# Polynomial regression

**Elsewhere in the Book**

## Polynomial Regression

Polynomial regressions may resemble non-linear regression in terms of the visual appearance of the regression line (i.e. with bends and curves), but they handle non-linearity by transforming the independent variable $X$ into higher powers (e.g., $X^2$, $X^3$), which are then included in the model along with coefficients that are linear in terms of estimation. For instance, a cubic (of *order*, *degree*, or *power* 3; denoted as $m$) polynomial regression model (**?@fig-plts** A) and is expressed as:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \epsilon_i \tag{1}$$

Where:

- $Y_i$ is the response variable for the $i$-th observation,
- $X_i$ is the predictor variable for the $i$-th observation,
- $\alpha$ is the intercept,
- $\beta_1$, $\beta_2$, and $\beta_3$ are the coefficients for the linear, quadratic, and cubic terms, respectively, and
- $\epsilon_i$ is the error term for the $i$-th observation (the residuals).

It is worth noting that higher order polynomials can lead to overfitting, where the model captures the noise in the data rather than the inherent pattern. This can result in poor generalisation to new data and poor predictive performance. Overfitting becomes more likely as $m$ increases. The $m$ of a polynomial should not exceed $n-1$, where $n$ is the number of data points. An $m$ greater than 4 or 5 is rarely justified.[1] If $m = n-1$, the polynomial will fit the data perfectly (i.e., $R^2 = 1$). For example, a linear regression ($m = 1$) fits two data points exactly, a quadratic regression ($m = 2$) fits three data points perfectly, and so on.

---

[1]The appropriate maximum $m$ can be determined using methods such as the backward-elimination or forward-selection multiple-regression procedure.

Therefore, always consider the trade-off between model complexity and generalisation when using polynomial regression.

Another complication is that the biological interpretation of more complex (higher order) models may be lacking. However, polynomial regression are more often than not used for prediction rather than their interpretability.