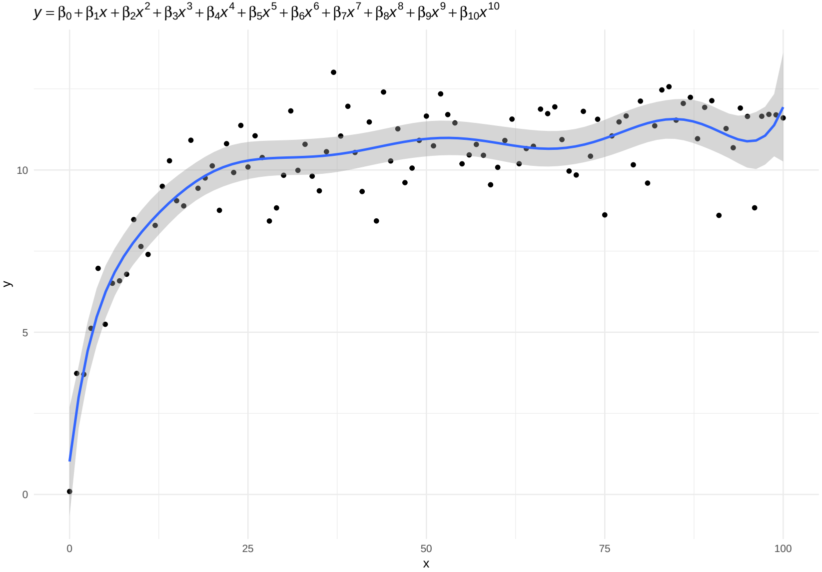
LM with non-linear data



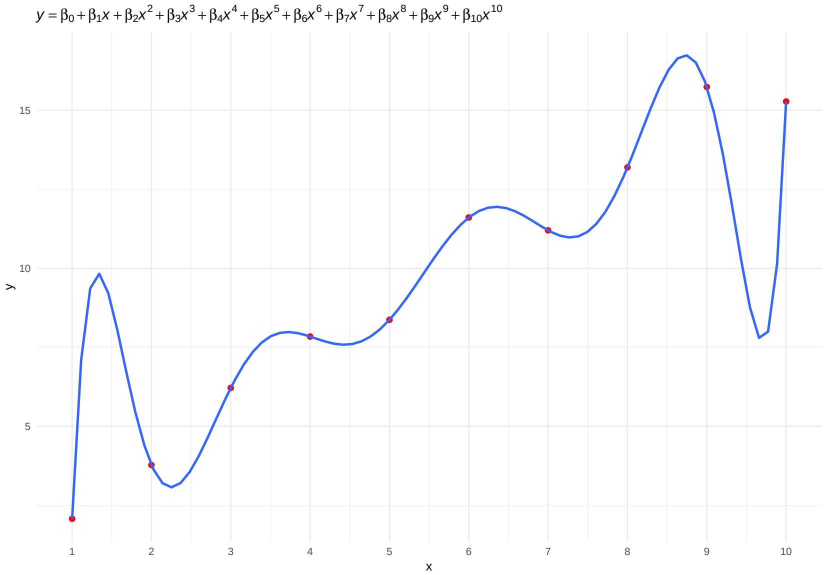
LM with non-linear data



LM with non-linear data



LM with non-linear data



Generalised additive models

- **G**enrealised **A**dditive **M**odels
- $y = f(x) + \epsilon$
  - $f(x) =$ 'some function of x' (or *smooth function*)

Smooth terms

- LMs have **parametric terms**
  - $\beta_n x_n$ (`x` in `R`)
  - linear effects
- GAMs add (non-parametric) **smooth terms** (or simply smooths)
  - $f(x)$, `s(x)` in `R`
  - non-linear effects
- `gam(y ~ s(x), data)`, '*y* as *some* function of *x*'

Basis functions

- polynomials are a type of basis function
  - linear regression is the simplest polynomial (degree 1)
- **smoothing splines** are another type
  - there are several kinds of splines
  - each with their own basis functions