

# *Polynomial regression*

Mark Andrews

Psychology Department, Nottingham Trent University

✉ `mark.andrews@ntu.ac.uk`

## *Normal polynomial regression*

- In normal polynomial regression, of degree  $K$ , with one predictor  $x$ , we have

$$y_i \sim N(\mu_i, \sigma^2), \quad \text{for } i \in 1 \dots n$$

$$\mu_i = \beta_0 + \sum_{k=1}^K \beta_k x_i^k.$$

- In other words, we assume that the data are generated as follows:

$$y_i \sim N(f(x_i), \sigma^2), \quad \text{for } i = 1, 2 \dots n,$$

where  $f$  is a nonlinear function that we will approximate with a polynomial of degree  $K$ , i.e. each  $x_i$ , we assume that  $f(x_i)$  can be approximated by  $\sum_{k=0}^K \beta_k x_i^k$  for some unknown values of  $\beta_0, \beta_1 \dots \beta_K$ .

## Polynomials of degree $k \in 0, 1, \dots, 5$

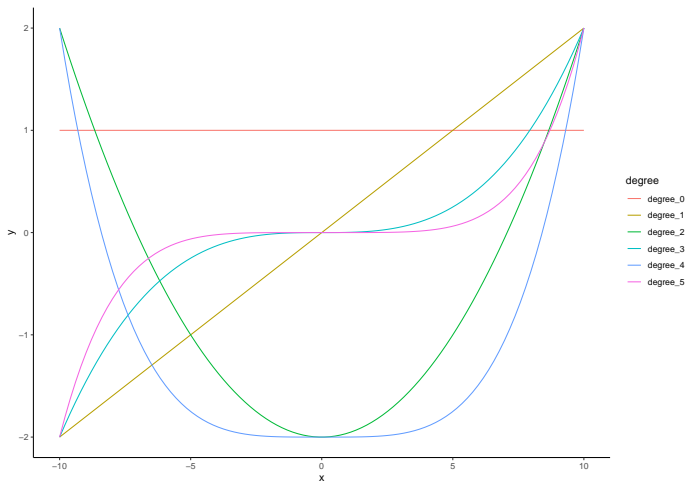


Figure 1: Plots of polynomial functions. For each degree  $k \in 0, 1, \dots, 5$ , we plot  $y = x^k$ .

## Weighted sums of polynomials

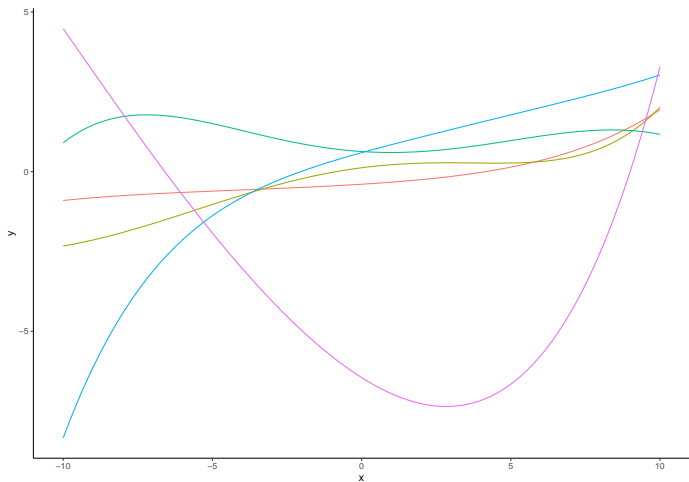


Figure 2: Examples of random polynomial functions. We have five random polynomials of degree  $K = 5$ . In other words, each function shown in each subplot is defined as  $y = \sum_{k=0}^5 \beta_k x^k$  for some random vector  $\beta_0, \beta_1 \dots \beta_5$ .

## Polynomial regression in practice

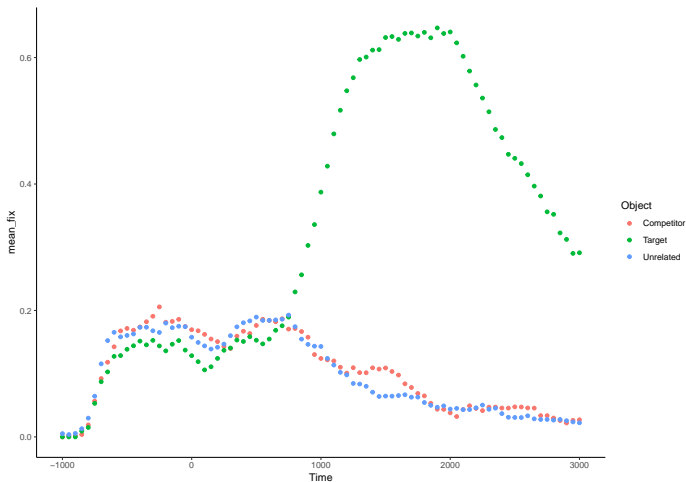


Figure 3: Average proportion of eye fixations at different types of objects (named *Competitor*, *Target*, *Unrelated*) in each time window in multisecond experimental trial.

## Polynomial regression in practice

- ▶ A ninth order polynomial of this data is obtained as follows.
- ▶ This can be compared to competitor model.

```
M_eyefix_null <- lm(mean_fix ~ Object + poly(Time, 9), data=eyefix)
anova(M_eyefix_null, M_eyefix)
#> Analysis of Variance Table
#>
#> Model 1: mean_fix ~ Object + poly(Time, 9)
#> Model 2: mean_fix ~ poly(Time, 9) * Object
#>   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
#> 1      231 3.3021
#> 2      213 0.0470 18      3.2551 819.82 < 2.2e-16 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- ▶ This shows that the polynomial functions for the categories do not differ simply in terms of their intercept terms.
- ▶ However, beyond that, it is not simple matter to say where and how the three different polynomial functions differ from one another.

## *Overfitting in polynomial regression*

- ▶ As the degree of the polynomial model increases, so too can its fit to the data.
- ▶ However, this fit to the data may be essentially an *overfit*.
- ▶ There is no precise general definition of overfitting, but in some cases, it is clearly the case that the overfitted model is fitting the noise rather than the true underlying function, which we know in this case is just a linear model.
- ▶ We can see how the functions in higher order polynomials are bending to fit individual data points.
- ▶ Overfitted models do not generalise well to new data.
- ▶ We may measure this with cross validation, and related methods.

## Cross validation

- ▶ For leave one out cross-validation, the procedure is as follows, with the procedure for any K-fold cross-validation being similarly defined.
- ▶ Assuming our data is  $\mathcal{D} = y_1, y_2 \dots y_n$ , we divide the data into  $n$  sets:

$$(y_1, y_{-1}), (y_2, y_{-2}), \dots (y_i, y_{-i}) \dots (y_n, y_{-n}),$$

where  $y_i$  is data point  $i$  and  $y_{-i}$  is all the remaining data except for data point  $i$ .

- ▶ Then, for each  $i$ , we fit the model using  $y_{-i}$  and test how well the fitted model can predict  $y_i$ .
- ▶ In classical inference based models, for each  $i$ , we calculate  $\hat{\theta}^{-i}$ , which is the maximum likelihood or other estimator of the parameter vector  $\theta$  based on  $y_{-i}$ .
- ▶ We then calculate

$$\text{elpd} = \sum_{i=1}^n \log P(y_i | \hat{\theta}^{-i})$$

as the overall measure of the model's out-of-sample predictive performance, which we refer to as *expected log predictive density*.



# AIC

- ▶ Akaike Information Criterion (AIC) is an approximation to leave one out cross validation.
- ▶ It is defined as follows.

$$\begin{aligned} \text{AIC} &= 2k - 2 \log P(\mathcal{D}|\hat{\theta}), \\ &= 2k + \text{Deviance}, \end{aligned}$$

- ▶ A small sample correction for AIC is

$$\text{AIC}_c = \text{AIC} + \frac{2k(k+1)}{n-k-1}.$$

- ▶ This correction is generally advised when the ratio of sample size  $n$  to number of parameters  $k$  is relatively low, as it would be in the case.

## Overfitting in polynomial regression

- ▶ Overfitting is a general problem in statistical modelling.
- ▶ However, polynomial regression is especially prone to overfitting.
- ▶ This is because higher order polynomials are too unconstrained.
- ▶ The higher the order of the polynomial, the more it can twist and bend to fit the data.
- ▶ This is not always avoided by simply sticking to lower order polynomials because lower order polynomials *underfit* the data, having insufficient flexibility to fit the function.
- ▶ Thus a common problem with polynomial regression is that the lower order polynomials are not flexible enough, and the higher order ones are too unconstrained.
- ▶ Moreover, polynomial regression is also prone to a pathology related *Runge's phenomenon*, which is where there is excessive oscillation in the polynomial function particularly at its edges.