Generalization

MLAI: Week 3

Neil D. Lawrence

Department of Computer Science
Sheffield University

15th October 2013

Review

- Last time: Looked at univariate and multivariate linear regression.
- Showed how to maximize the likelihood of a multivariate model.
- Introduced basis functions to make the model non-linear.

Outline

Basis Functions

Fitting Basis Functions

Generalization

Review: Overdetermined Systems

Underdetermined Systems

Bayesian Perspective

Bayesian Regression

- Problem with Linear Regression—$\mathbf{x}$ may not be linearly related to $\mathbf{y}$.
- Potential solution: create a feature space: define $\phi(\mathbf{x})$ where $\phi(\cdot)$ is a nonlinear function of $\mathbf{x}$.
- Model for target is a linear combination of these nonlinear functions

$$f(\mathbf{x}) = \sum_{j=1}^{K} w_j \phi_j(\mathbf{x}) \tag{1}$$

Quadratic Basis

- Basis functions can be global. E.g. quadratic basis:

$$[1, x, x^2]$$

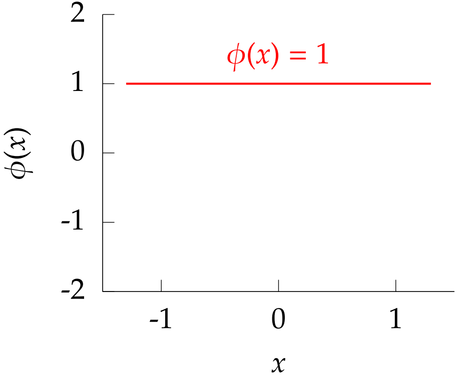


Figure: A quadratic basis.

Quadratic Basis

- Basis functions can be global. E.g. quadratic basis:

$$[1, x, x^2]$$

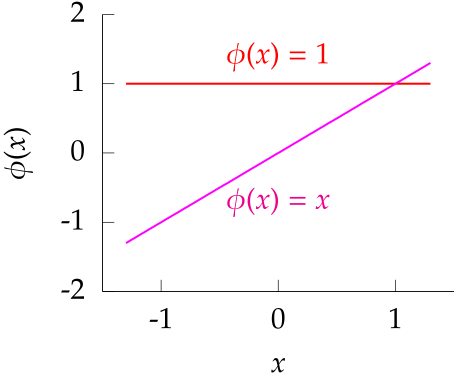


Figure: A quadratic basis.

Quadratic Basis

- Basis functions can be global. E.g. quadratic basis:

$$[1, x, x^2]$$

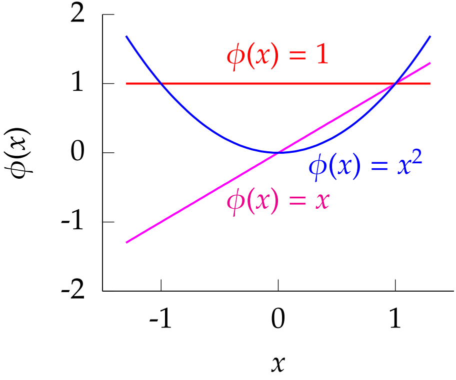


Figure: A quadratic basis.

Functions Derived from Quadratic Basis

$$f(x) = w_1 + w_2 x + w_3 x^2$$

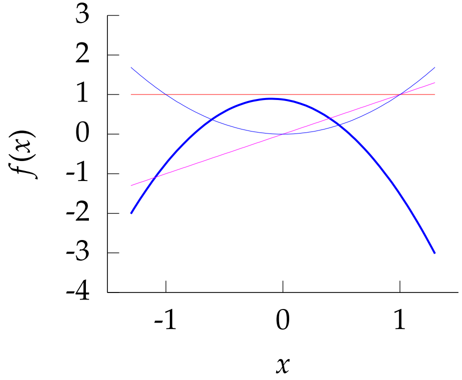


Figure: Function from quadratic basis with weights $w_1 = 0.87466$, $w_2 = -0.38835$, $w_3 = -2.0058$.

Functions Derived from Quadratic Basis

$$f(x) = w_1 + w_2 x + w_3 x^2$$

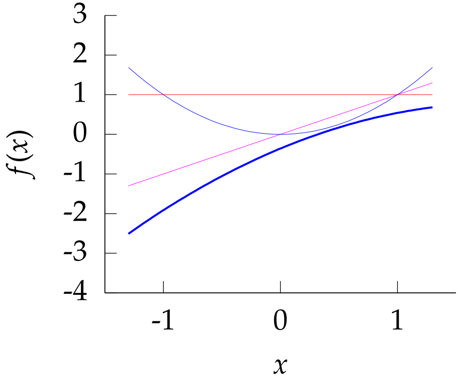


Figure: Function from quadratic basis with weights $w_1 = -0.35908$, $w_2 = 1.2274$, $w_3 = -0.32825$.

Functions Derived from Quadratic Basis

$$f(x) = w_1 + w_2 x + w_3 x^2$$

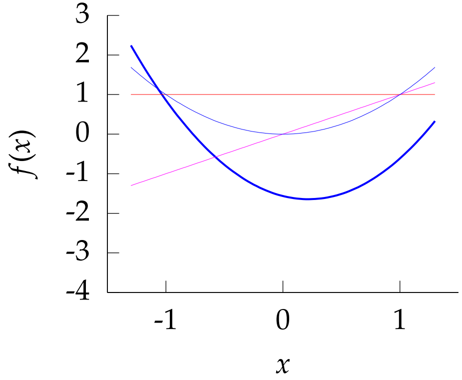


Figure: Function from quadratic basis with weights $w_1 = -1.5638$, $w_2 = -0.73577$, $w_3 = 1.6861$.

Radial Basis Functions

- Or they can be local. E.g. radial (or Gaussian) basis $\phi_j(x) = \exp\left(-\frac{(x-\mu_j)^2}{\ell^2}\right)$

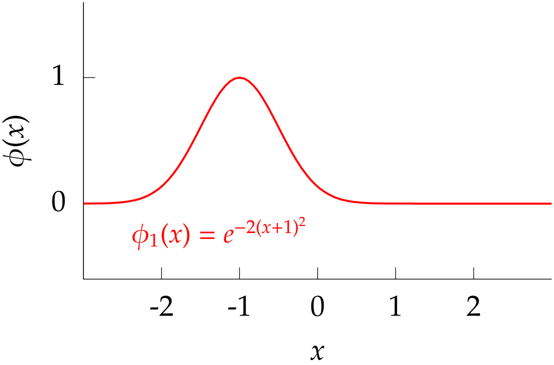


Figure: Radial basis functions.

Radial Basis Functions

- Or they can be local. E.g. radial (or Gaussian) basis $\phi_j(x) = \exp\left(-\frac{(x-\mu_j)^2}{\ell^2}\right)$

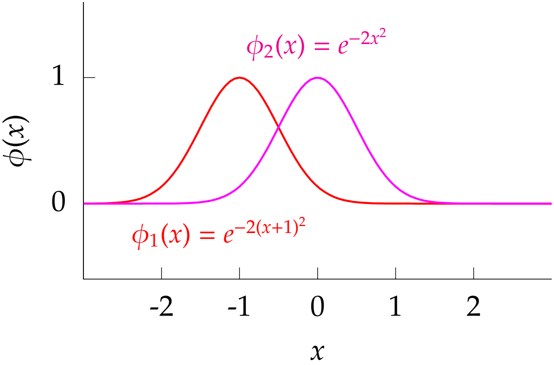


Figure: Radial basis functions.

Radial Basis Functions

- Or they can be local. E.g. radial (or Gaussian) basis $\phi_j(x) = \exp\left(-\frac{(x-\mu_j)^2}{\ell^2}\right)$

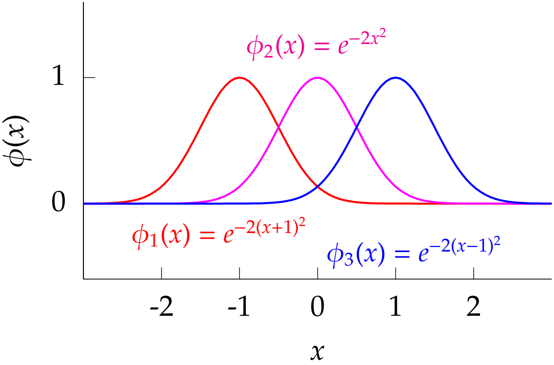


Figure: Radial basis functions.

Functions Derived from Radial Basis

$$f(x) = w_1 e^{-2(x+1)^2} + w_2 e^{-2x^2} + w_3 e^{-2(x-1)^2}$$

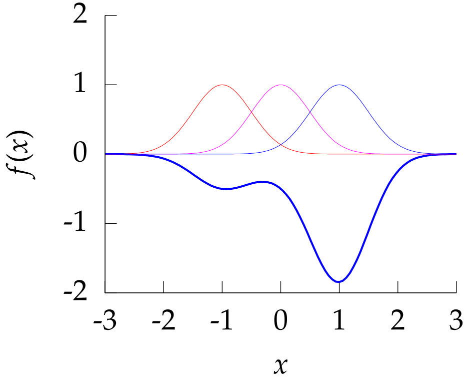


Figure: Function from radial basis with weights $w_1 = -0.47518$, $w_2 = -0.18924$, $w_3 = -1.8183$.

Functions Derived from Radial Basis

$$f(x) = w_1 e^{-2(x+1)^2} + w_2 e^{-2x^2} + w_3 e^{-2(x-1)^2}$$

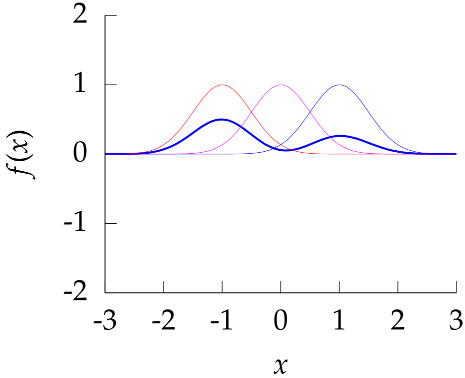


Figure: Function from radial basis with weights $w_1 = 0.50596$, $w_2 = -0.046315$, $w_3 = 0.26813$.

Functions Derived from Radial Basis

$$f(x) = w_1 e^{-2(x+1)^2} + w_2 e^{-2x^2} + w_3 e^{-2(x-1)^2}$$

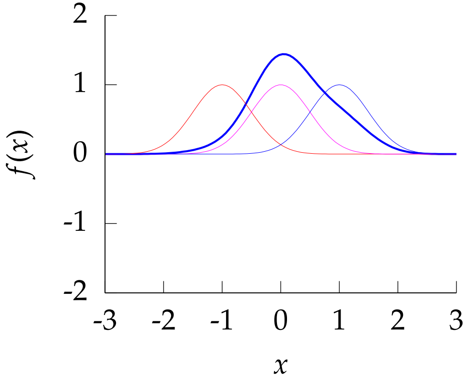


Figure: Function from radial basis with weights $w_1 = 0.07179$, $w_2 = 1.3591$, $w_3 = 0.50604$.

Outline

Basis Function Models

- A Basis function mapping is now defined as

$$f(\mathbf{x}_i) = \sum_{j=1}^{m} w_j \phi_{i,j} + c$$

Vector Notation

- Write in vector notation,

$$f(\mathbf{x}_i) = \mathbf{w}^\top \boldsymbol{\phi}_i + c$$

Log Likelihood for Basis Function Model

- The likelihood of a single data point is

$$p(y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mathbf{w}^\top \boldsymbol{\phi}_i)^2}{2\sigma^2}\right).$$

- Leading to a log likelihood for the data set of

$$L(\mathbf{w}, \sigma^2) = -\frac{N}{2}\log\sigma^2 - \frac{N}{2}\log 2\pi - \frac{\sum_{i=1}^{N}(y_i - \mathbf{w}^\top \boldsymbol{\phi}_i)^2}{2\sigma^2}.$$

- And a corresponding error function of

$$E(\mathbf{w}, \sigma^2) = \frac{N}{2}\log\sigma^2 + \frac{\sum_{i=1}^{N}(y_i - \mathbf{w}^\top \boldsymbol{\phi}_i)^2}{2\sigma^2}.$$

Expand the Brackets

$$
\begin{aligned}
E(\mathbf{w}, \sigma^2) = & \frac{N}{2} \log \sigma^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{N} y_i^2 - \frac{1}{\sigma^2} \sum_{i=1}^{N} y_i \mathbf{w}^\top \boldsymbol{\phi}_i \\
& + \frac{1}{2\sigma^2} \sum_{i=1}^{N} \mathbf{w}^\top \boldsymbol{\phi}_i \boldsymbol{\phi}_i^\top \mathbf{w} + \text{const.} \\
= & \frac{N}{2} \log \sigma^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{N} y_i^2 - \frac{1}{\sigma^2} \mathbf{w}^\top \sum_{i=1}^{N} \boldsymbol{\phi}_i y_i \\
& + \frac{1}{2\sigma^2} \mathbf{w}^\top \left[\sum_{i=1}^{N} \boldsymbol{\phi}_i \boldsymbol{\phi}_i^\top \right] \mathbf{w} + \text{const.}
\end{aligned}
$$

Multivariate Derivatives Reminder

- We will need some multivariate calculus.

$$\frac{\mathrm{d}\mathbf{a}^\top \mathbf{w}}{\mathrm{d}\mathbf{w}} = \mathbf{a}$$

and

$$\frac{\mathrm{d}\mathbf{w}^\top \mathbf{A}\mathbf{w}}{\mathrm{d}\mathbf{w}} = \left(\mathbf{A} + \mathbf{A}^\top\right)\mathbf{w}$$

or if $\mathbf{A}$ is symmetric (*i.e.* $\mathbf{A} = \mathbf{A}^\top$)

$$\frac{\mathrm{d}\mathbf{w}^\top \mathbf{A}\mathbf{w}}{\mathrm{d}\mathbf{w}} = 2\mathbf{A}\mathbf{w}.$$

Differentiate

Differentiating with respect to the vector **w** we obtain

$$\frac{\partial L(\mathbf{w}, \beta)}{\partial \mathbf{w}} = \beta \sum_{i=1}^{N} \phi_i y_i - \beta \left[\sum_{i=1}^{N} \phi_i \phi_i^\top \right] \mathbf{w}$$

Leading to

$$\mathbf{w}^* = \left[\sum_{i=1}^{N} \phi_i \phi_i^\top \right]^{-1} \sum_{i=1}^{N} \phi_i y_i,$$

Rewrite in matrix notation:

$$\sum_{i=1}^{N} \phi_i \phi_i^\top = \mathbf{\Phi}^\top \mathbf{\Phi}$$

$$\sum_{i=1}^{N} \phi_i y_i = \mathbf{\Phi}^\top \mathbf{y}$$

Update Equations

- Update for $\mathbf{w}^*$.

$$\mathbf{w}^* = \left(\boldsymbol{\Phi}^\top \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^\top \mathbf{y}$$

- The equation for ${\sigma^2}^*$ may also be found

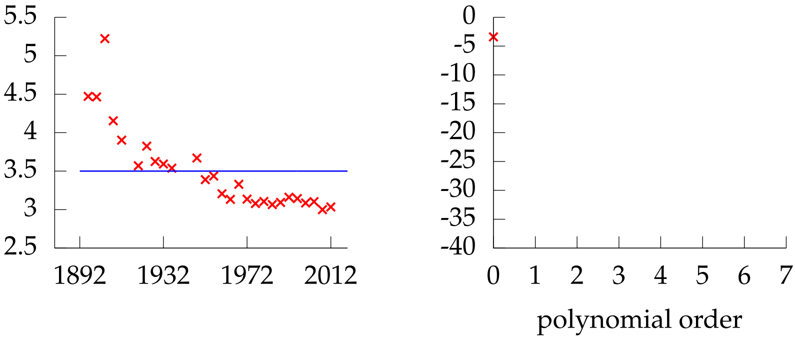
$$ {\sigma^2}^* = \frac{\sum_{i=1}^{N} \left(y_i - {\mathbf{w}^*}^\top \boldsymbol{\phi}_i\right)^2}{N}. $$

Reading

- Section 1.4 of Rogers and Girolami.
- Chapter 1, pg 1-6 of Bishop.
- Chapter 3, Section 3.1 of Bishop up to pg 143.

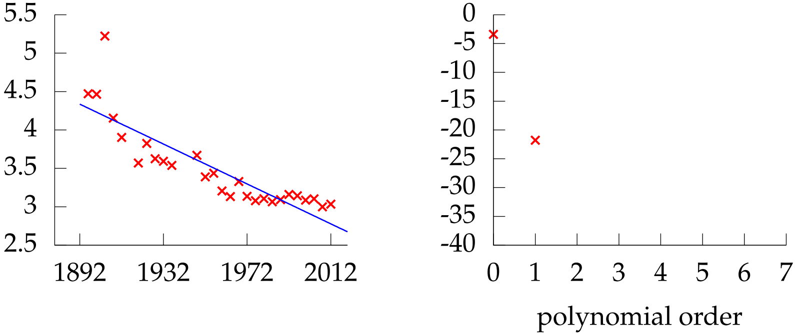
Outline

Polynomial Fits to Olympics Data



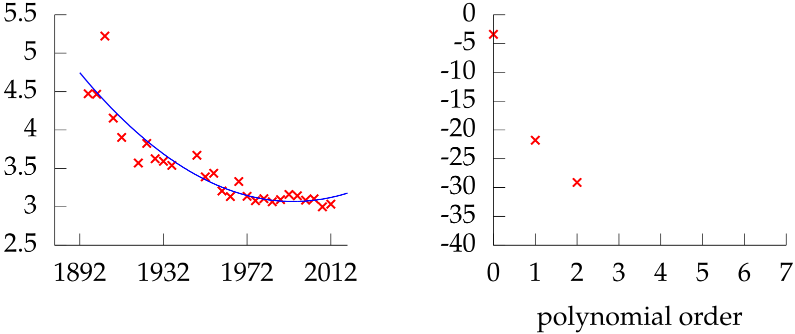
Left: fit to data, *Right*: model error. Polynomial order 0, model error -3.3989, $\sigma^2 = 0.286$, $\sigma = 0.535$.

Polynomial Fits to Olympics Data



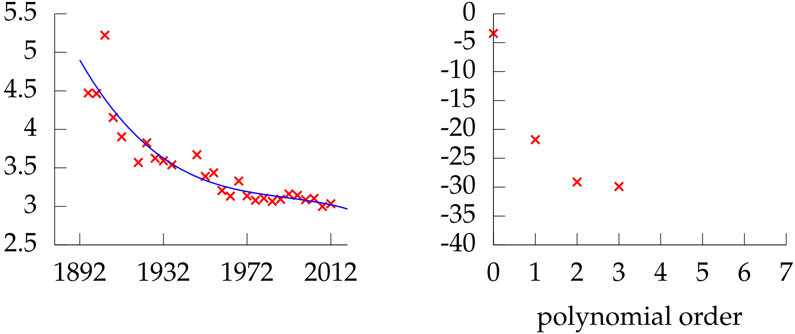
Left: fit to data, *Right*: model error. Polynomial order 1, model error -21.772, $\sigma^2 = 0.0733$, $\sigma = 0.271$.

Polynomial Fits to Olympics Data



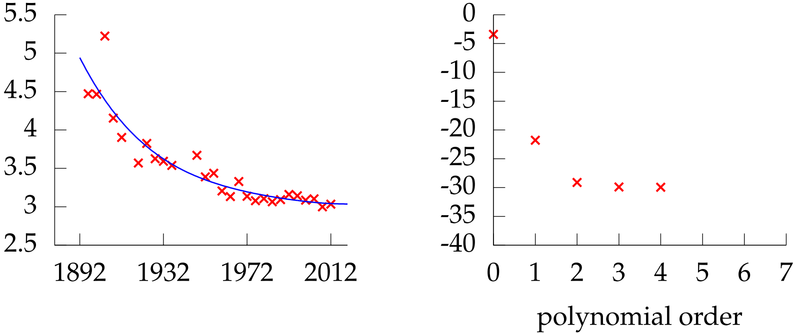
Left: fit to data, *Right*: model error. Polynomial order 2, model error -29.101, $\sigma^2 = 0.0426$, $\sigma = 0.206$.

Polynomial Fits to Olympics Data



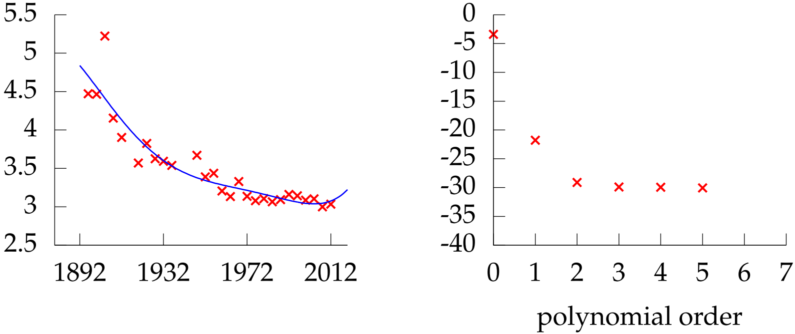
Left: fit to data, *Right*: model error. Polynomial order 3, model error -29.907, $\sigma^2 = 0.0401$, $\sigma = 0.200$.

Polynomial Fits to Olympics Data



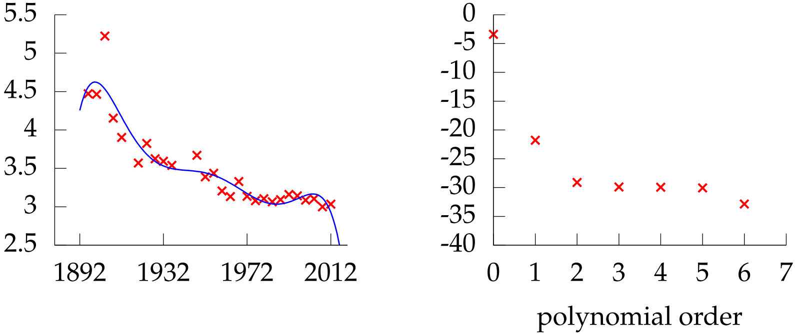
Left: fit to data, *Right*: model error. Polynomial order 4, model error -29.943, $\sigma^2 = 0.0400$, $\sigma = 0.200$.

Polynomial Fits to Olympics Data



Left: fit to data, *Right*: model error. Polynomial order 5, model error -30.056, $\sigma^2 = 0.0397$, $\sigma = 0.199$.

Polynomial Fits to Olympics Data



Left: fit to data, *Right*: model error. Polynomial order 6, model error -32.866, $\sigma^2 = 0.0322$, $\sigma = 0.180$.

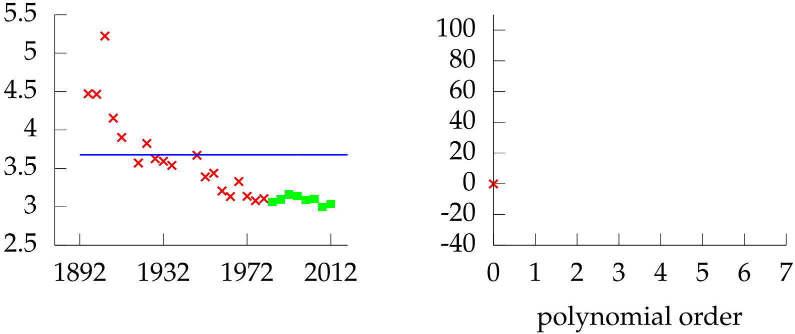
Overfitting

- Increase number of basis functions, we obtain a better ‘fit’ to the data.
- How will the model perform on previously unseen data?

Training and Test Sets

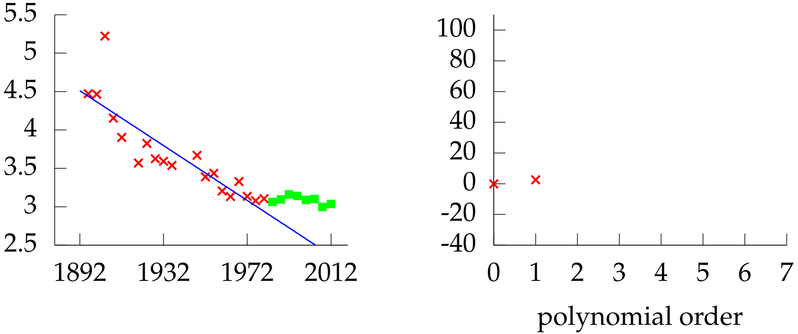
- We call the data used for fitting the model the 'training set'.
- Data not used for training, but when the model is applied 'in the field' is called the 'test data'.
- Challenge for generalization is to ensure a good performance on test data given only training data.

Validation Set



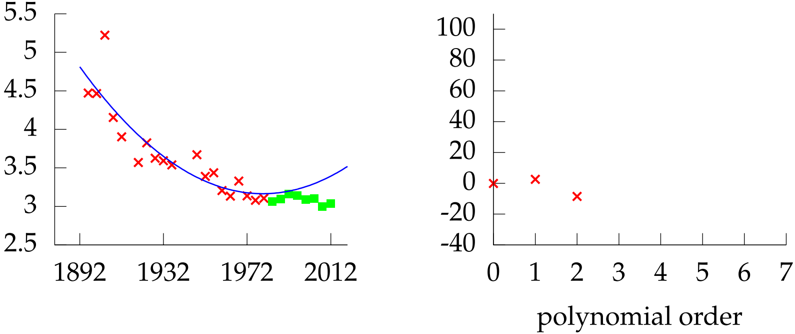
Left: fit to data, *Right*: model error. Polynomial order 0, training error -1.8774, validation error -0.13132, $\sigma^2 = 0.302$, $\sigma = 0.549$.

Validation Set



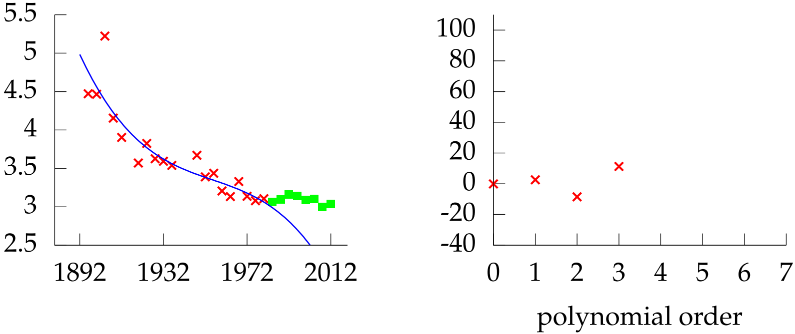
Left: fit to data, *Right*: model error. Polynomial order 1, training error -15.325, validation error 2.5863, $\sigma^2 = 0.0733$, $\sigma = 0.271$.

Validation Set



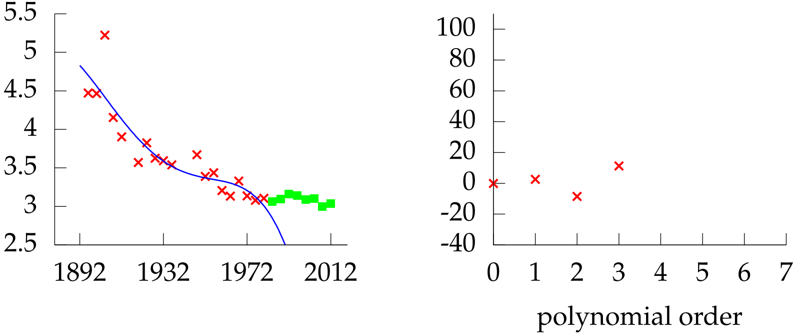
Left: fit to data, *Right*: model error. Polynomial order 2, training error -17.579, validation error -8.4831, σ^2 = 0.0578, σ = 0.240.

Validation Set



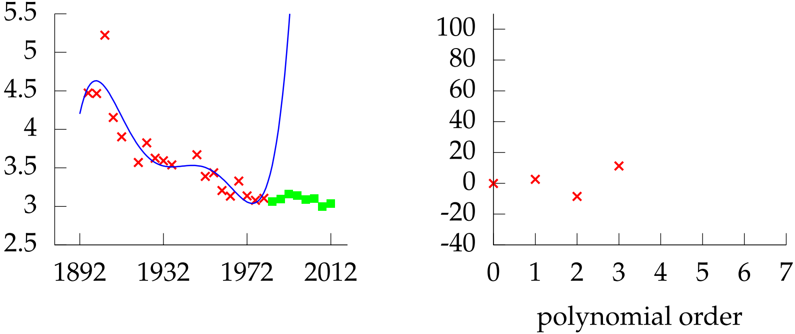
Left: fit to data, *Right*: model error. Polynomial order 3, training error -18.064, validation error 11.27, $\sigma^2 = 0.0549$, $\sigma = 0.234$.

Validation Set



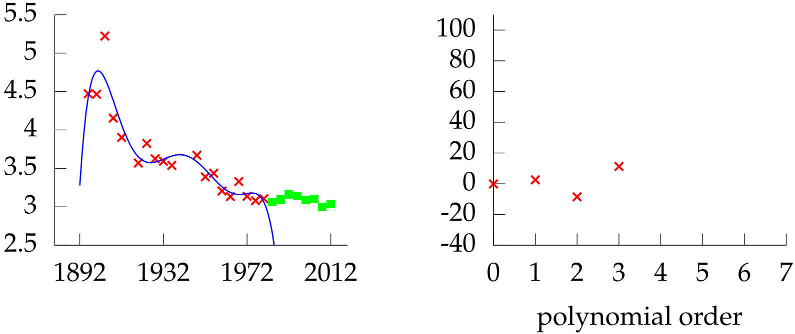
Left: fit to data, *Right*: model error. Polynomial order 4, training error -18.245, validation error 232.92, $\sigma^2 = 0.0539$, $\sigma = 0.232$.

Validation Set



Left: fit to data, *Right*: model error. Polynomial order 5, training error -20.471, validation error 9898.1, $\sigma^2 = 0.0426$, $\sigma = 0.207$.

Validation Set

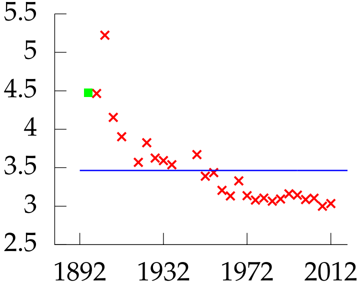


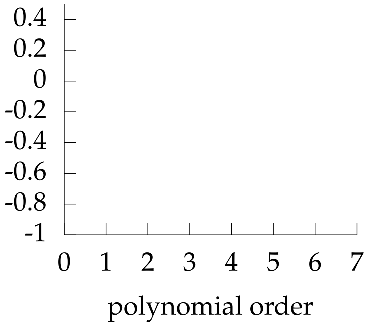
Left: fit to data, *Right*: model error. Polynomial order 6, training error -22.881, validation error 67775, $\sigma^2 = 0.0331$, $\sigma = 0.182$.

Leave One Out Error

- Take training set and remove one point.
- Train on the remaining data.
- Compute the error on the point you removed (which wasn't in the training data).
- Do this for each point in the training set in turn.
- Average the resulting error. This is the leave one out error.

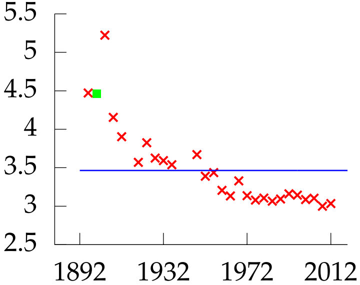
Leave One Out Error

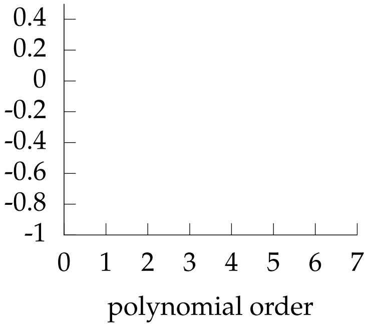




Polynomial order 0, training error -3.346, leave one out error 0.045811.

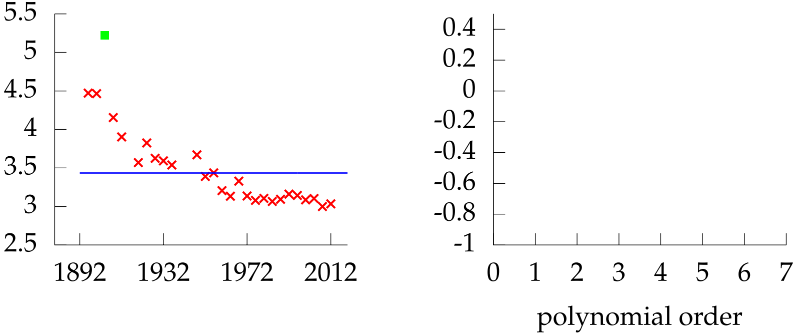
Leave One Out Error





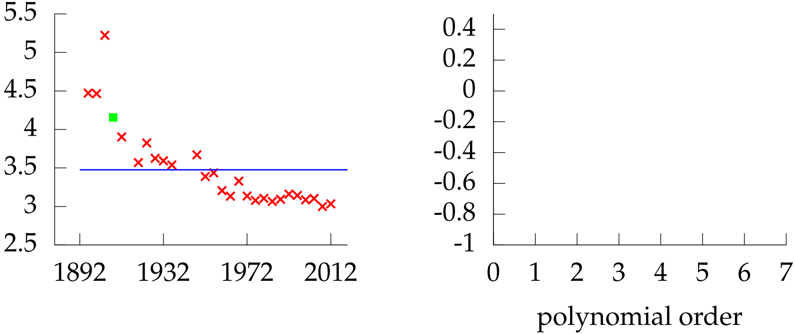
Polynomial order 0, training error -3.346, leave one out error 0.045811.

Leave One Out Error



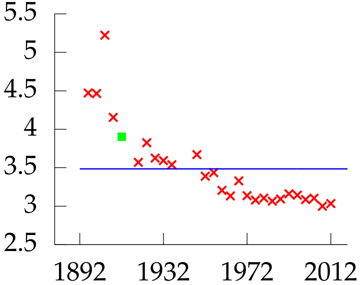
Polynomial order 0, training error -3.346, leave one out error 0.045811.

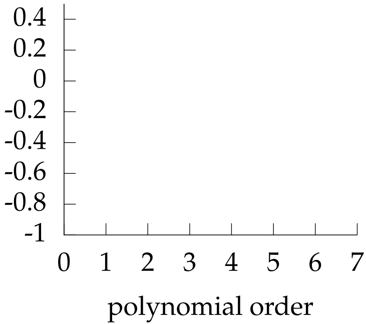
Leave One Out Error



Polynomial order 0, training error -3.346, leave one out error 0.045811.

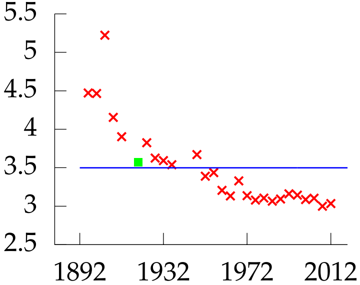
Leave One Out Error

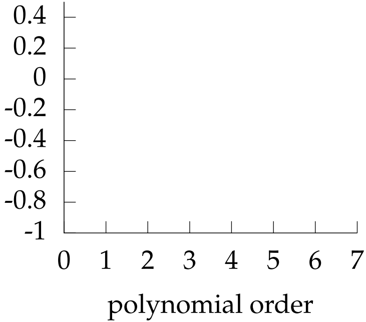




Polynomial order 0, training error -3.346, leave one out error 0.045811.

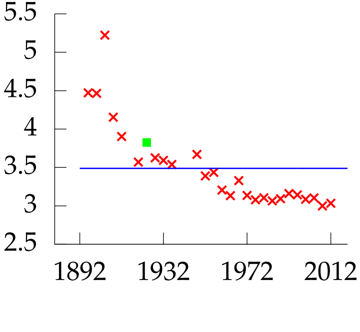
Leave One Out Error





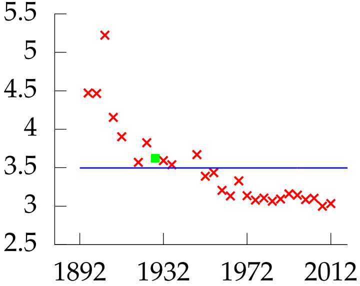
Polynomial order 0, training error -3.346, leave one out error 0.045811.

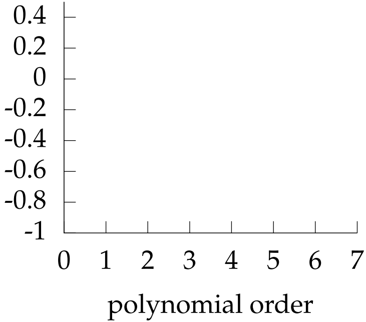
Leave One Out Error



Polynomial order 0, training error -3.346, leave one out error 0.045811.

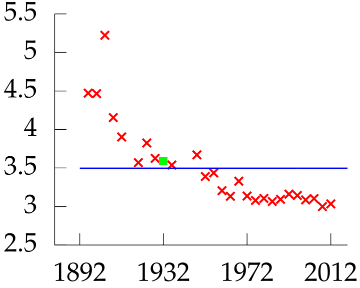
Leave One Out Error

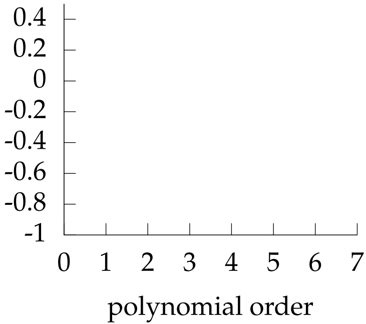




Polynomial order 0, training error -3.346, leave one out error 0.045811.

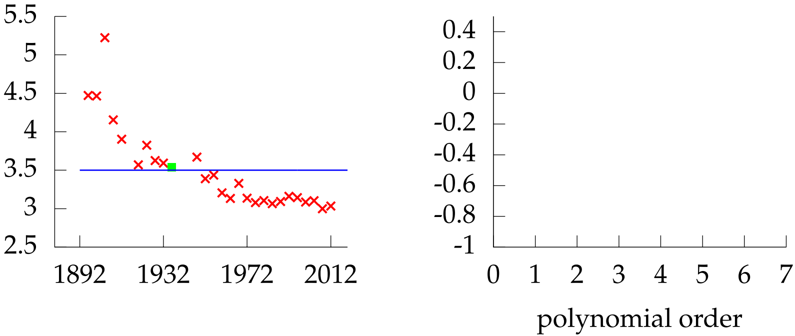
Leave One Out Error





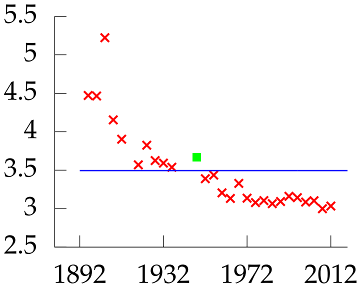
Polynomial order 0, training error -3.346, leave one out error 0.045811.

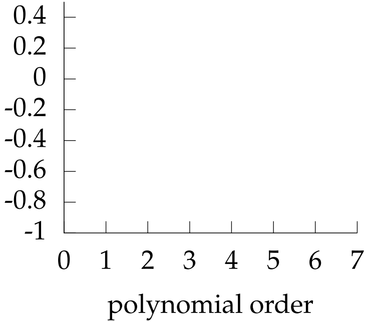
Leave One Out Error



Polynomial order 0, training error -3.346, leave one out error 0.045811.

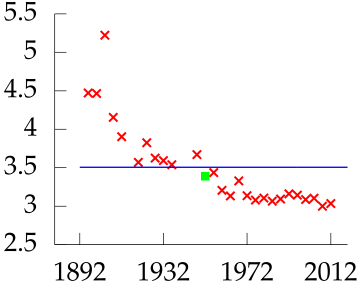
Leave One Out Error

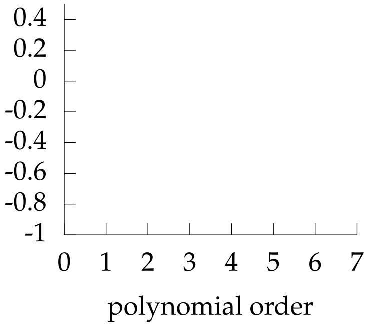




Polynomial order 0, training error -3.346, leave one out error 0.045811.

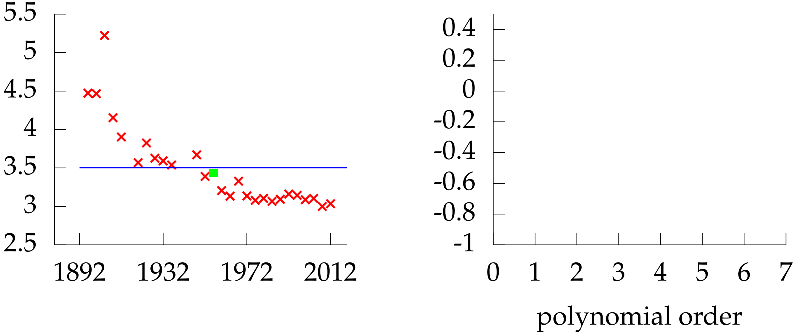
Leave One Out Error





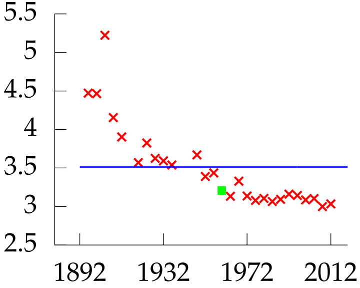
Polynomial order 0, training error -3.346, leave one out error 0.045811.

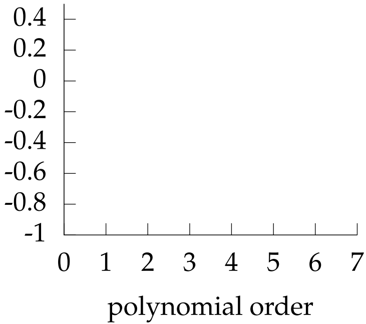
Leave One Out Error



Polynomial order 0, training error -3.346, leave one out error 0.045811.

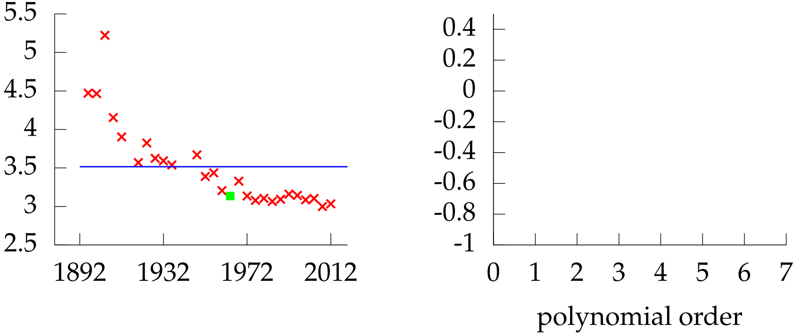
Leave One Out Error





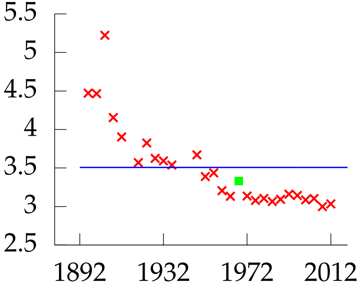
Polynomial order 0, training error -3.346, leave one out error 0.045811.

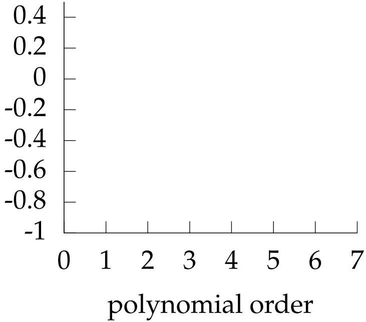
Leave One Out Error



Polynomial order 0, training error -3.346, leave one out error 0.045811.

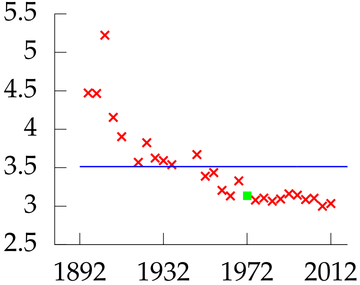
Leave One Out Error

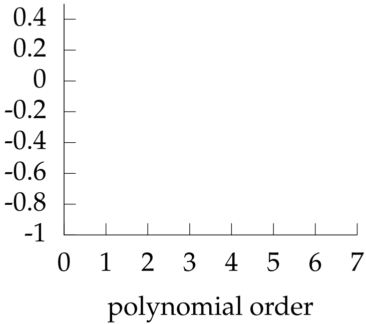




Polynomial order 0, training error -3.346, leave one out error 0.045811.

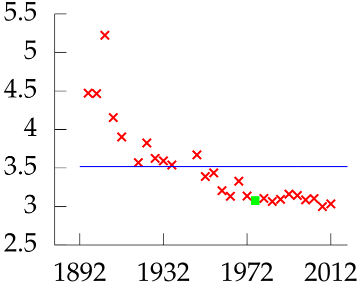
Leave One Out Error

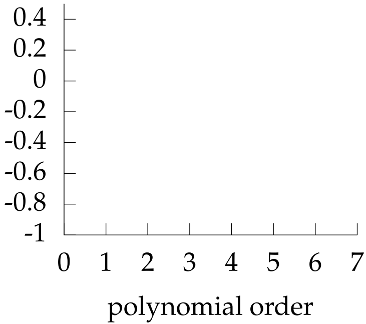




Polynomial order 0, training error -3.346, leave one out error 0.045811.

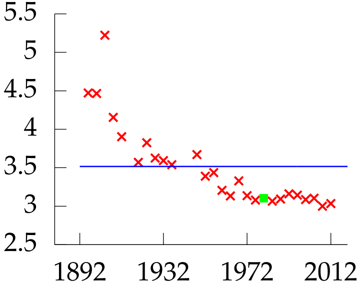
Leave One Out Error

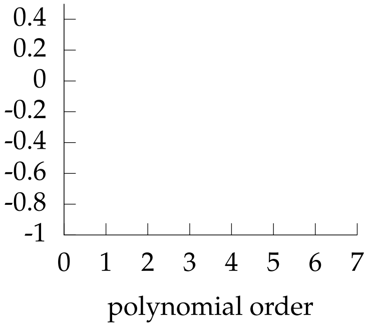




Polynomial order 0, training error -3.346, leave one out error 0.045811.

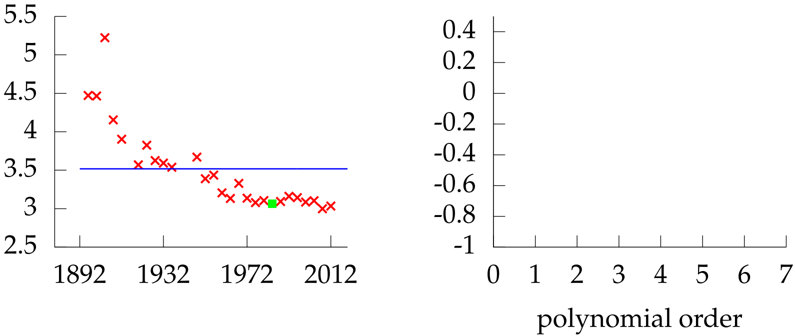
Leave One Out Error





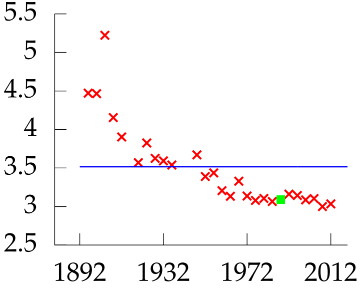
Polynomial order 0, training error -3.346, leave one out error 0.045811.

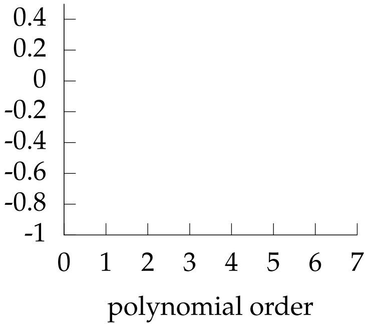
Leave One Out Error



Polynomial order 0, training error -3.346, leave one out error 0.045811.

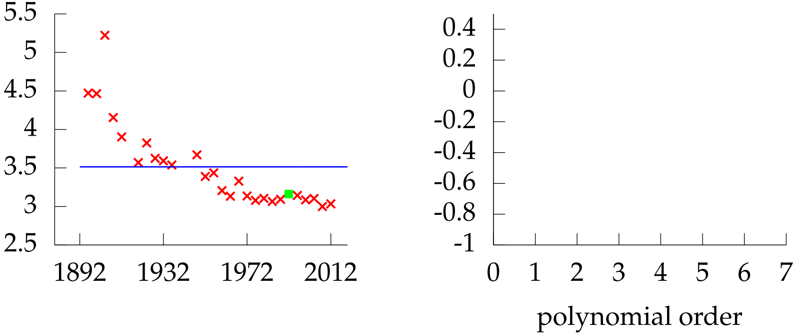
Leave One Out Error





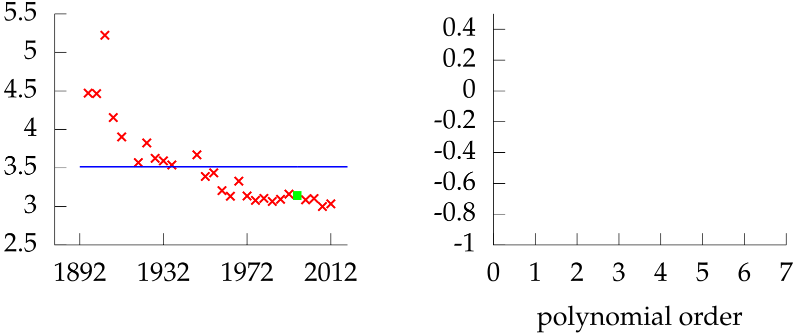
Polynomial order 0, training error -3.346, leave one out error 0.045811.

Leave One Out Error



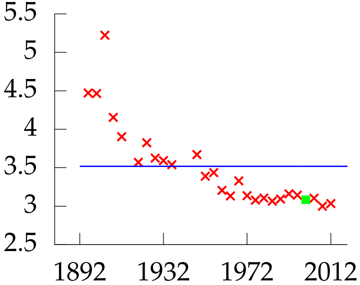
Polynomial order 0, training error -3.346, leave one out error 0.045811.

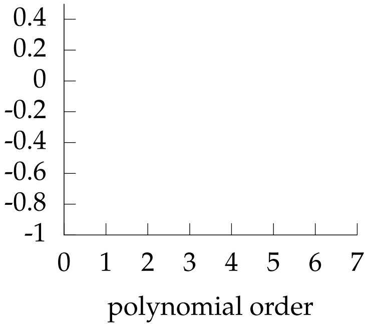
Leave One Out Error



Polynomial order 0, training error -3.346, leave one out error 0.045811.

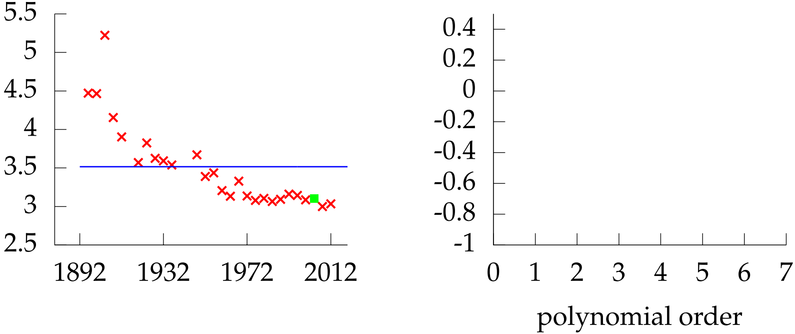
Leave One Out Error





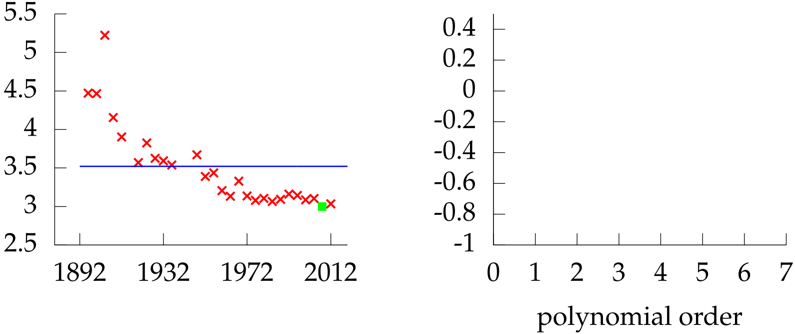
Polynomial order 0, training error -3.346, leave one out error 0.045811.

Leave One Out Error



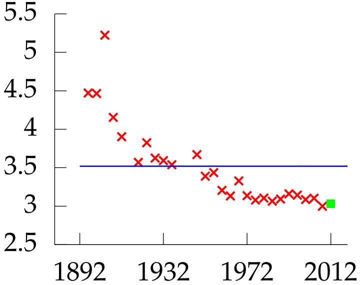
Polynomial order 0, training error -3.346, leave one out error 0.045811.

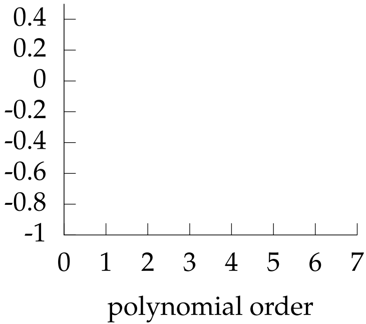
Leave One Out Error



Polynomial order 0, training error -3.346, leave one out error 0.045811.

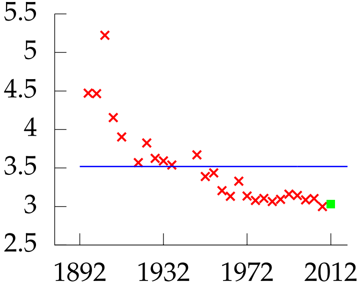
Leave One Out Error

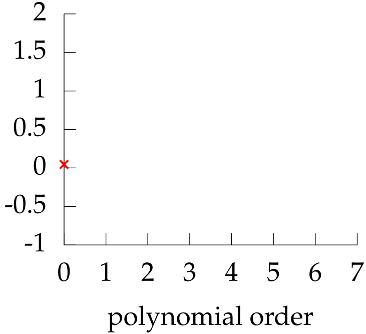




Polynomial order 0, training error -3.346, leave one out error 0.045811.

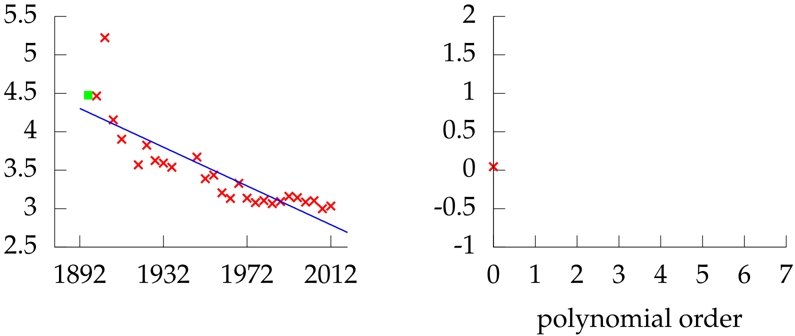
Leave One Out Error





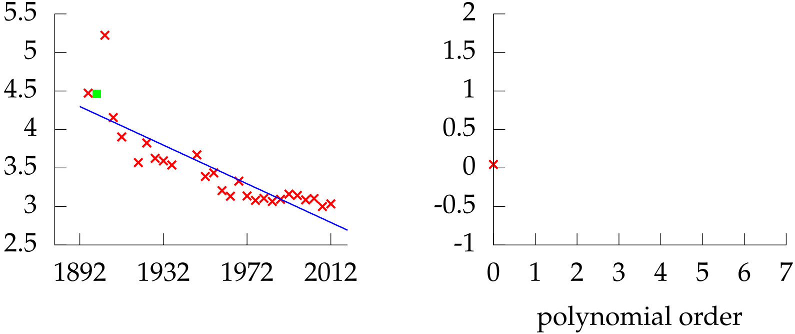
Polynomial order 0, training error -3.346, leave one out error 0.045811.

Leave One Out Error



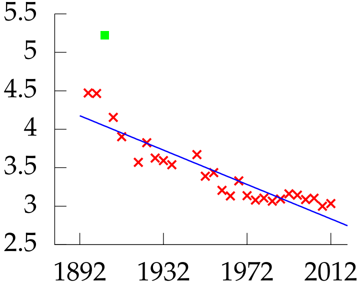
Polynomial order 1, training error -21.183, leave one out error -0.15413.

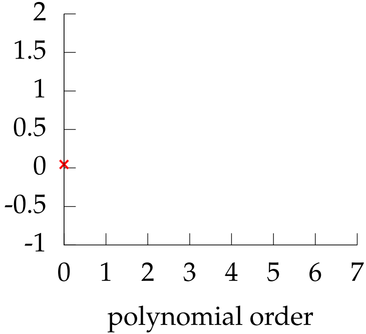
Leave One Out Error



Polynomial order 1, training error -21.183, leave one out error -0.15413.

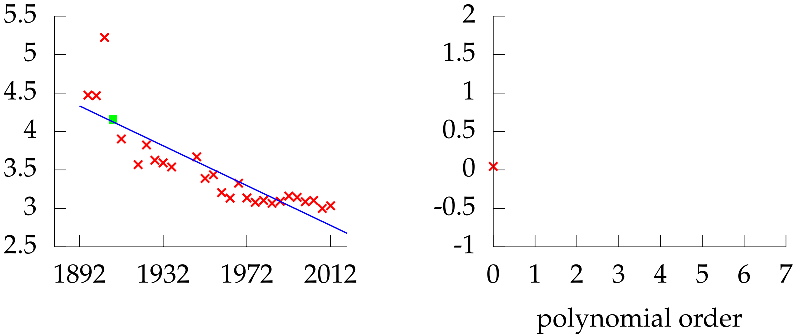
Leave One Out Error





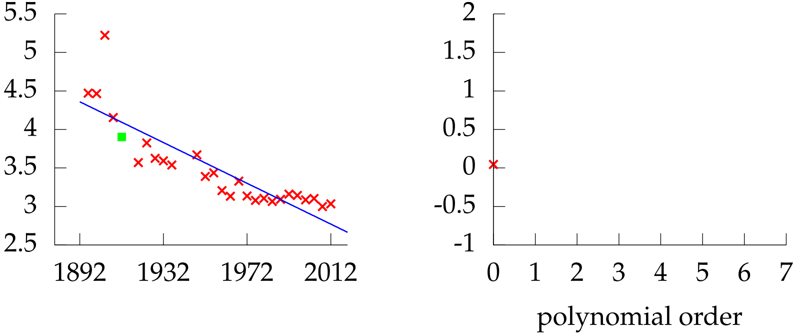
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



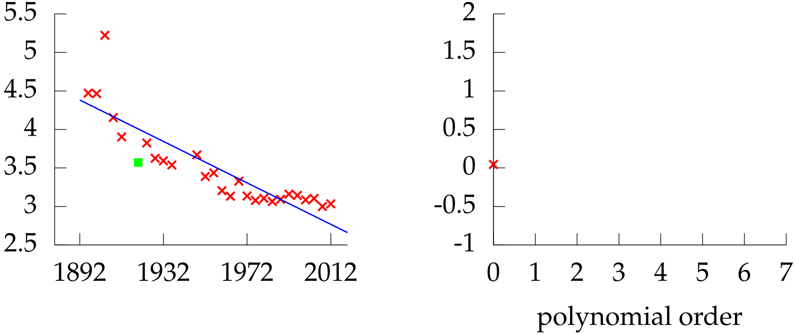
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



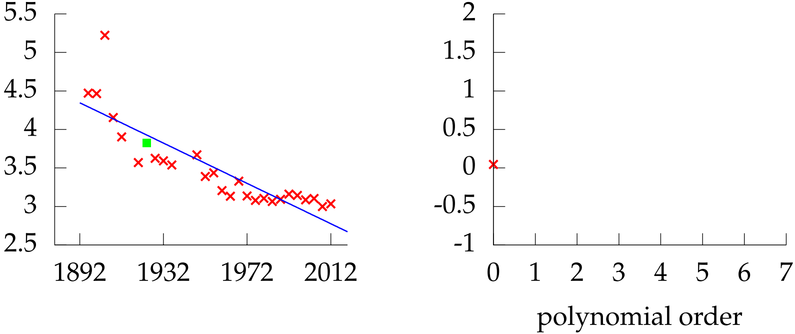
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



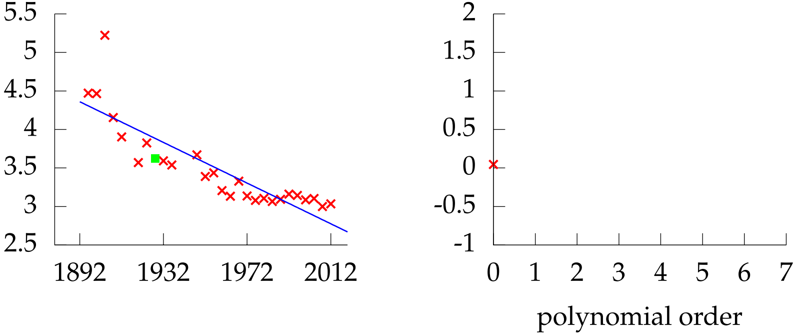
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



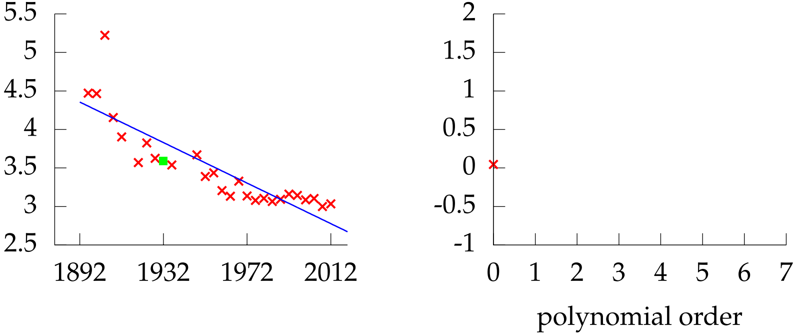
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



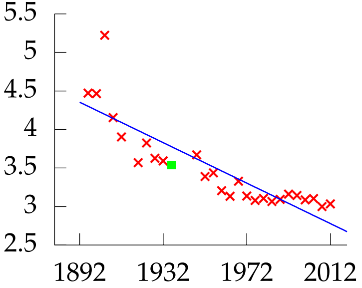
Polynomial order 1, training error -21.183, leave one out error -0.15413.

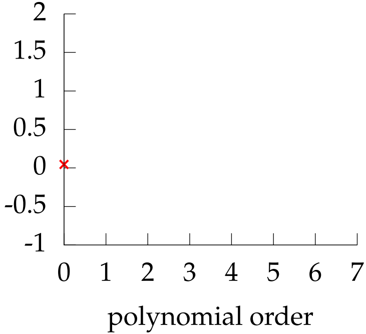
Leave One Out Error



Polynomial order 1, training error -21.183, leave one out error -0.15413.

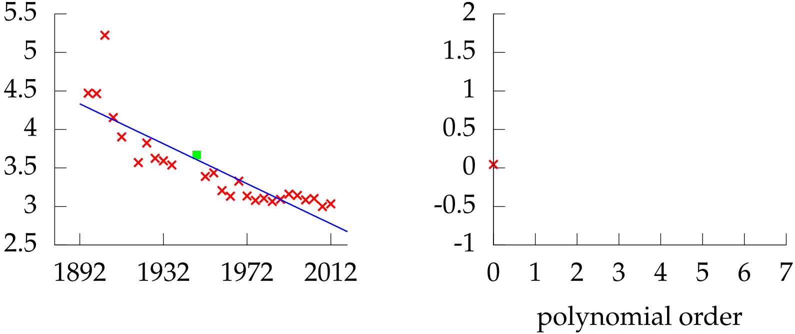
Leave One Out Error





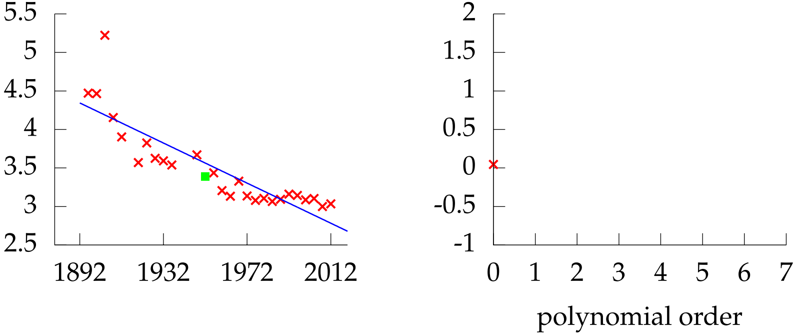
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



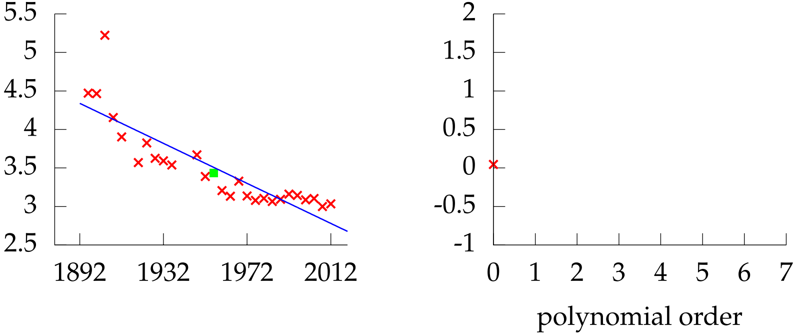
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



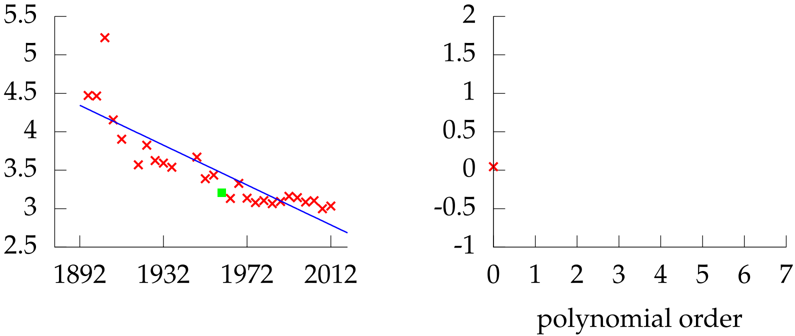
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



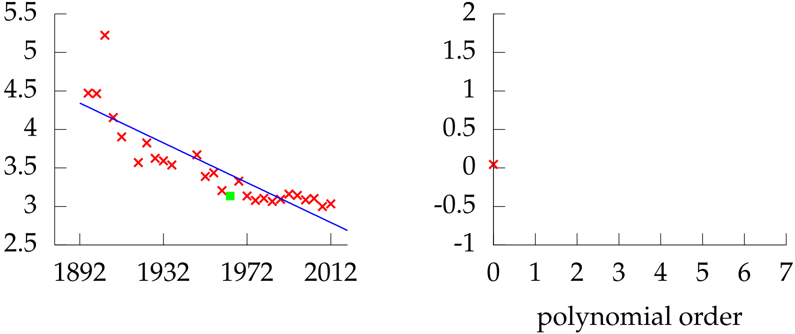
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



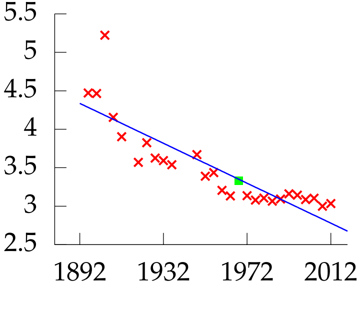
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



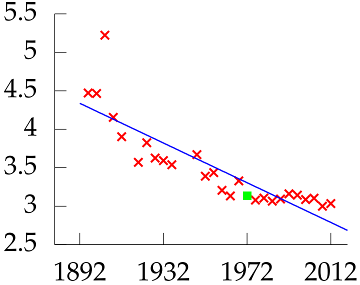
Polynomial order 1, training error -21.183, leave one out error -0.15413.

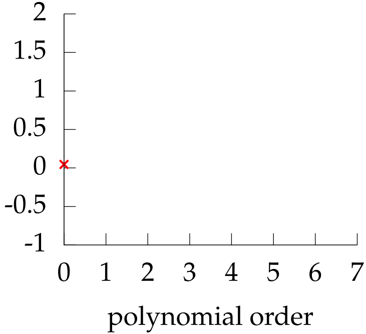
Leave One Out Error



Polynomial order 1, training error -21.183, leave one out error -0.15413.

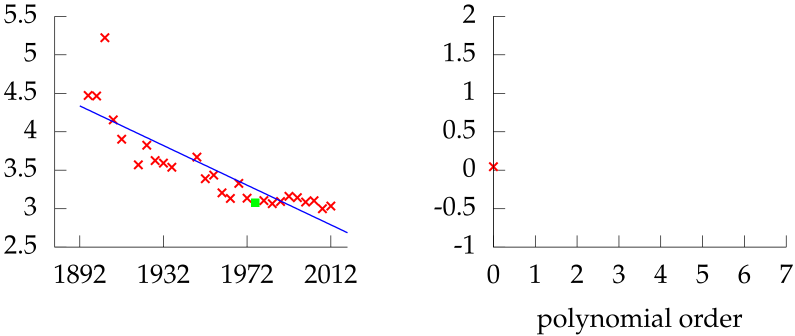
Leave One Out Error





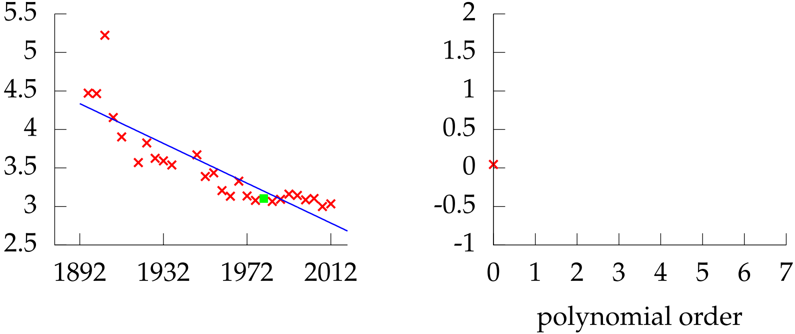
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



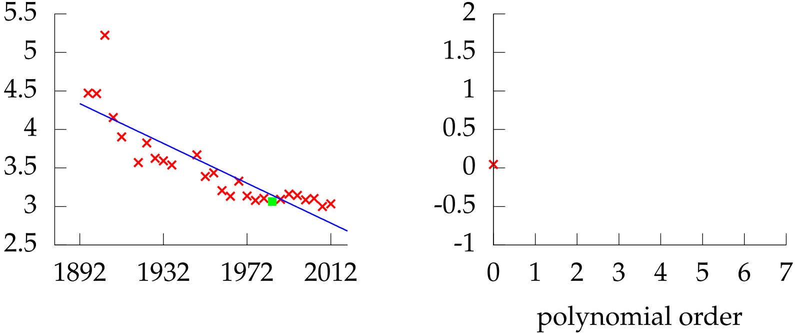
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



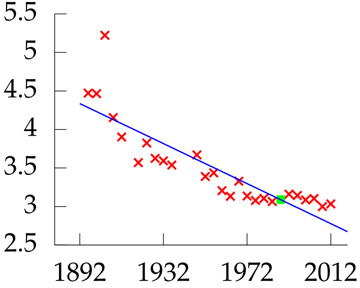
Polynomial order 1, training error -21.183, leave one out error -0.15413.

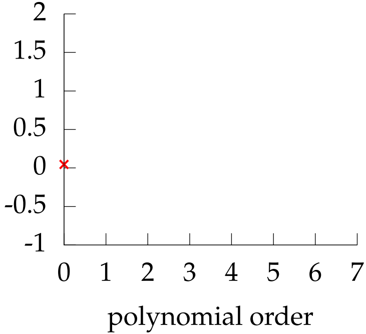
Leave One Out Error



Polynomial order 1, training error -21.183, leave one out error -0.15413.

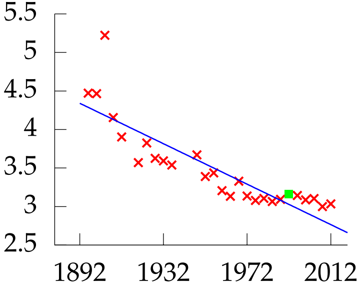
Leave One Out Error

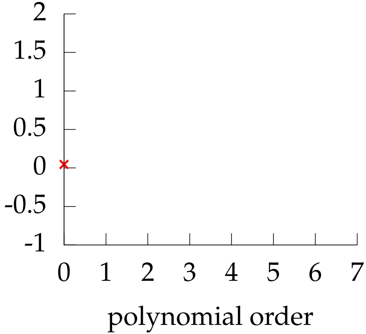




Polynomial order 1, training error -21.183, leave one out error -0.15413.

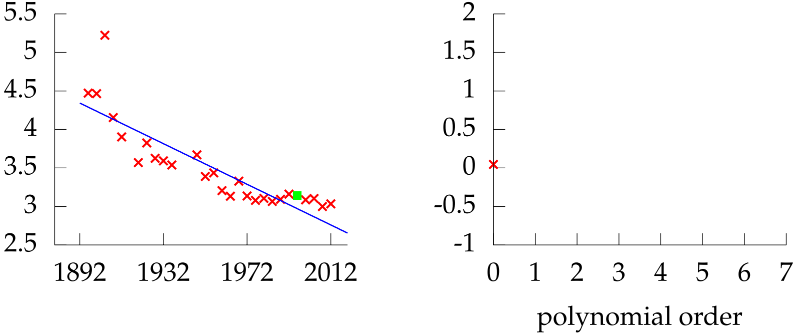
Leave One Out Error





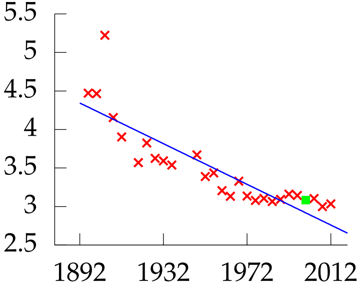
Polynomial order 1, training error -21.183, leave one out error -0.15413.

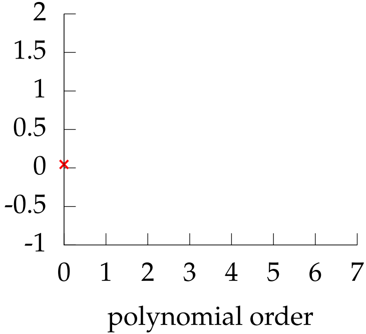
Leave One Out Error



Polynomial order 1, training error -21.183, leave one out error -0.15413.

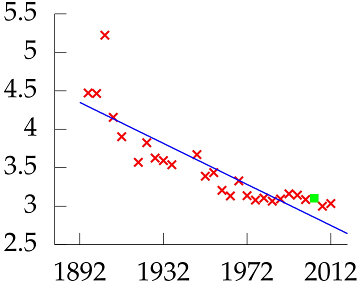
Leave One Out Error

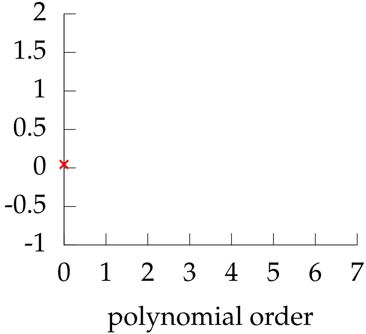




Polynomial order 1, training error -21.183, leave one out error -0.15413.

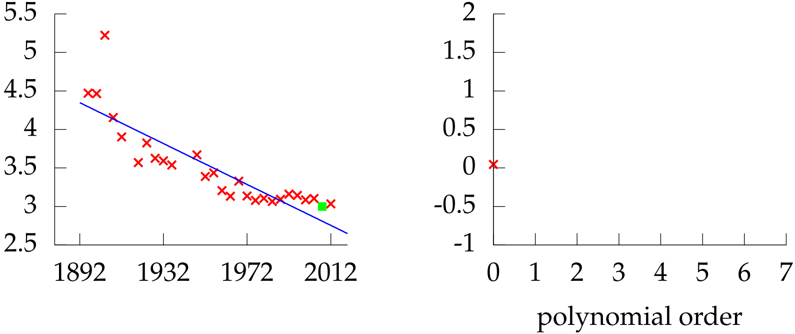
Leave One Out Error





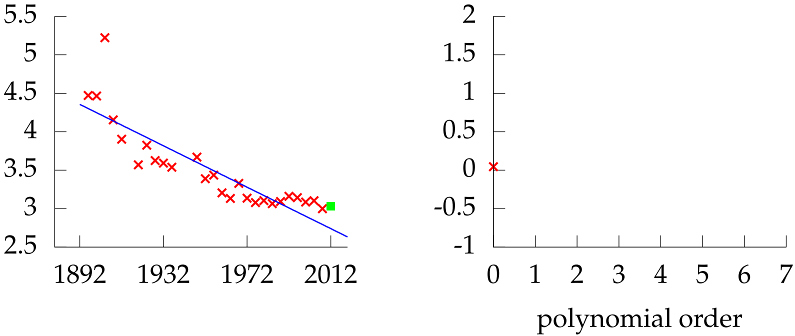
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



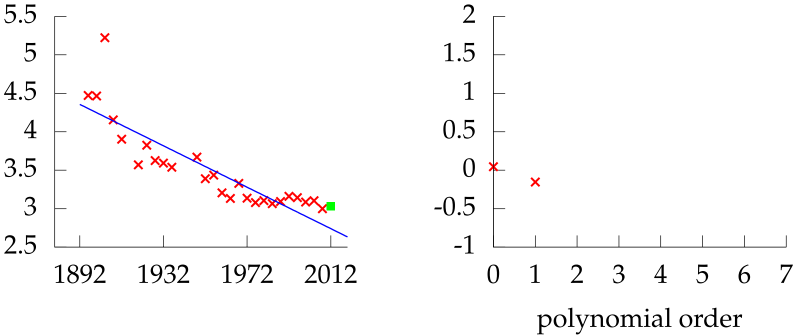
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



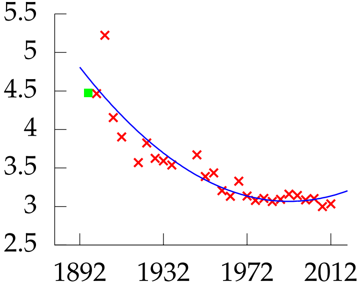
Polynomial order 1, training error -21.183, leave one out error -0.15413.

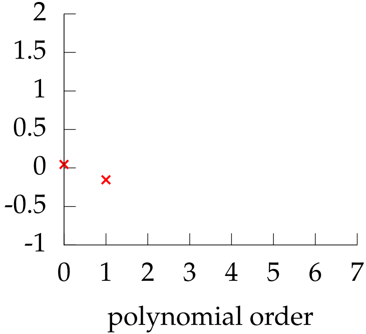
Leave One Out Error



Polynomial order 1, training error -21.183, leave one out error -0.15413.

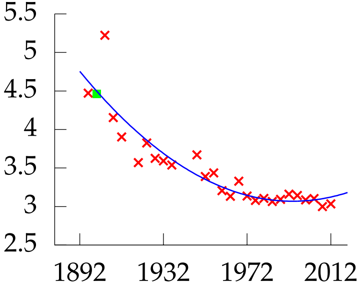
Leave One Out Error

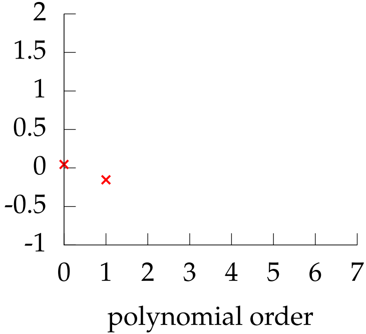




Polynomial order 2, training error -28.403, leave one out error 0.34669.

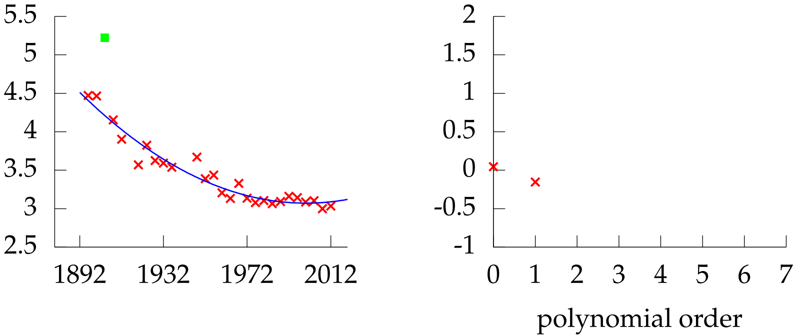
Leave One Out Error





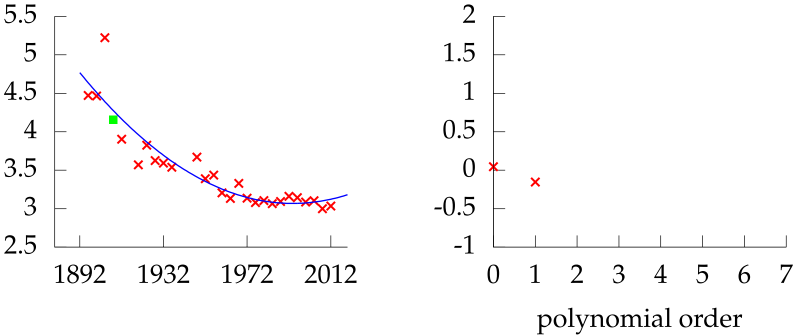
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



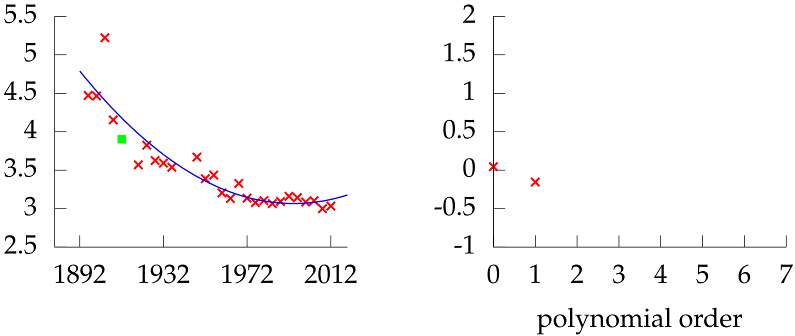
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



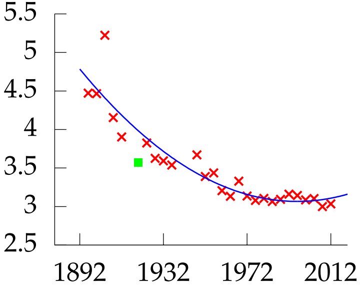
Polynomial order 2, training error -28.403, leave one out error 0.34669.

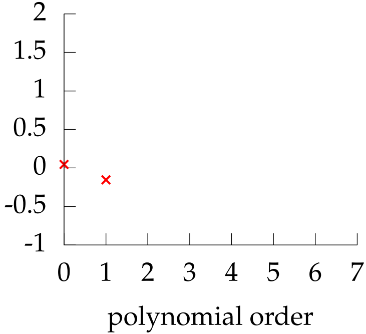
Leave One Out Error



Polynomial order 2, training error -28.403, leave one out error 0.34669.

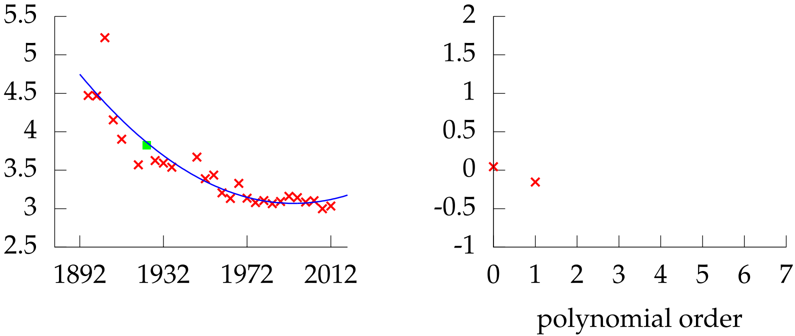
Leave One Out Error





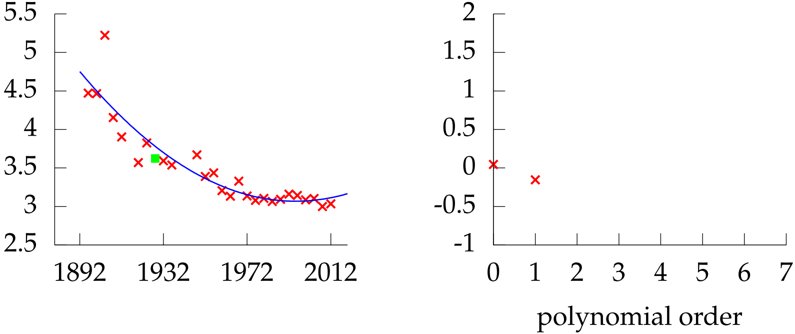
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



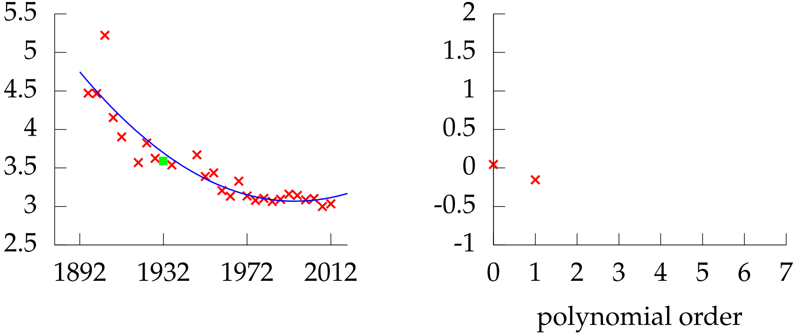
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



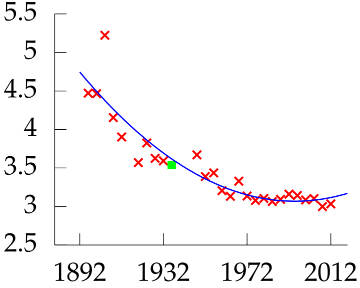
Polynomial order 2, training error -28.403, leave one out error 0.34669.

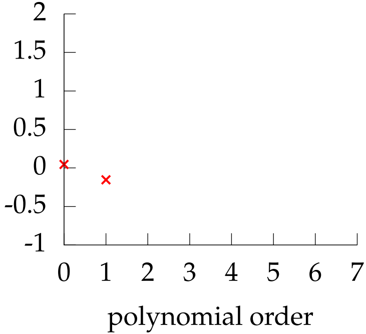
Leave One Out Error



Polynomial order 2, training error -28.403, leave one out error 0.34669.

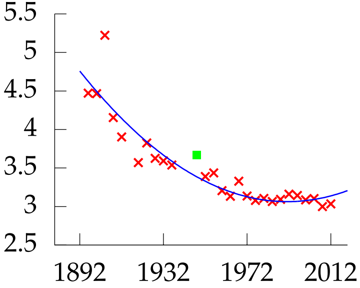
Leave One Out Error

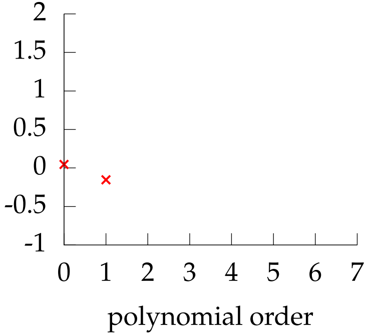




Polynomial order 2, training error -28.403, leave one out error 0.34669.

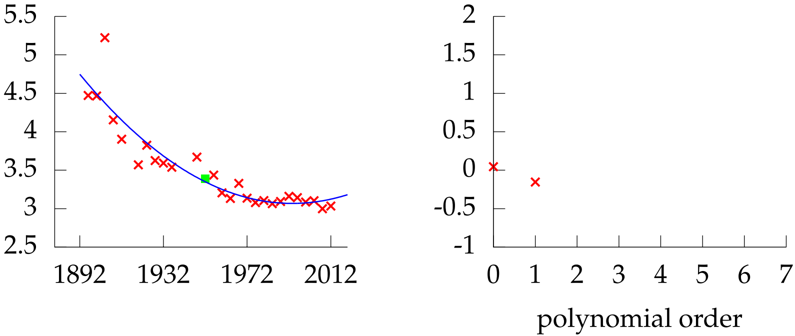
Leave One Out Error





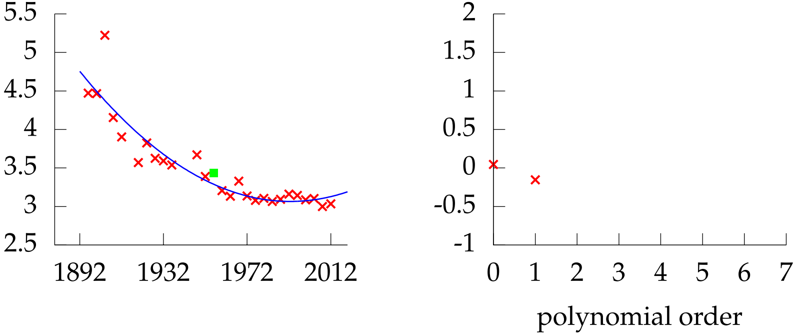
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



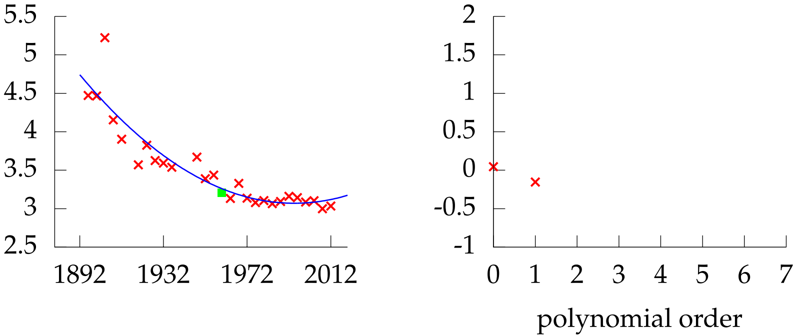
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



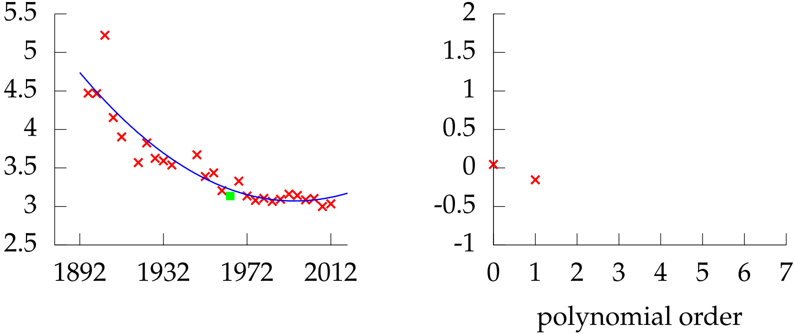
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



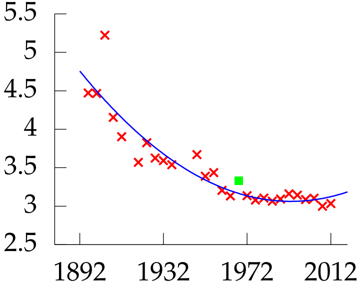
Polynomial order 2, training error -28.403, leave one out error 0.34669.

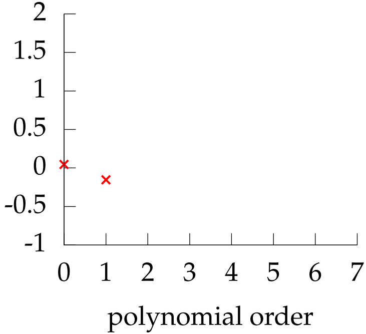
Leave One Out Error



Polynomial order 2, training error -28.403, leave one out error 0.34669.

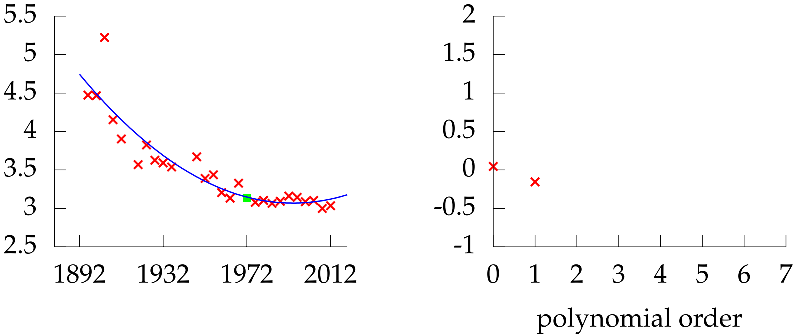
Leave One Out Error





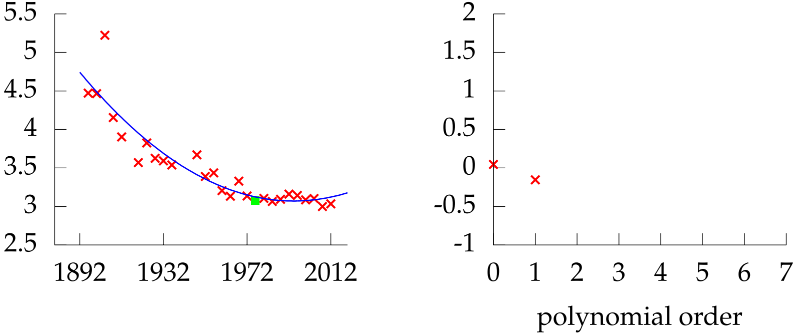
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



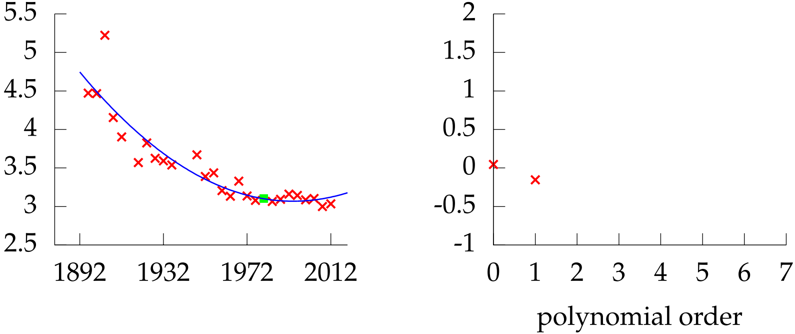
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



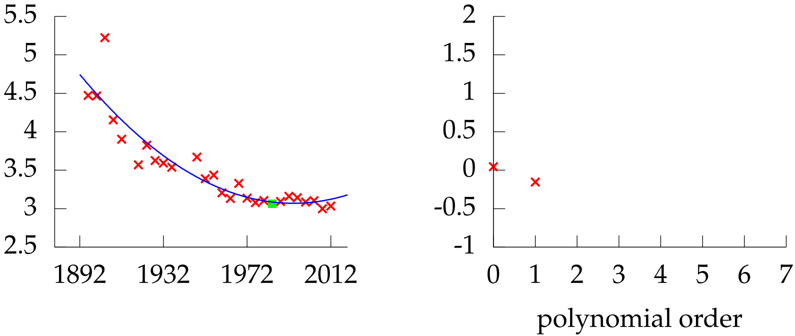
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



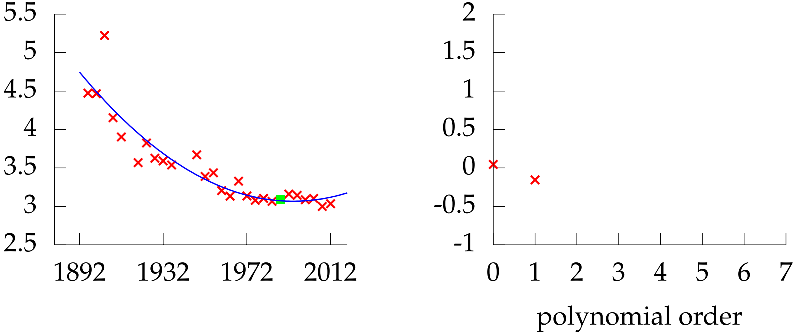
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



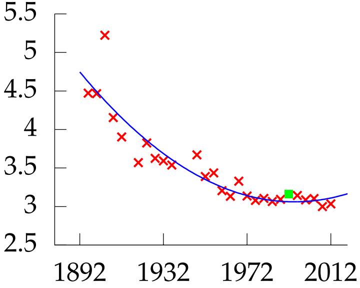
Polynomial order 2, training error -28.403, leave one out error 0.34669.

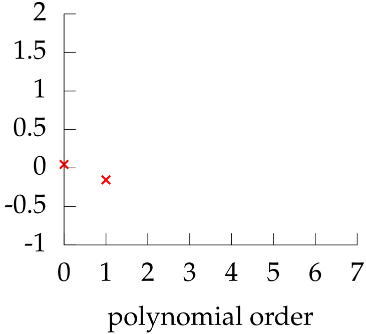
Leave One Out Error



Polynomial order 2, training error -28.403, leave one out error 0.34669.

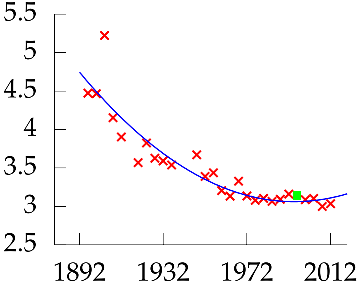
Leave One Out Error

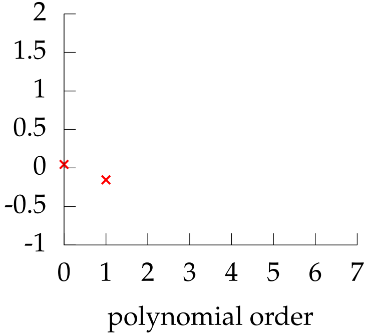




Polynomial order 2, training error -28.403, leave one out error 0.34669.

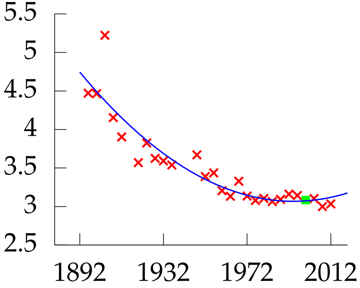
Leave One Out Error

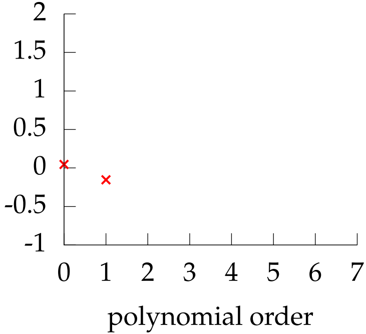




Polynomial order 2, training error -28.403, leave one out error 0.34669.

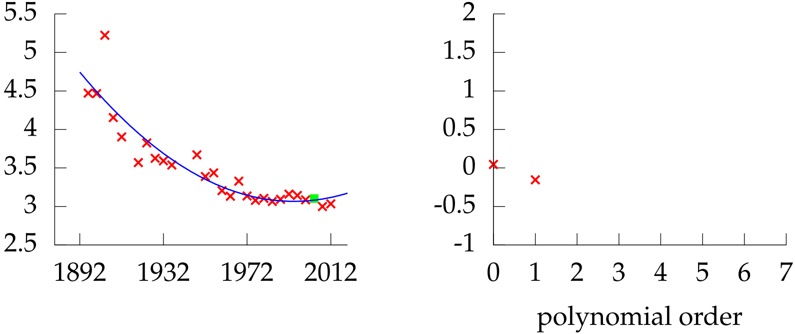
Leave One Out Error





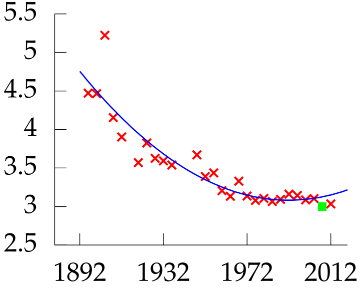
Polynomial order 2, training error -28.403, leave one out error 0.34669.

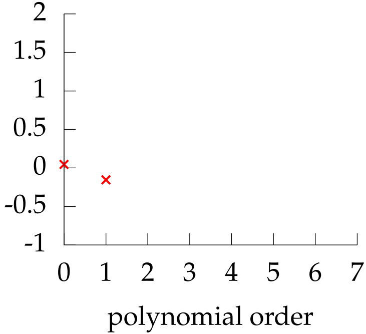
Leave One Out Error



Polynomial order 2, training error -28.403, leave one out error 0.34669.

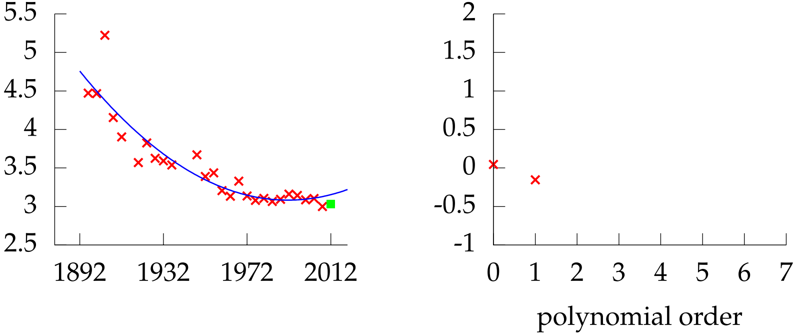
Leave One Out Error





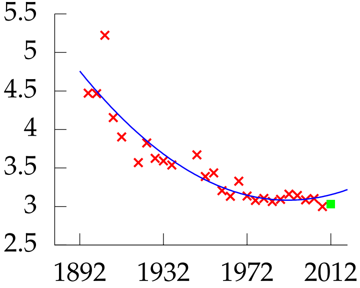
Polynomial order 2, training error -28.403, leave one out error 0.34669.

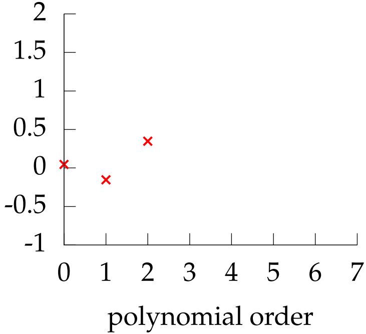
Leave One Out Error



Polynomial order 2, training error -28.403, leave one out error 0.34669.

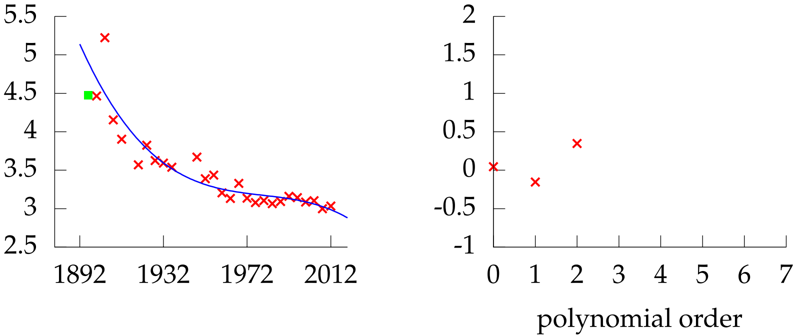
Leave One Out Error





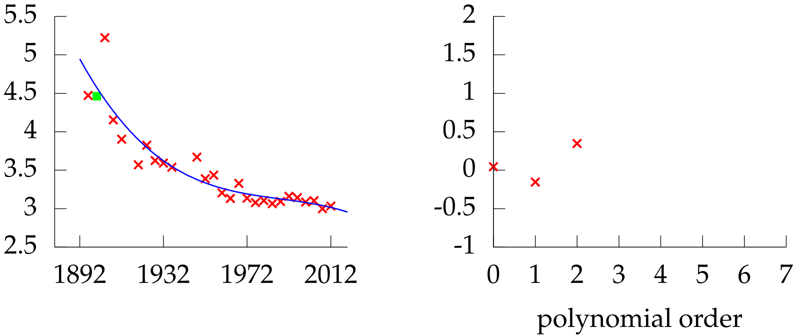
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



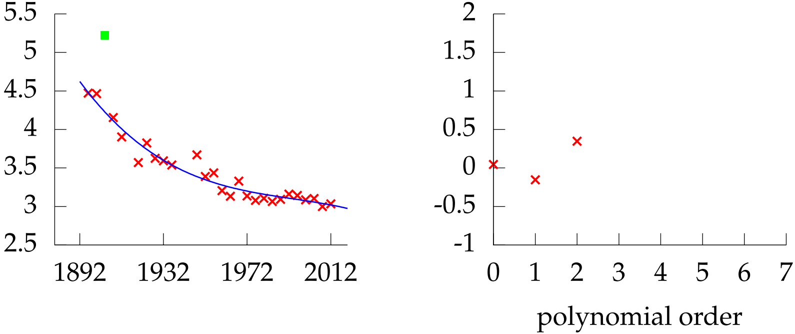
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



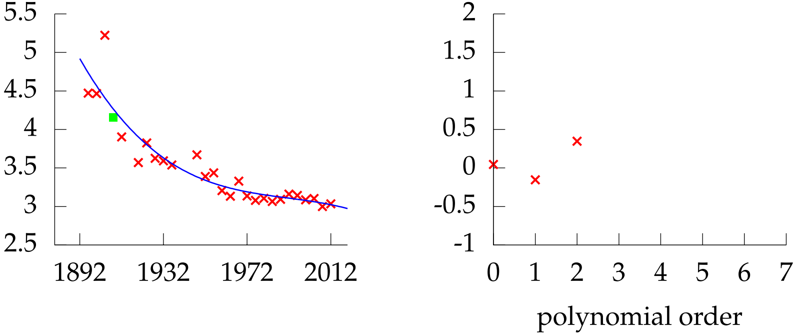
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



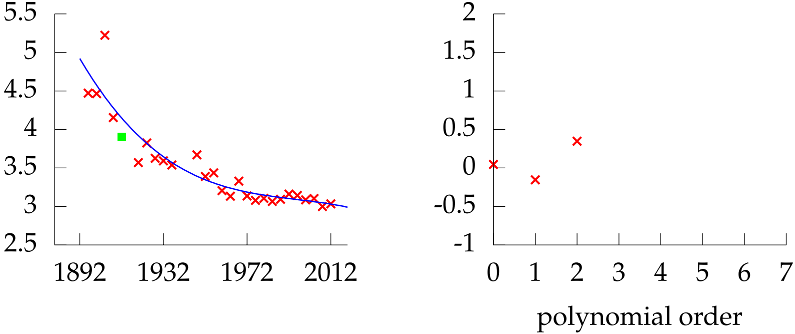
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



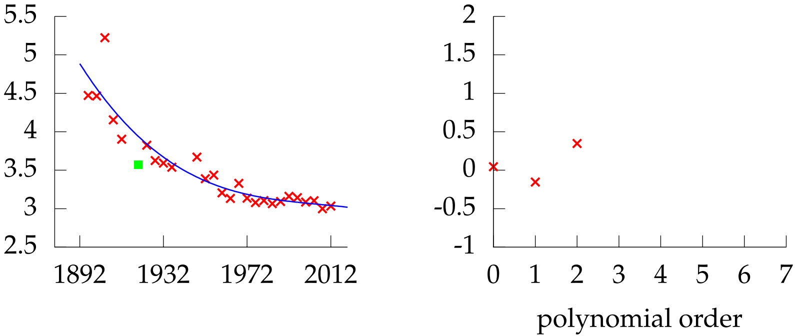
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



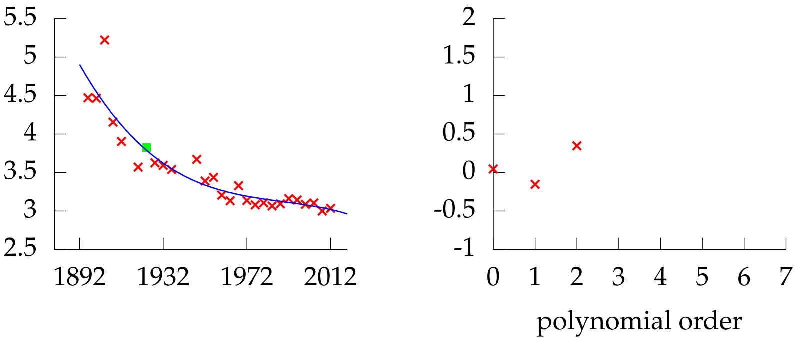
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



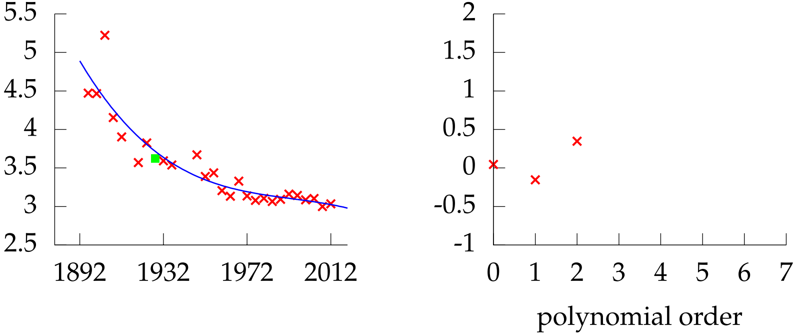
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



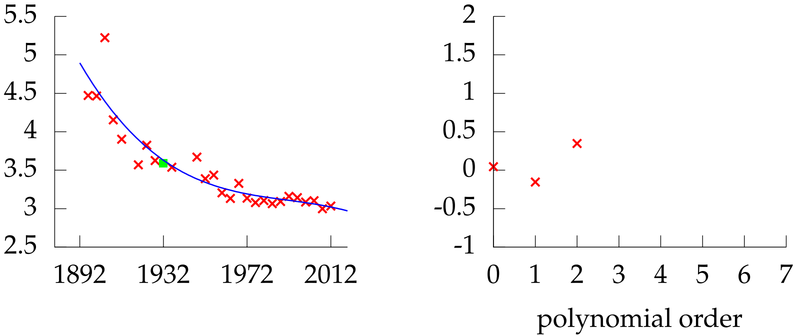
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



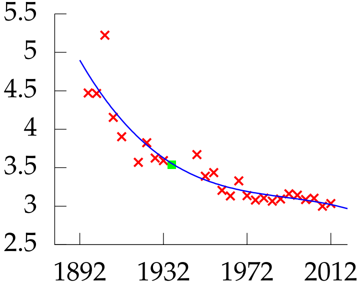
Polynomial order 3, training error -29.223, leave one out error 0.51621.

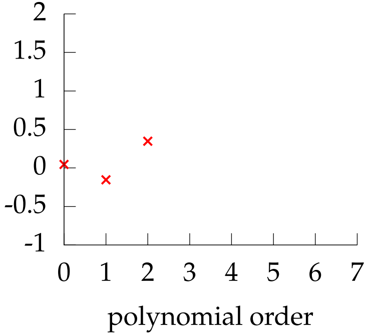
Leave One Out Error



Polynomial order 3, training error -29.223, leave one out error 0.51621.

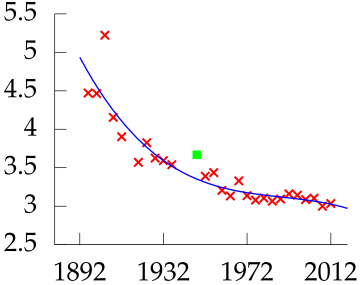
Leave One Out Error

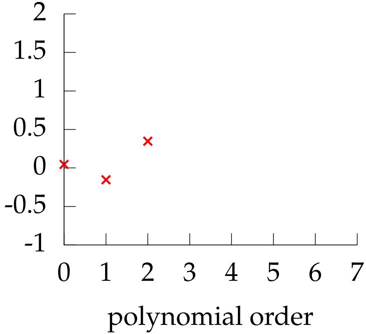




Polynomial order 3, training error -29.223, leave one out error 0.51621.

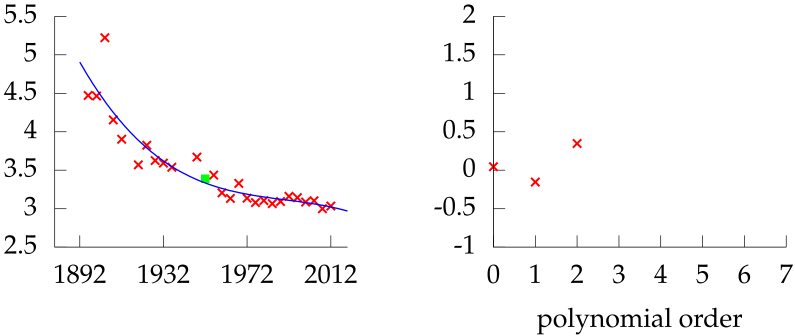
Leave One Out Error





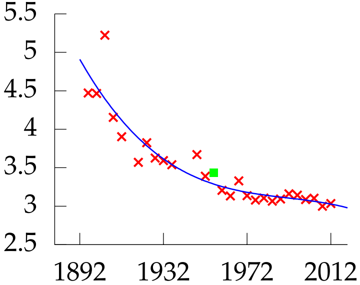
Polynomial order 3, training error -29.223, leave one out error 0.51621.

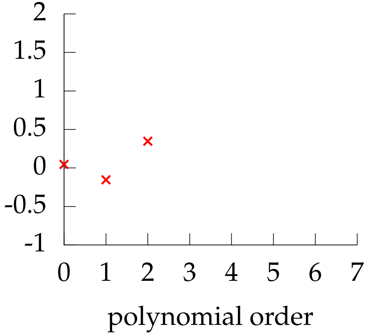
Leave One Out Error



Polynomial order 3, training error -29.223, leave one out error 0.51621.

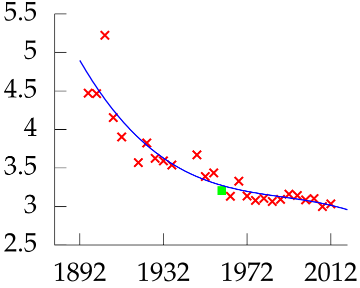
Leave One Out Error

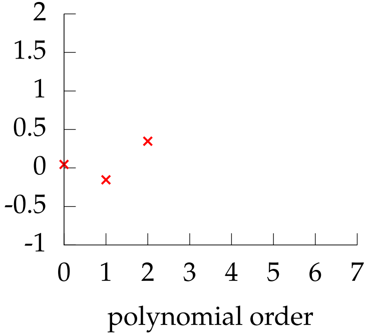




Polynomial order 3, training error -29.223, leave one out error 0.51621.

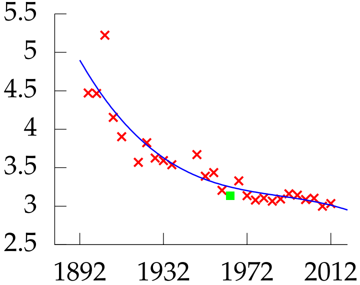
Leave One Out Error

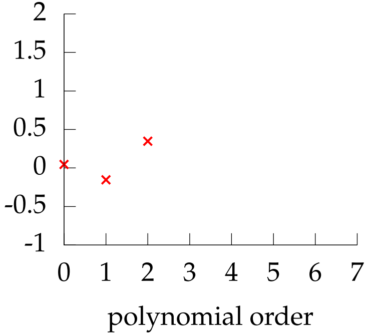




Polynomial order 3, training error -29.223, leave one out error 0.51621.

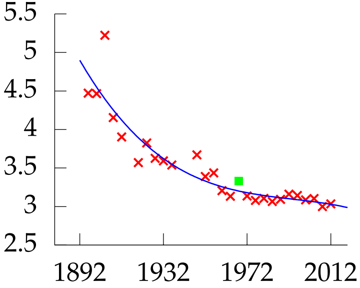
Leave One Out Error

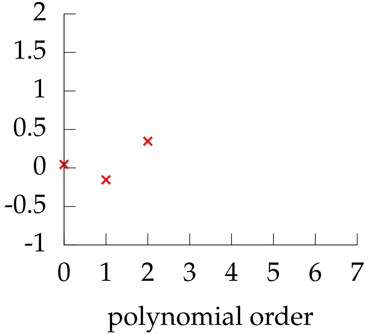




Polynomial order 3, training error -29.223, leave one out error 0.51621.

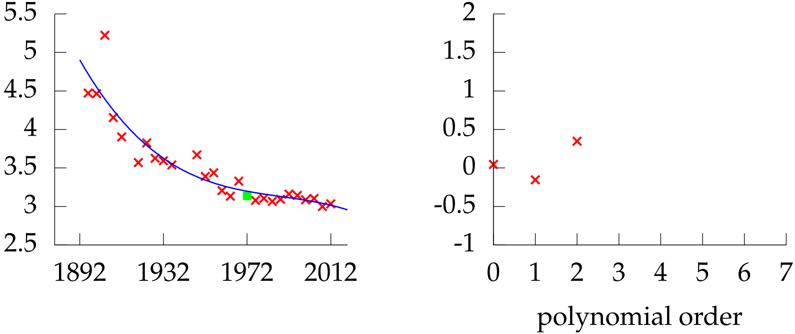
Leave One Out Error





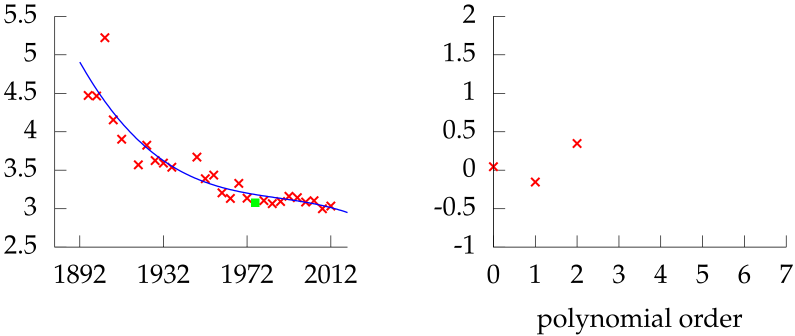
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



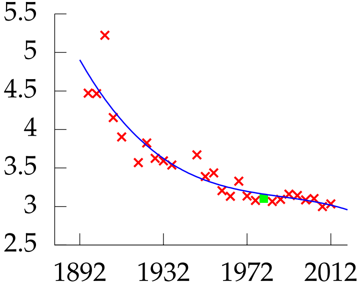
Polynomial order 3, training error -29.223, leave one out error 0.51621.

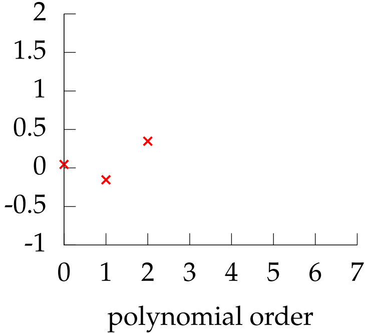
Leave One Out Error



Polynomial order 3, training error -29.223, leave one out error 0.51621.

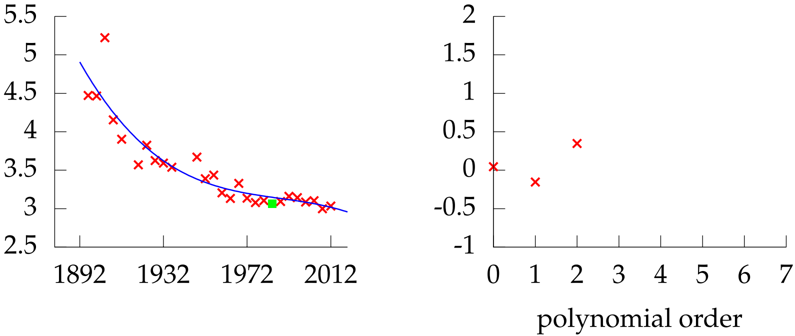
Leave One Out Error





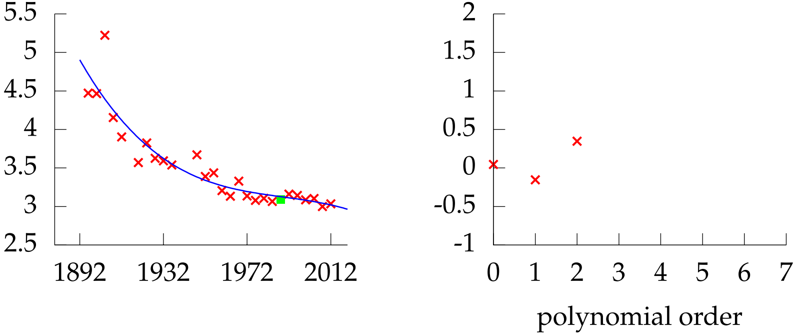
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



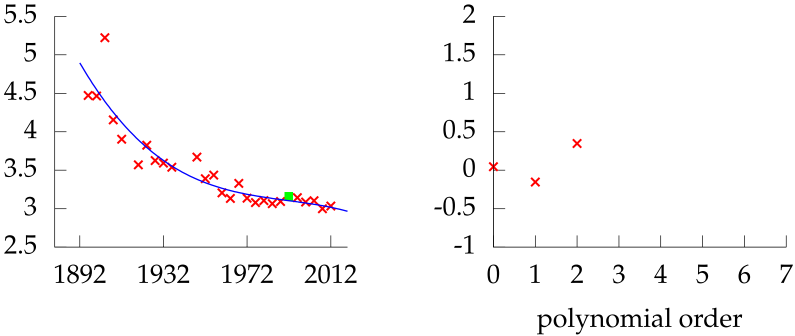
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



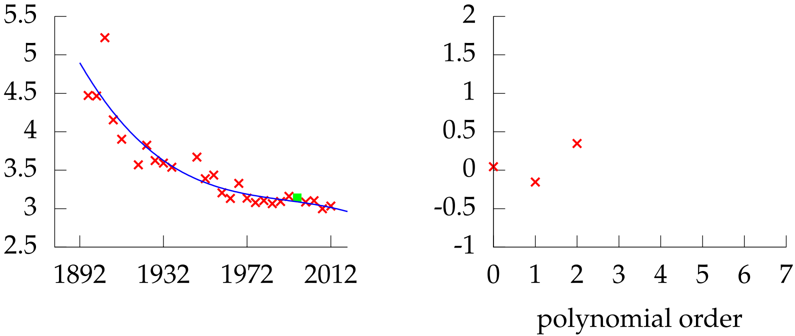
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



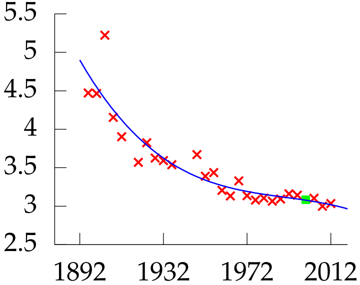
Polynomial order 3, training error -29.223, leave one out error 0.51621.

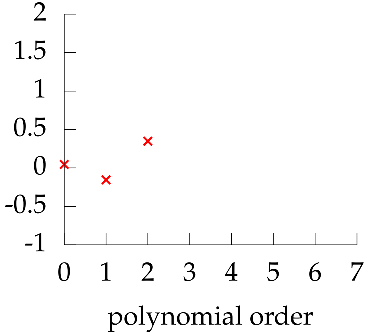
Leave One Out Error



Polynomial order 3, training error -29.223, leave one out error 0.51621.

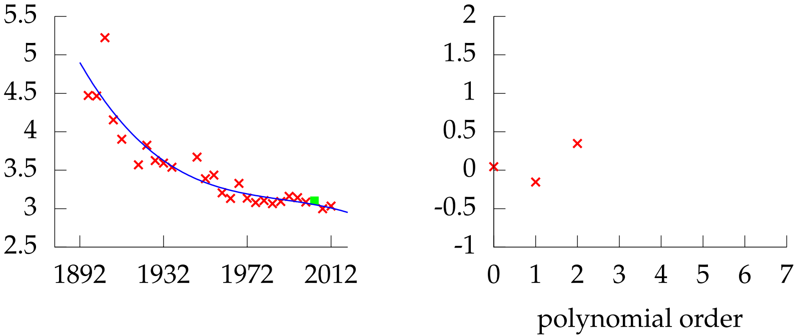
Leave One Out Error





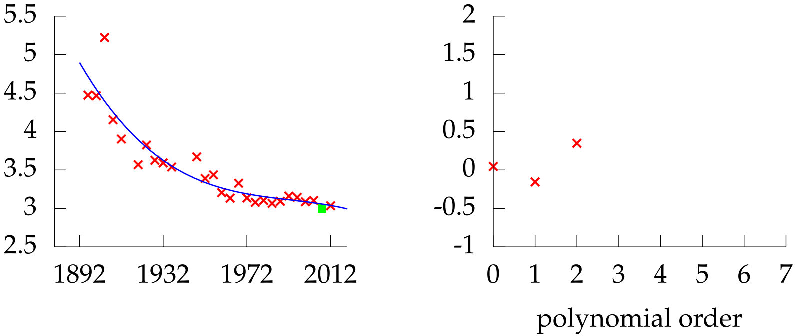
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



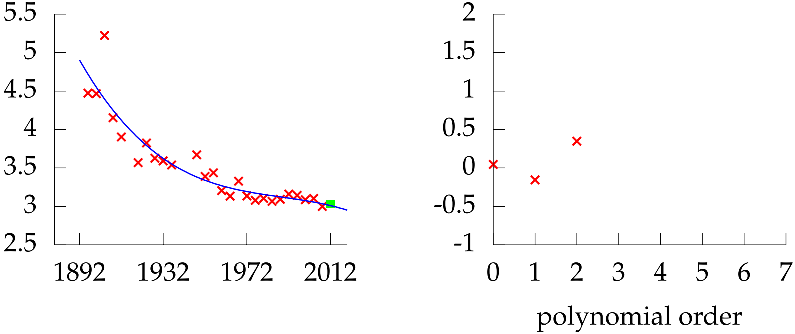
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



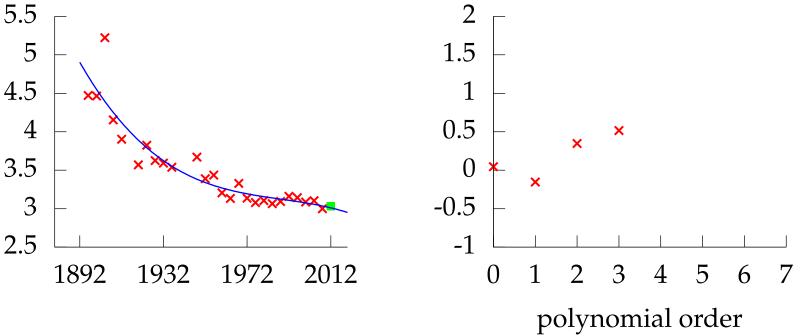
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



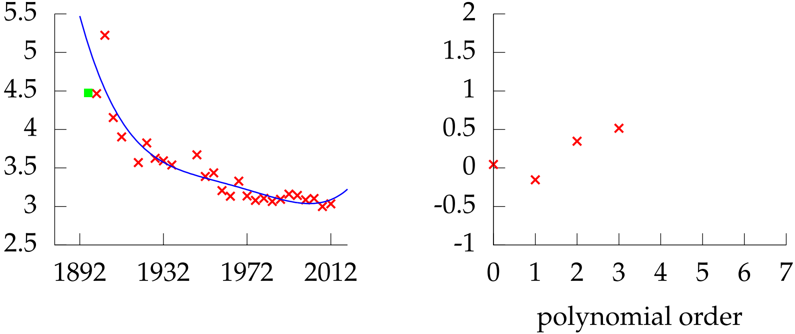
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



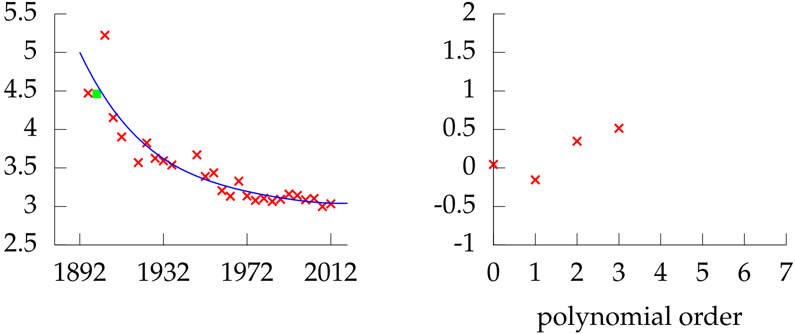
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



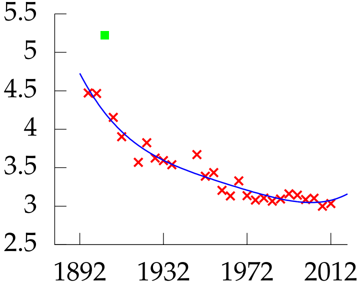
Polynomial order 4, training error -29.324, leave one out error 0.84844.

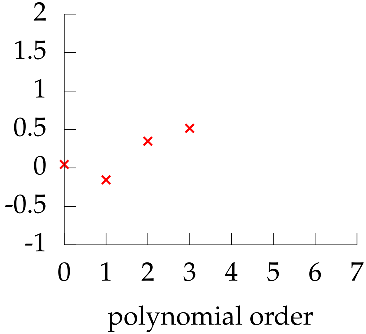
Leave One Out Error



Polynomial order 4, training error -29.324, leave one out error 0.84844.

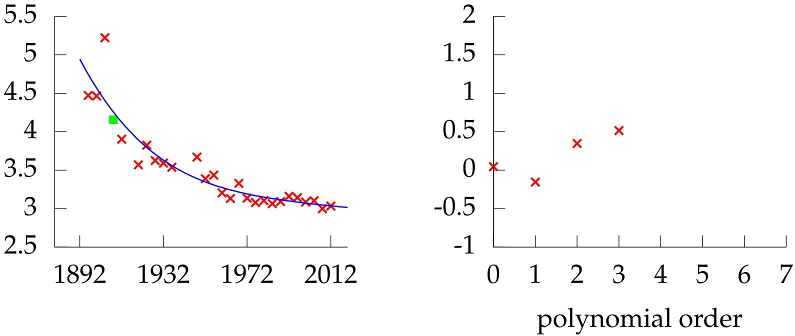
Leave One Out Error





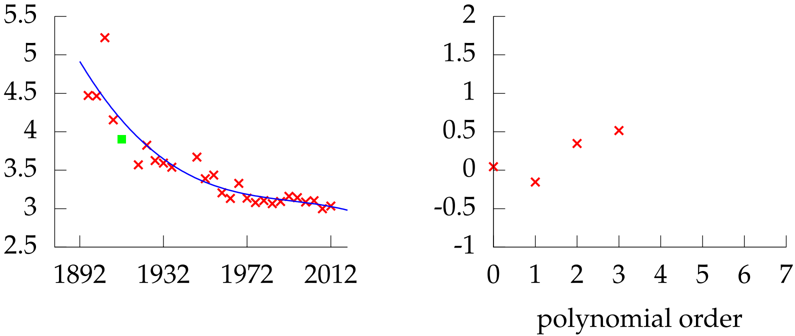
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



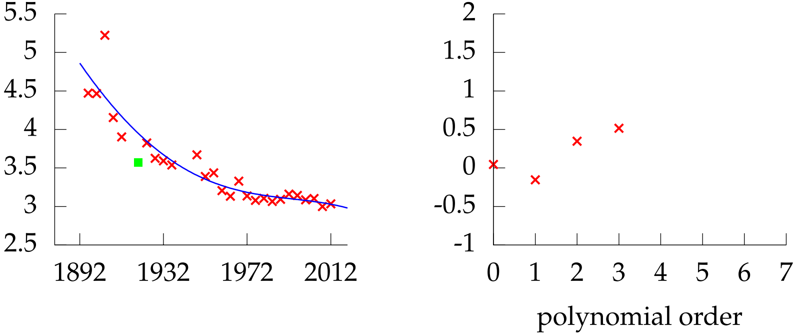
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



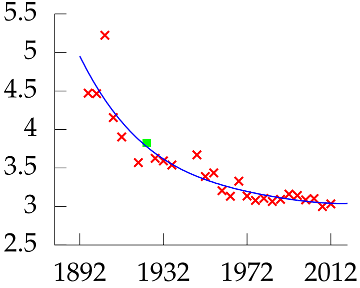
Polynomial order 4, training error -29.324, leave one out error 0.84844.

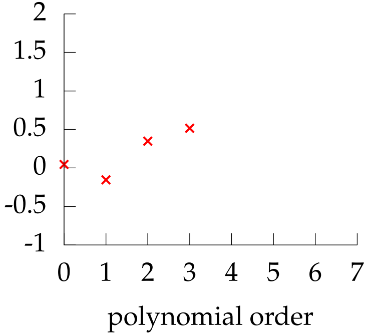
Leave One Out Error



Polynomial order 4, training error -29.324, leave one out error 0.84844.

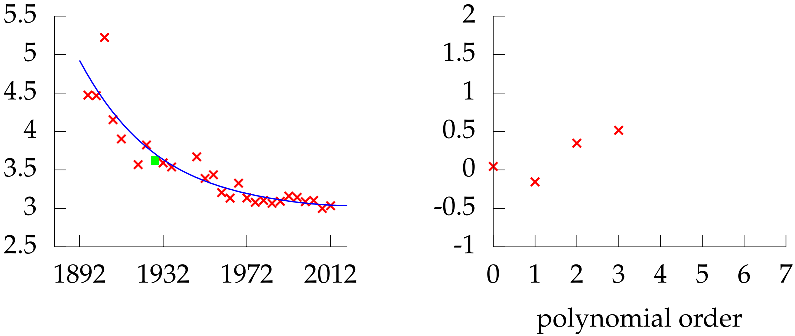
Leave One Out Error





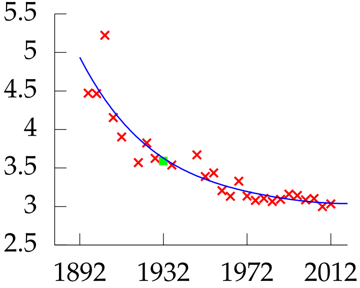
Polynomial order 4, training error -29.324, leave one out error 0.84844.

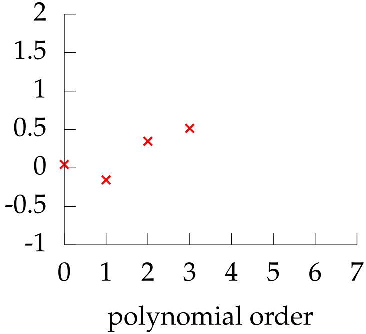
Leave One Out Error



Polynomial order 4, training error -29.324, leave one out error 0.84844.

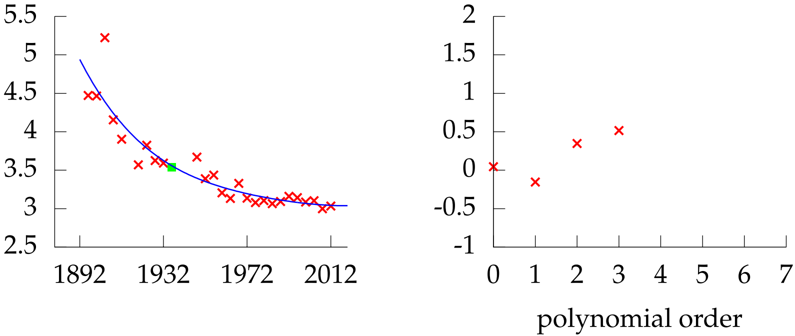
Leave One Out Error





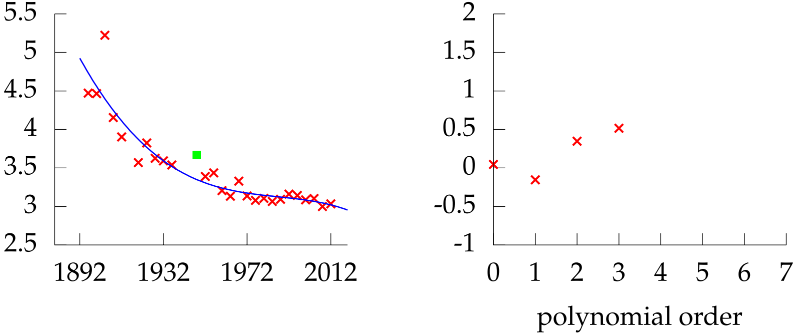
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



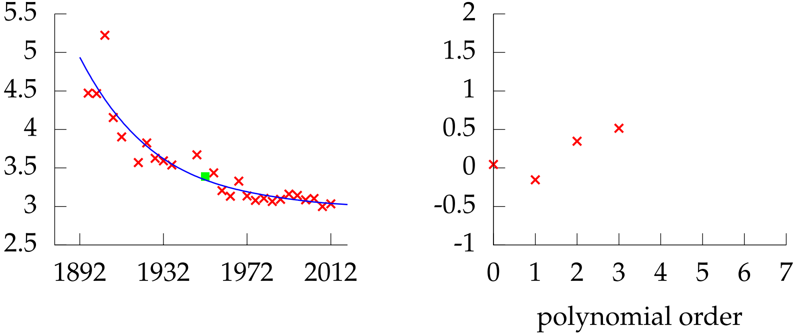
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



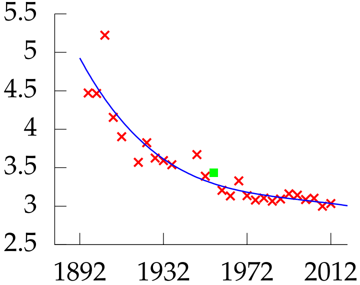
Polynomial order 4, training error -29.324, leave one out error 0.84844.

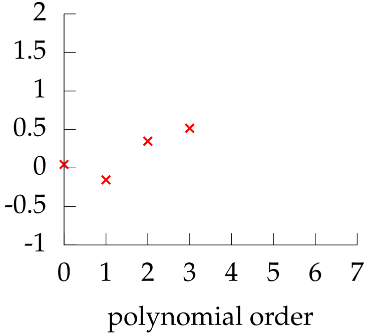
Leave One Out Error



Polynomial order 4, training error -29.324, leave one out error 0.84844.

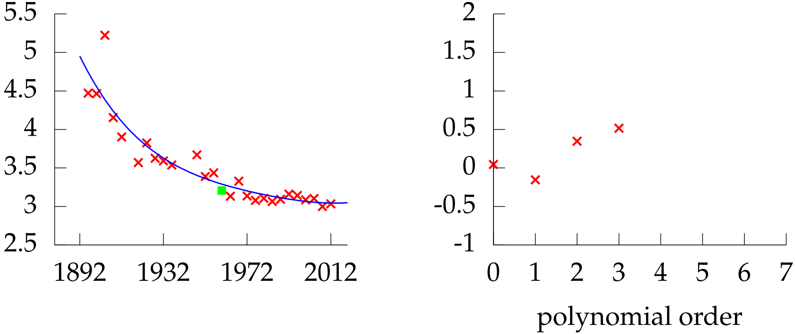
Leave One Out Error





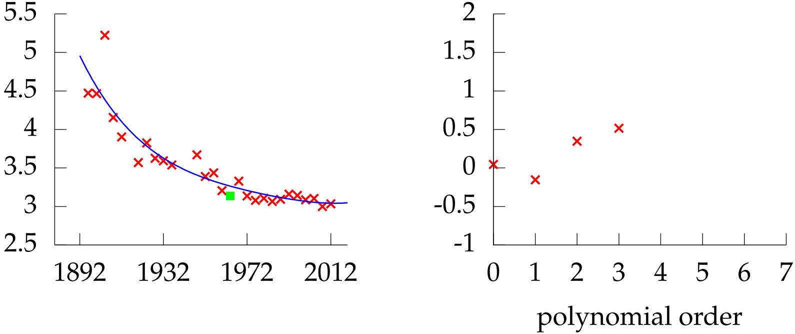
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



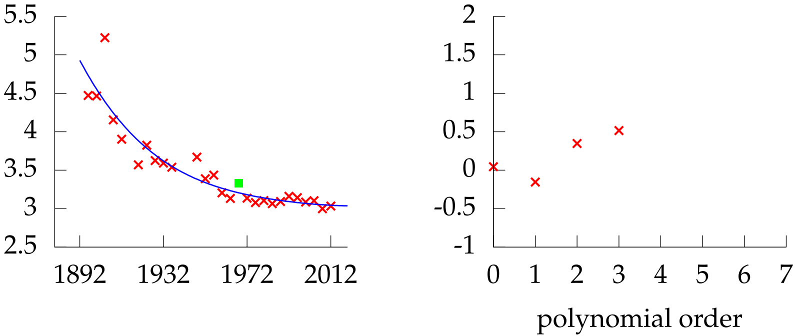
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



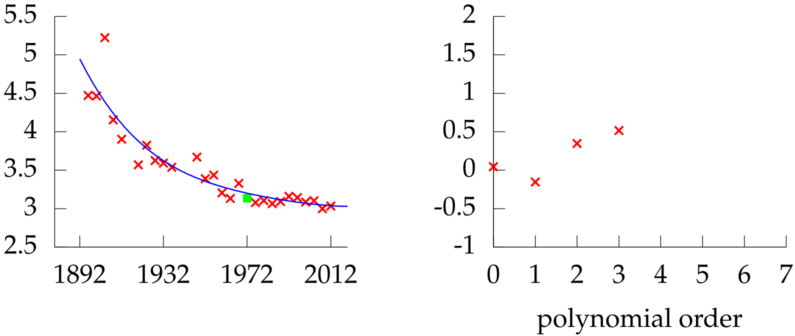
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



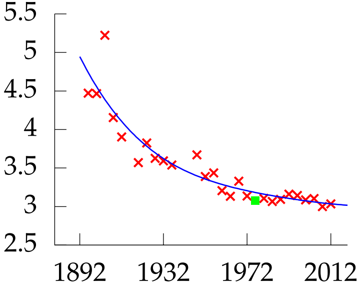
Polynomial order 4, training error -29.324, leave one out error 0.84844.

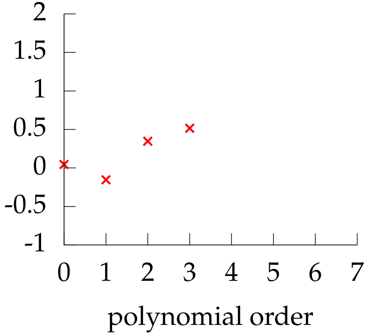
Leave One Out Error



Polynomial order 4, training error -29.324, leave one out error 0.84844.

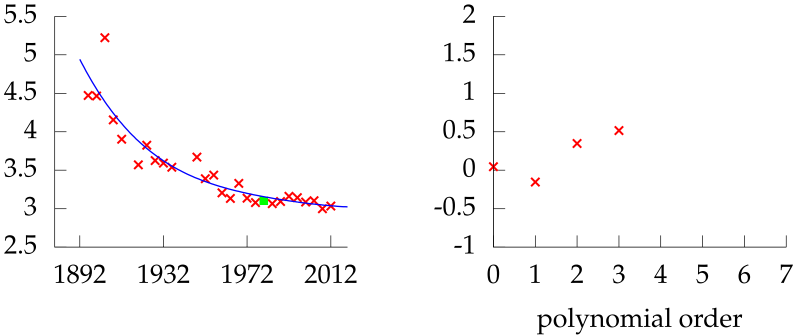
Leave One Out Error





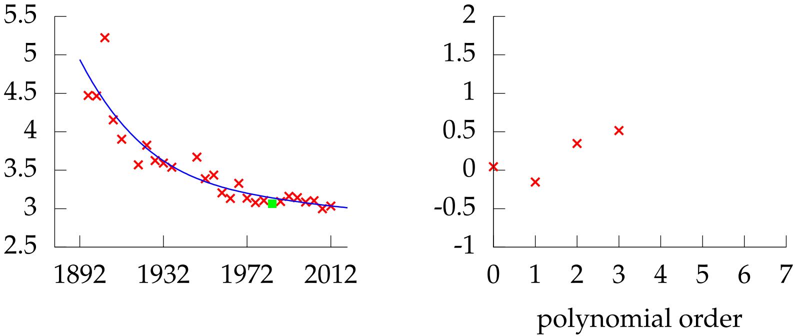
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



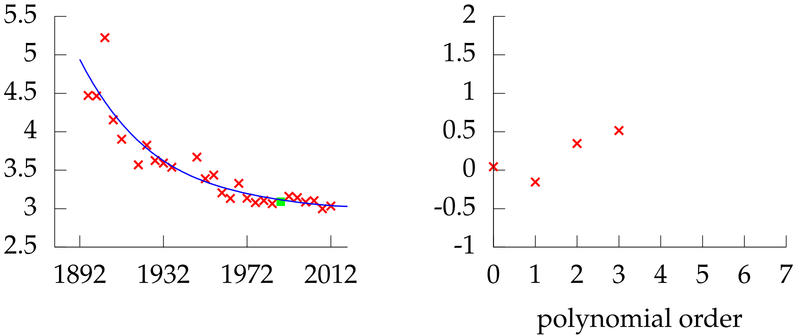
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



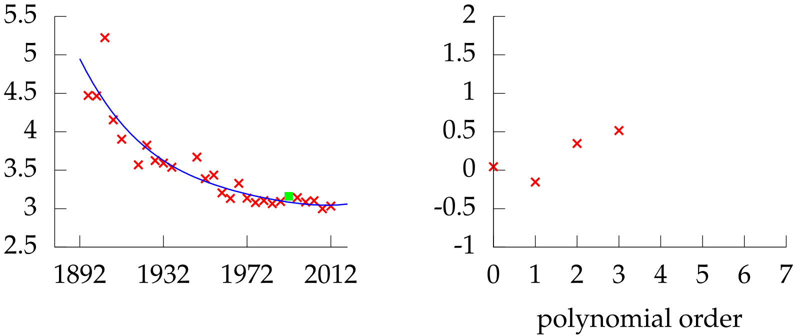
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



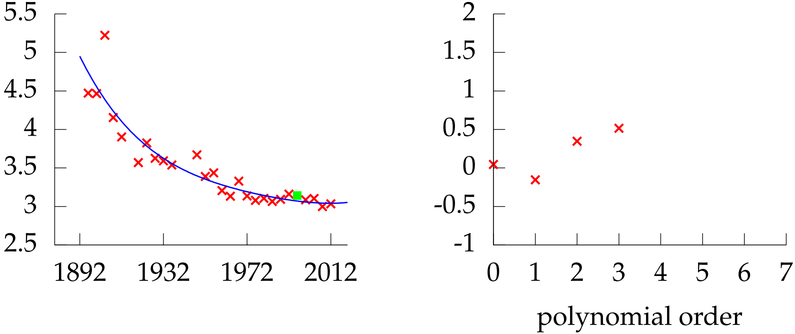
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



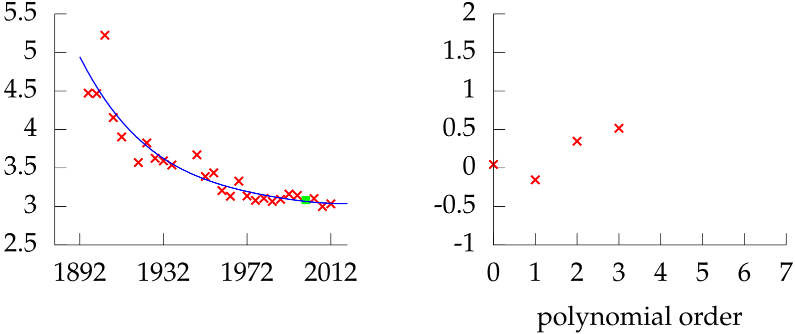
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



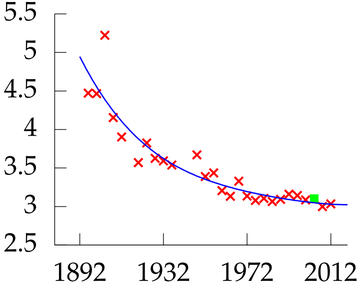
Polynomial order 4, training error -29.324, leave one out error 0.84844.

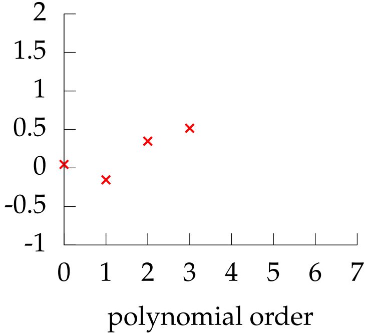
Leave One Out Error



Polynomial order 4, training error -29.324, leave one out error 0.84844.

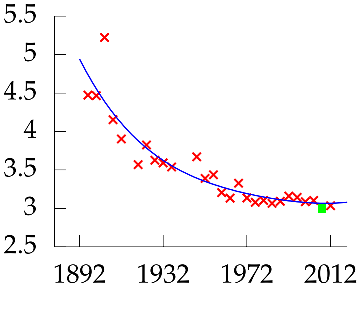
Leave One Out Error





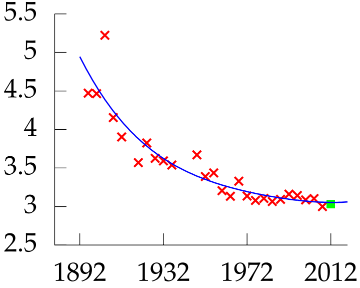
Polynomial order 4, training error -29.324, leave one out error 0.84844.

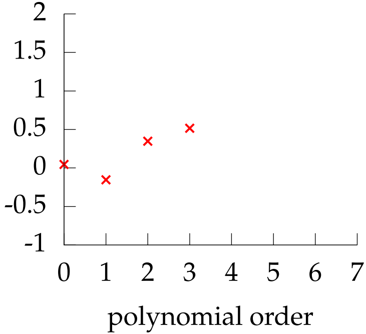
Leave One Out Error



Polynomial order 4, training error -29.324, leave one out error 0.84844.

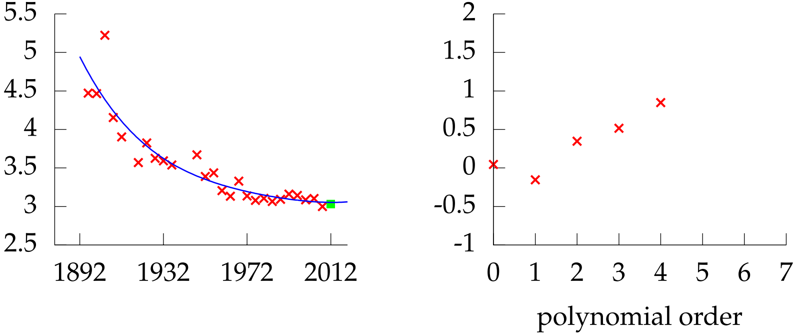
Leave One Out Error





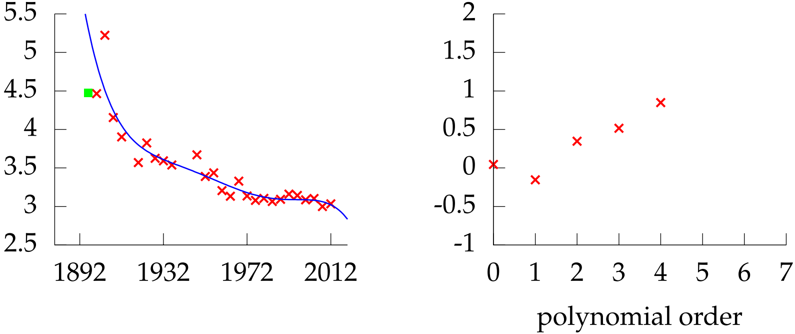
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



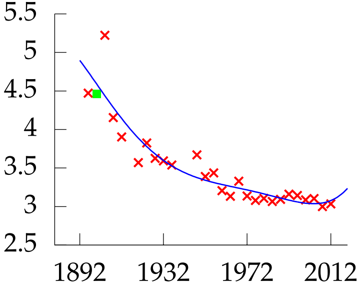
Polynomial order 4, training error -29.324, leave one out error 0.84844.

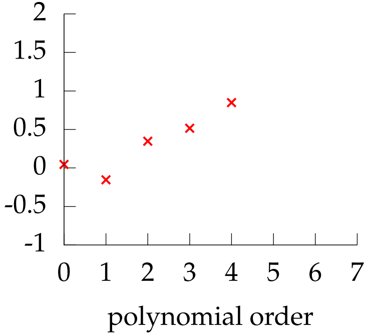
Leave One Out Error



Polynomial order 5, training error -29.524, leave one out error 1.48.

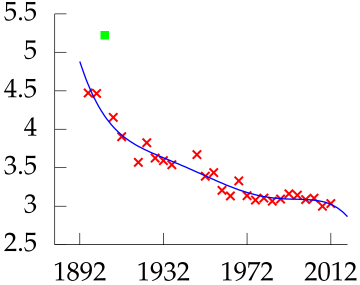
Leave One Out Error

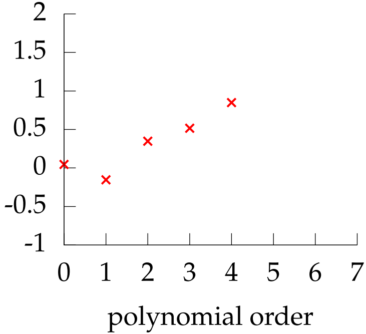




Polynomial order 5, training error -29.524, leave one out error 1.48.

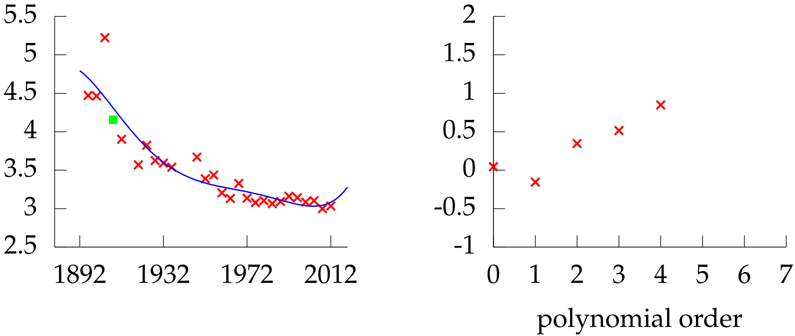
Leave One Out Error





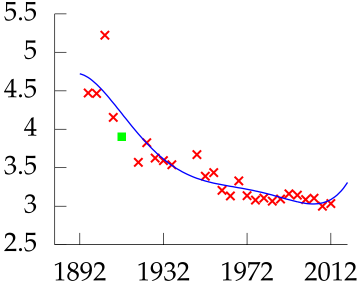
Polynomial order 5, training error -29.524, leave one out error 1.48.

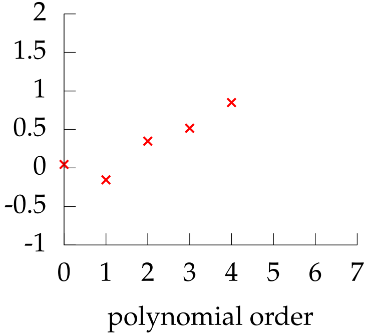
Leave One Out Error



Polynomial order 5, training error -29.524, leave one out error 1.48.

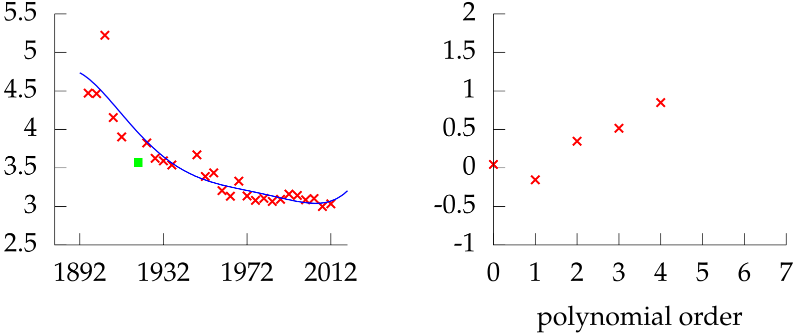
Leave One Out Error





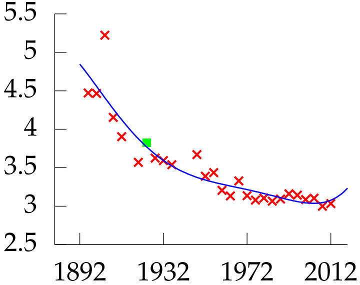
Polynomial order 5, training error -29.524, leave one out error 1.48.

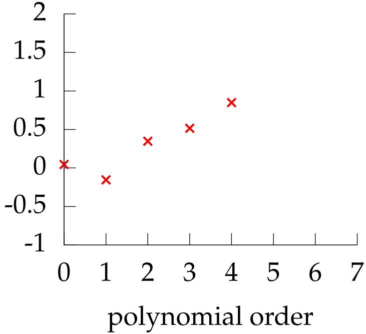
Leave One Out Error



Polynomial order 5, training error -29.524, leave one out error 1.48.

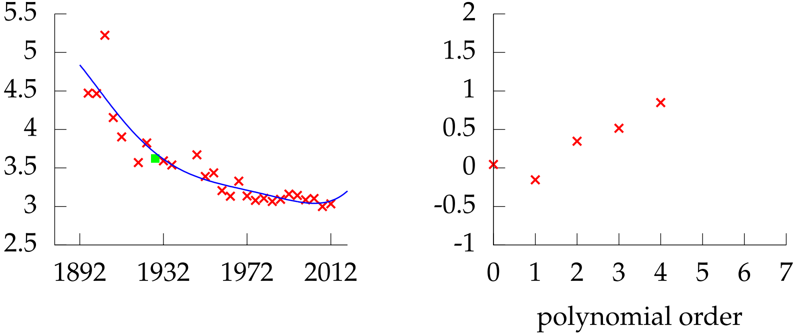
Leave One Out Error





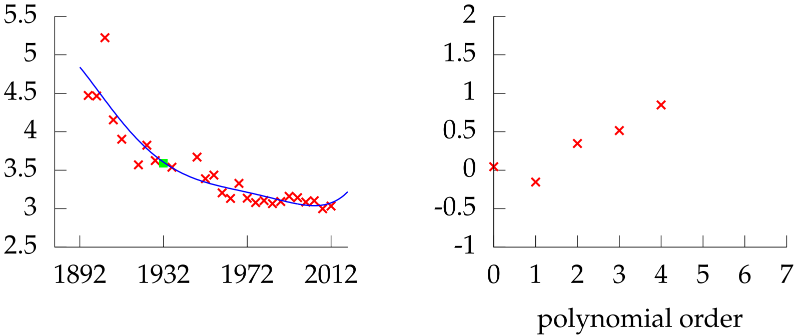
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



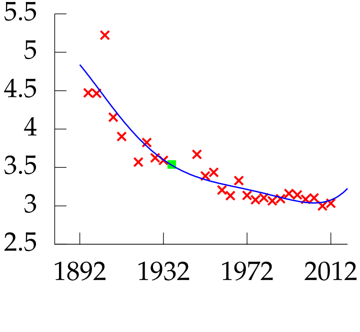
Polynomial order 5, training error -29.524, leave one out error 1.48.

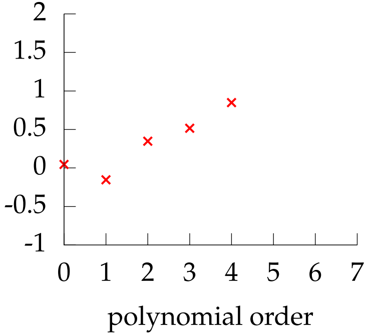
Leave One Out Error



Polynomial order 5, training error -29.524, leave one out error 1.48.

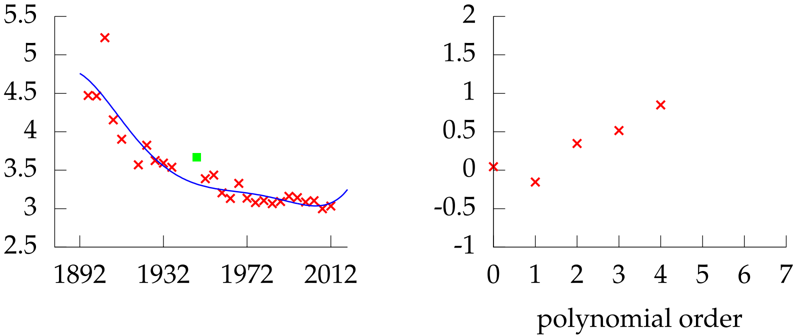
Leave One Out Error





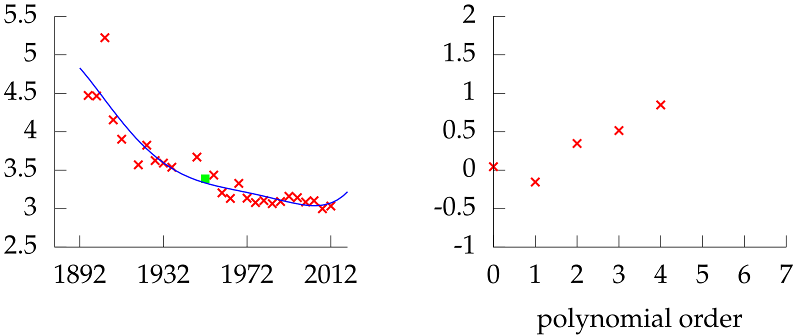
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



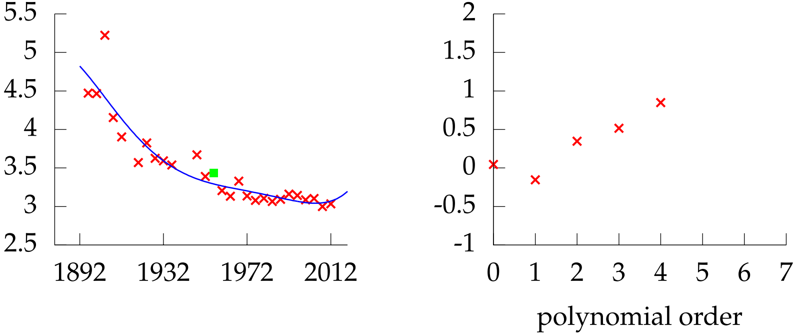
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



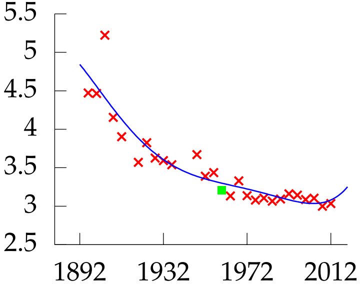
Polynomial order 5, training error -29.524, leave one out error 1.48.

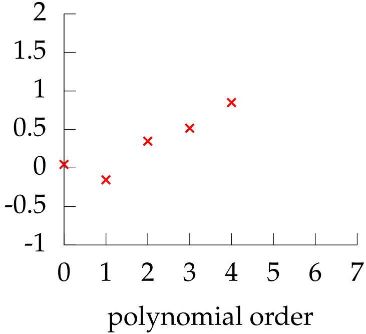
Leave One Out Error



Polynomial order 5, training error -29.524, leave one out error 1.48.

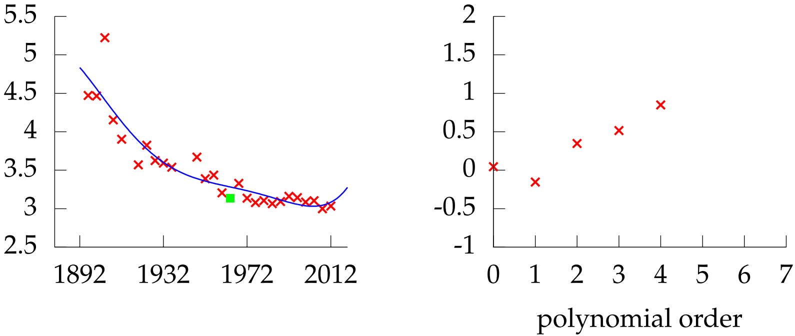
Leave One Out Error





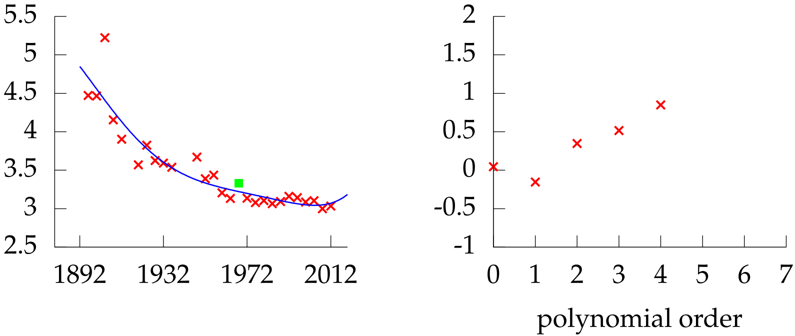
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



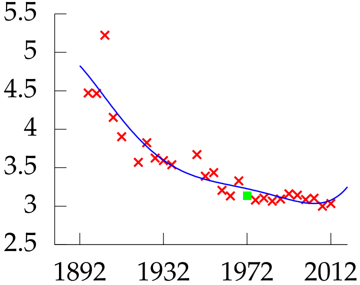
Polynomial order 5, training error -29.524, leave one out error 1.48.

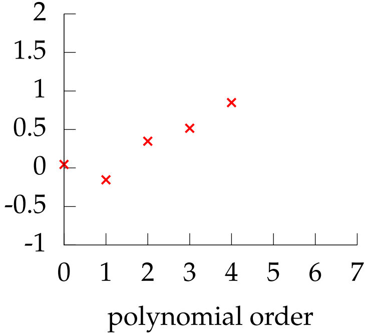
Leave One Out Error



Polynomial order 5, training error -29.524, leave one out error 1.48.

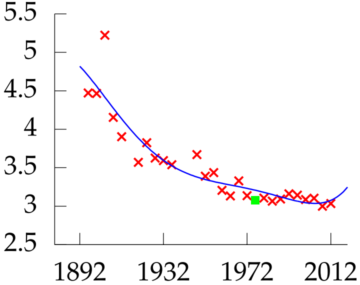
Leave One Out Error

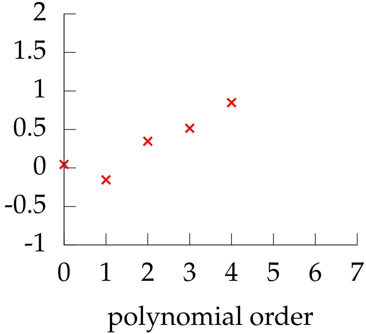




Polynomial order 5, training error -29.524, leave one out error 1.48.

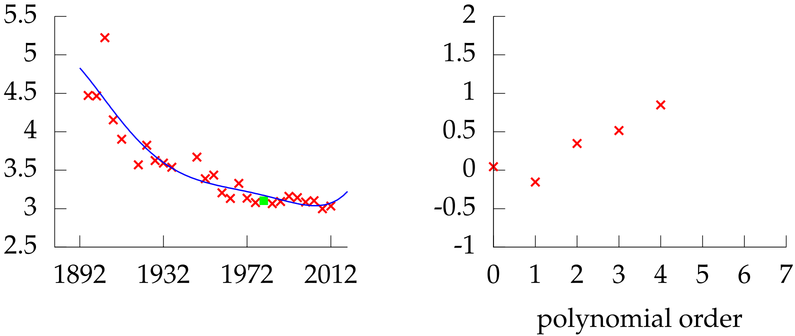
Leave One Out Error





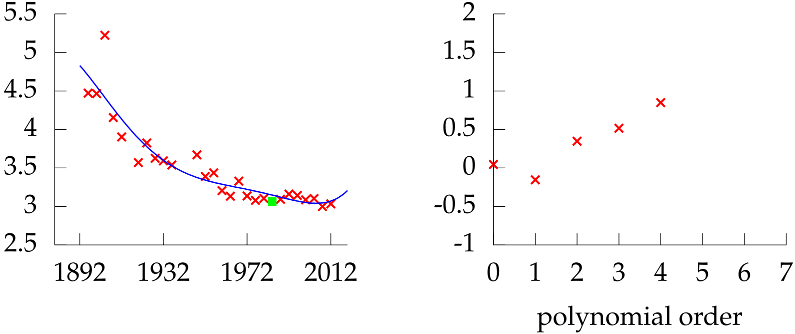
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



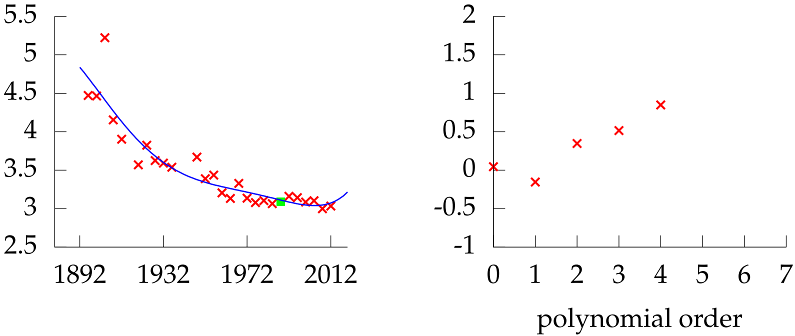
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



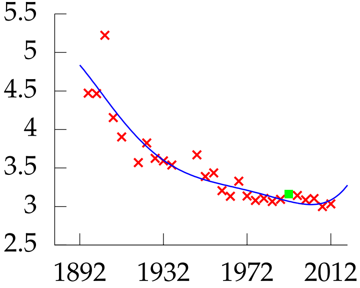
Polynomial order 5, training error -29.524, leave one out error 1.48.

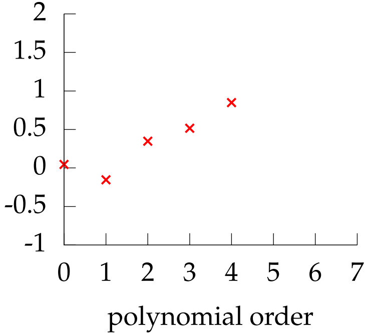
Leave One Out Error



Polynomial order 5, training error -29.524, leave one out error 1.48.

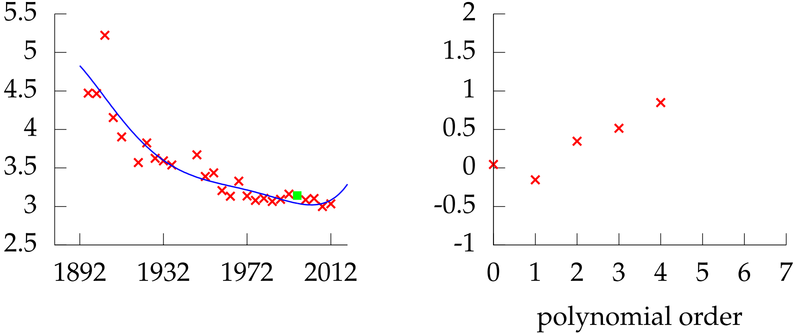
Leave One Out Error





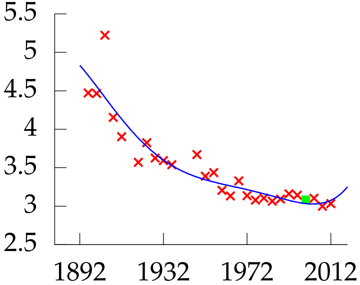
Polynomial order 5, training error -29.524, leave one out error 1.48.

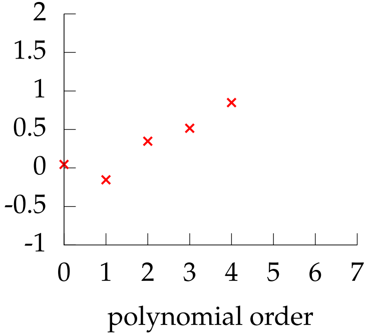
Leave One Out Error



Polynomial order 5, training error -29.524, leave one out error 1.48.

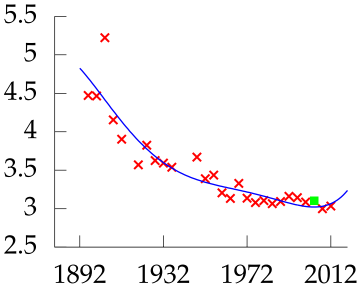
Leave One Out Error

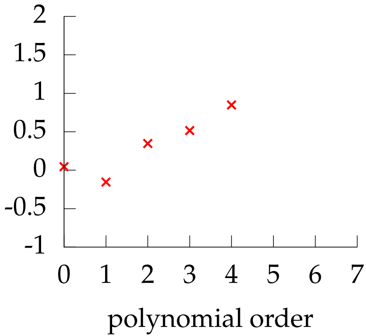




Polynomial order 5, training error -29.524, leave one out error 1.48.

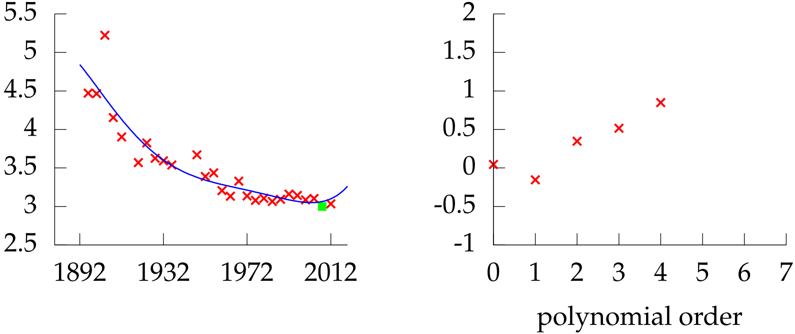
Leave One Out Error





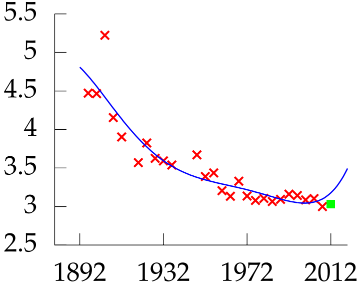
Polynomial order 5, training error -29.524, leave one out error 1.48.

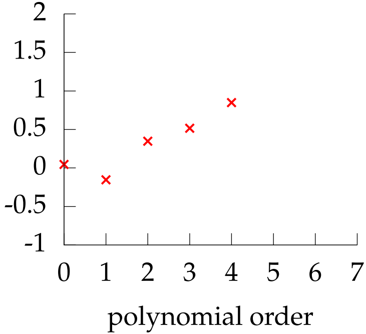
Leave One Out Error



Polynomial order 5, training error -29.524, leave one out error 1.48.

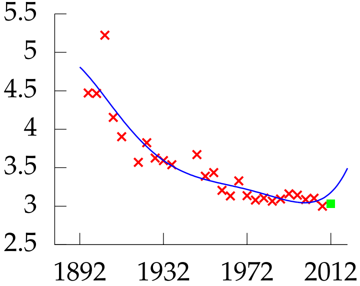
Leave One Out Error

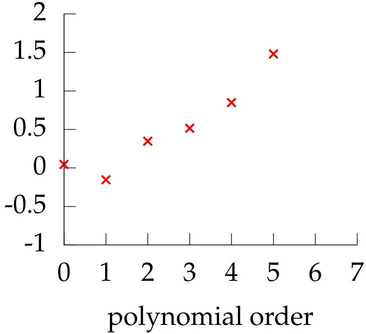




Polynomial order 5, training error -29.524, leave one out error 1.48.

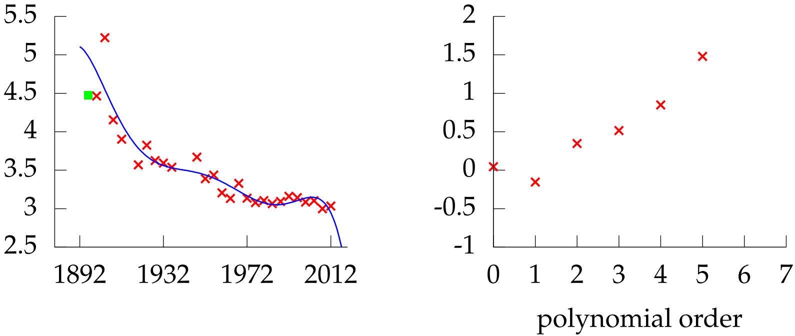
Leave One Out Error





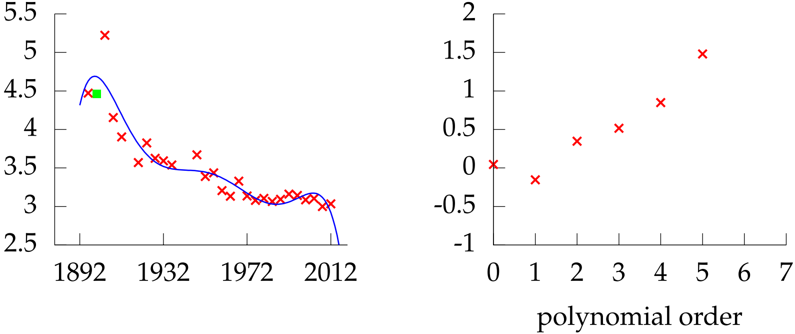
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



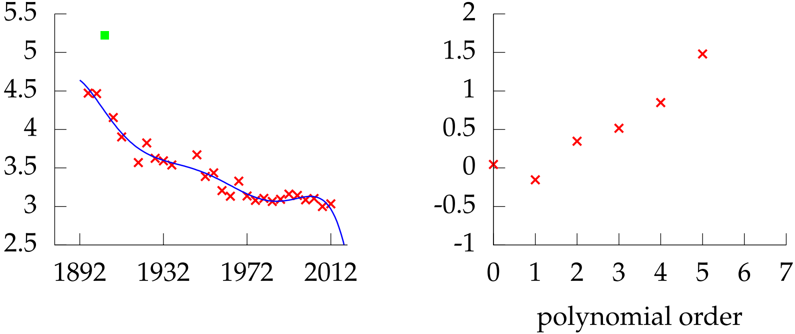
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



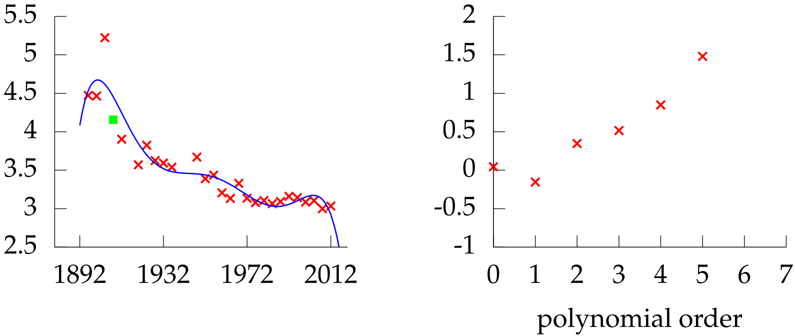
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



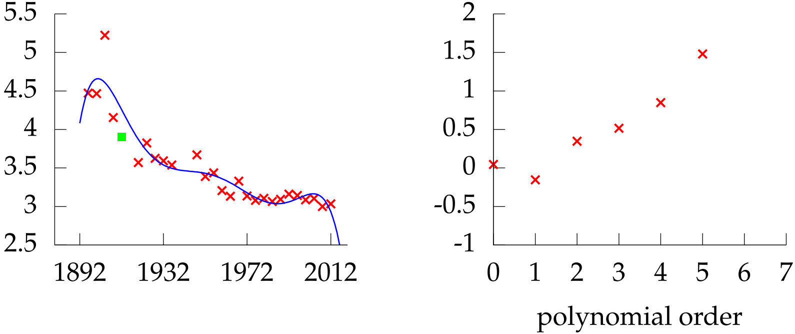
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



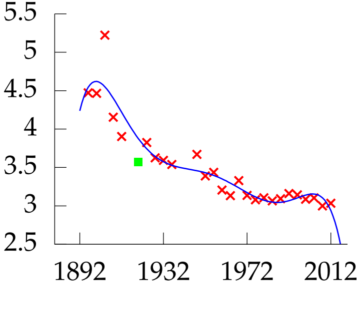
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



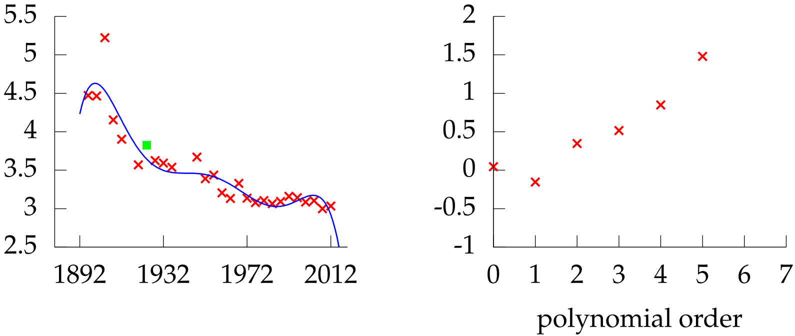
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



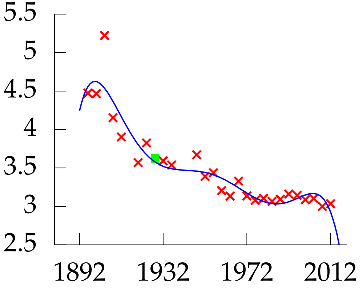
Polynomial order 6, training error -32.237, leave one out error 1.5047.

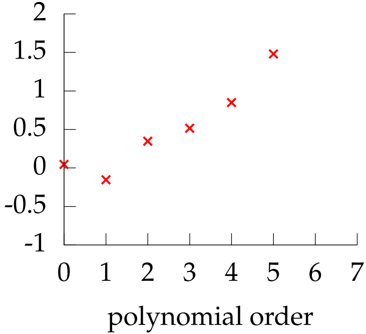
Leave One Out Error



Polynomial order 6, training error -32.237, leave one out error 1.5047.

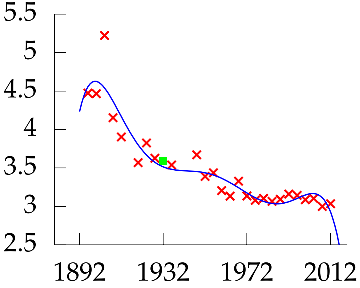
Leave One Out Error

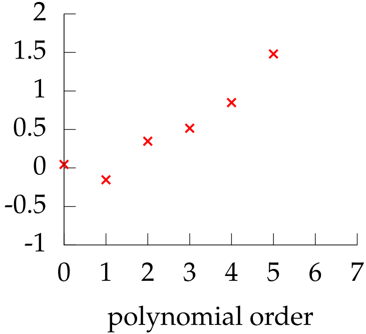




Polynomial order 6, training error -32.237, leave one out error 1.5047.

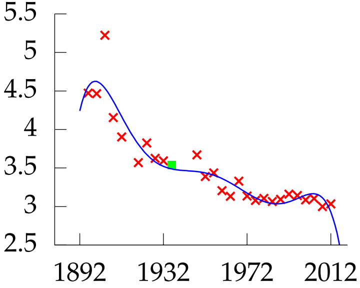
Leave One Out Error

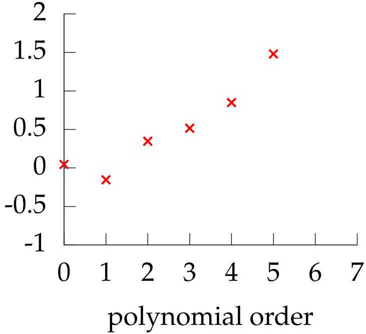




Polynomial order 6, training error -32.237, leave one out error 1.5047.

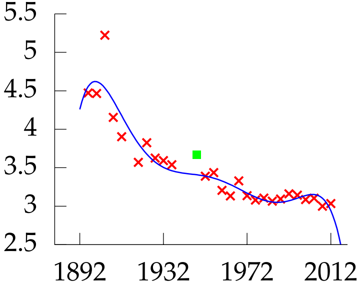
Leave One Out Error

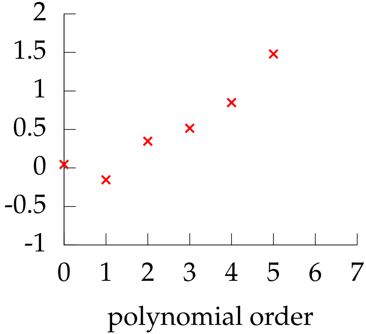




Polynomial order 6, training error -32.237, leave one out error 1.5047.

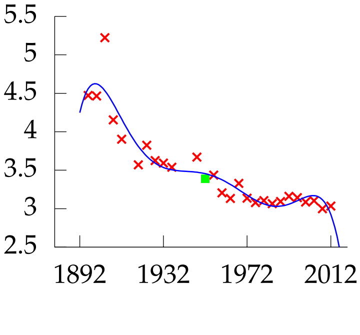
Leave One Out Error

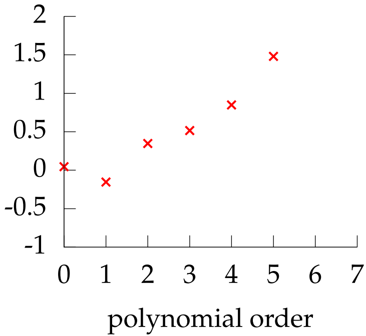




Polynomial order 6, training error -32.237, leave one out error 1.5047.

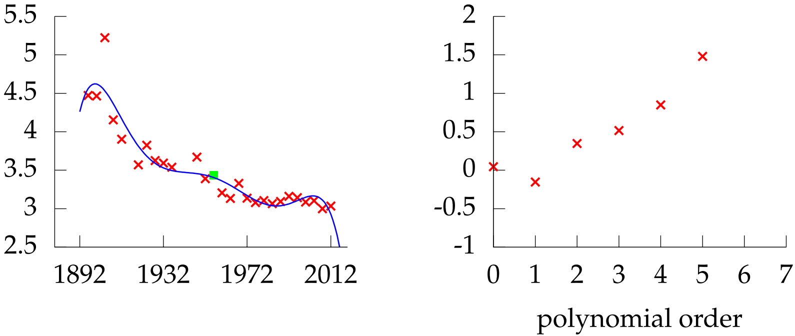
Leave One Out Error





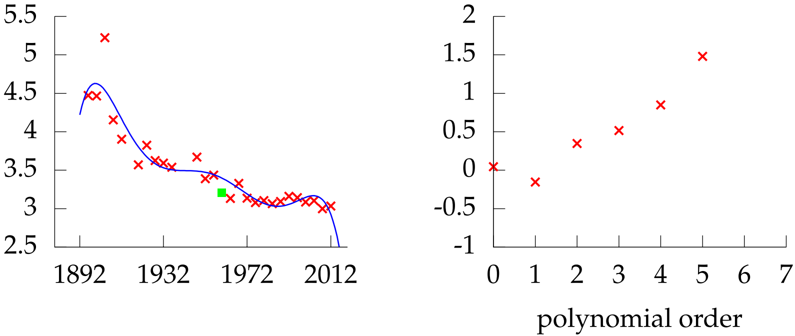
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



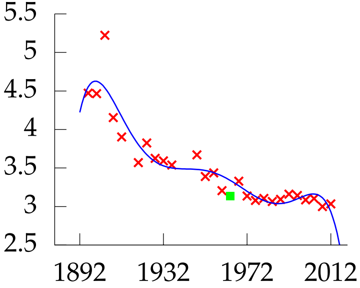
Polynomial order 6, training error -32.237, leave one out error 1.5047.

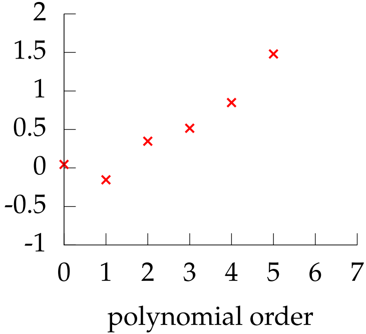
Leave One Out Error



Polynomial order 6, training error -32.237, leave one out error 1.5047.

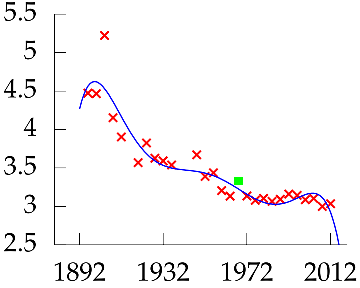
Leave One Out Error

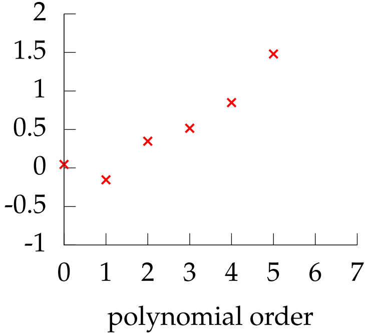




Polynomial order 6, training error -32.237, leave one out error 1.5047.

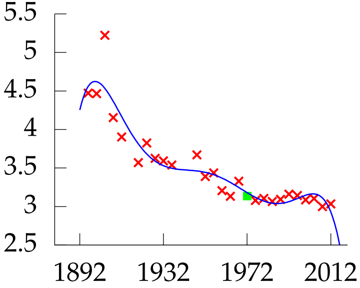
Leave One Out Error

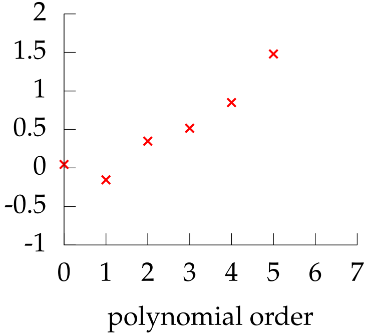




Polynomial order 6, training error -32.237, leave one out error 1.5047.

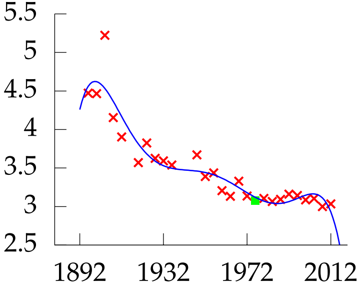
Leave One Out Error

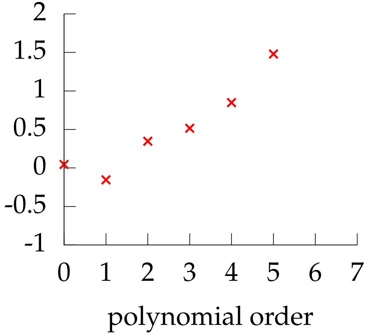




Polynomial order 6, training error -32.237, leave one out error 1.5047.

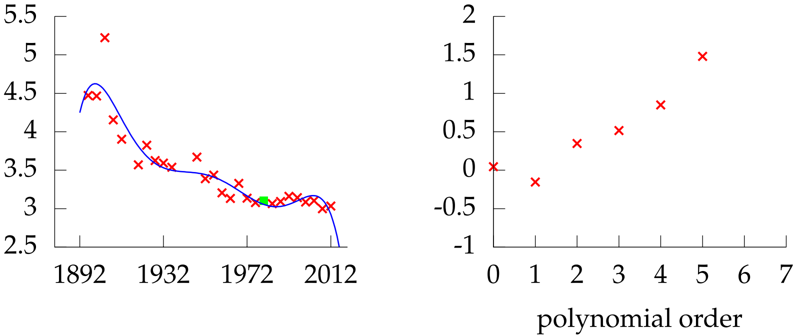
Leave One Out Error





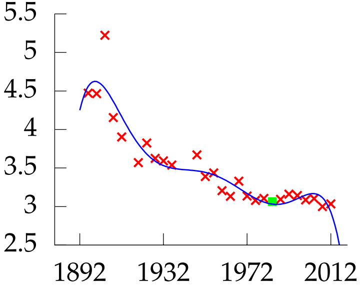
Polynomial order 6, training error -32.237, leave one out error 1.5047.

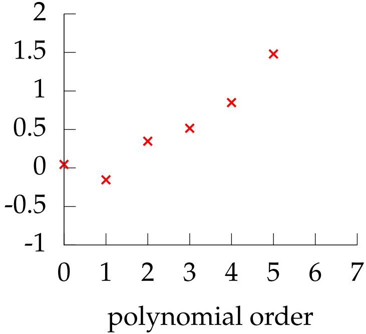
Leave One Out Error



Polynomial order 6, training error -32.237, leave one out error 1.5047.

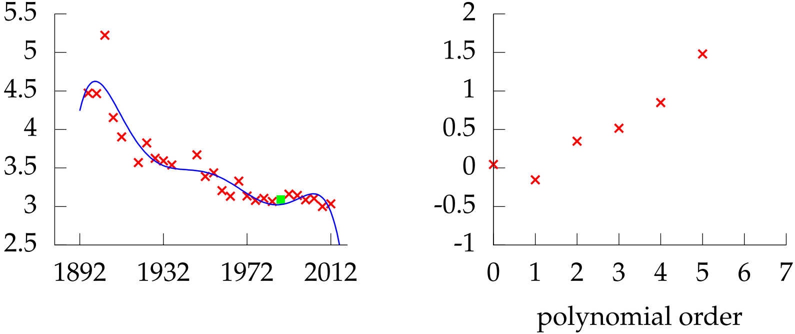
Leave One Out Error





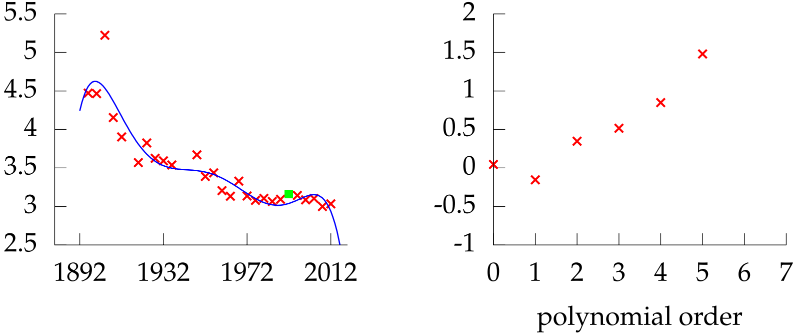
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



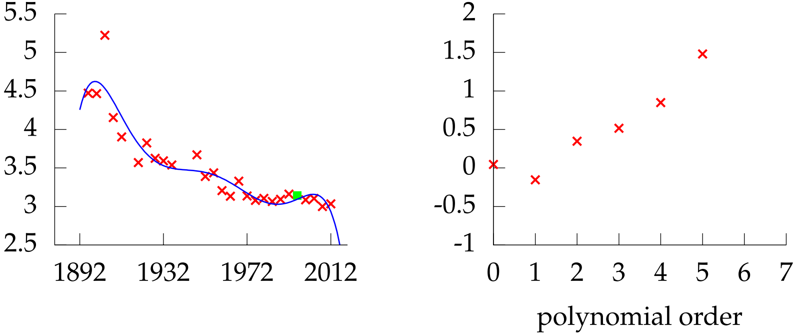
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



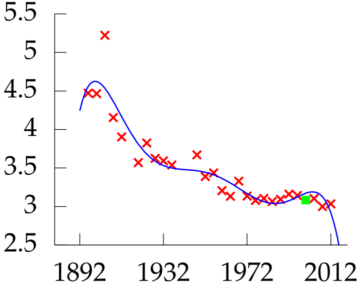
Polynomial order 6, training error -32.237, leave one out error 1.5047.

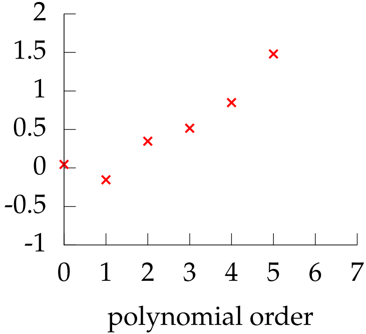
Leave One Out Error



Polynomial order 6, training error -32.237, leave one out error 1.5047.

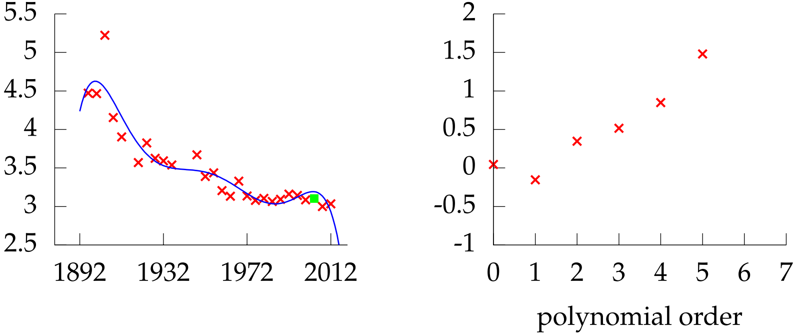
Leave One Out Error





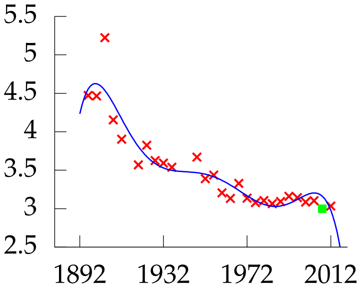
Polynomial order 6, training error -32.237, leave one out error 1.5047.

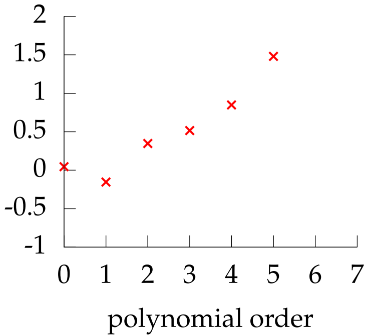
Leave One Out Error



Polynomial order 6, training error -32.237, leave one out error 1.5047.

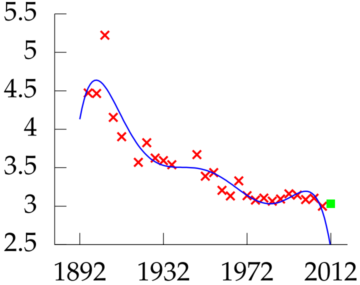
Leave One Out Error

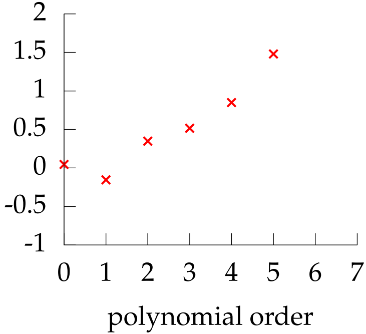




Polynomial order 6, training error -32.237, leave one out error 1.5047.

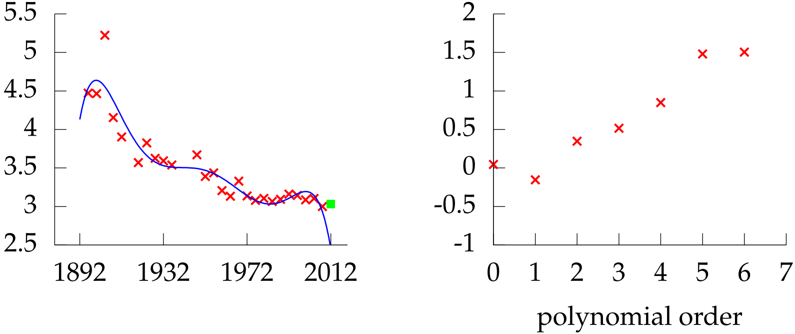
Leave One Out Error





Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error

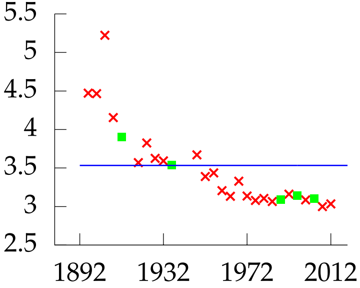


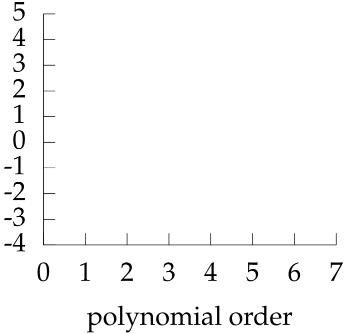
Polynomial order 6, training error -32.237, leave one out error 1.5047.

k Fold Cross Validation

- Leave one out cross validation can be very time consuming!
- Need to train your algorithm N times.
- An alternative: k fold cross validation.

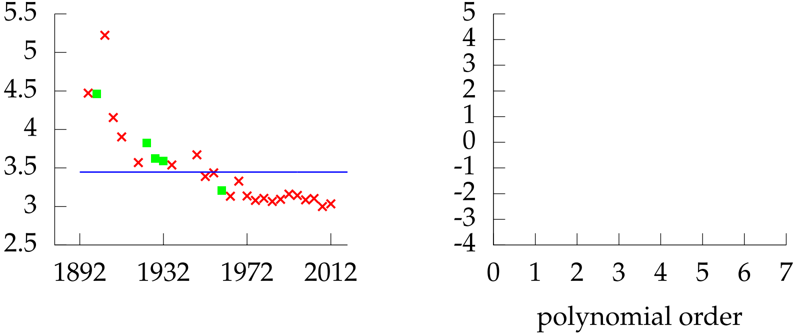
Cross Validation Error





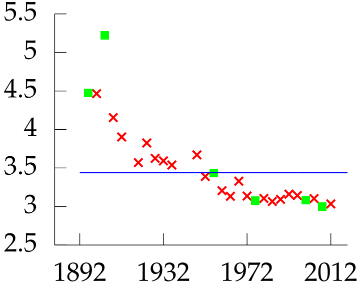
Polynomial order 0, training error -3.2644, leave one out error 0.045811.

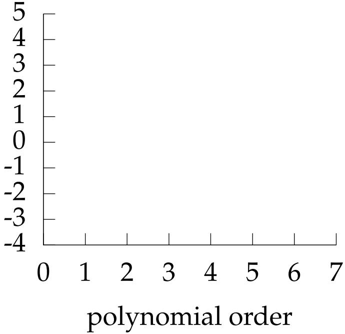
Cross Validation Error



Polynomial order 0, training error -3.2644, leave one out error 0.045811.

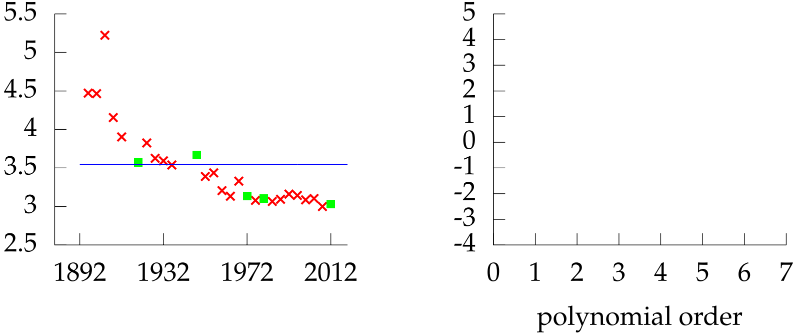
Cross Validation Error





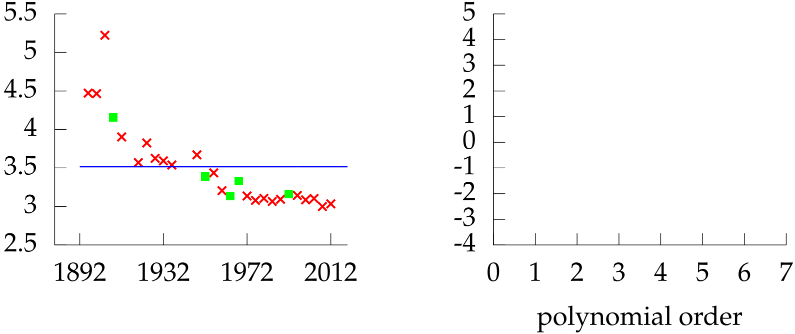
Polynomial order 0, training error -3.2644, leave one out error 0.045811.

Cross Validation Error



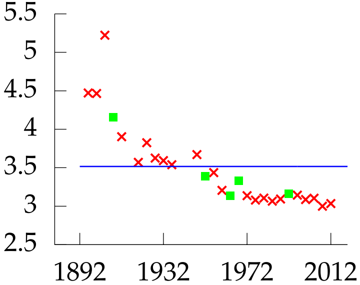
Polynomial order 0, training error -3.2644, leave one out error 0.045811.

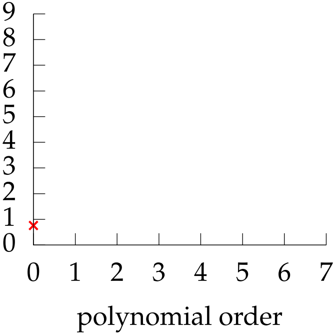
Cross Validation Error



Polynomial order 0, training error -3.2644, leave one out error 0.045811.

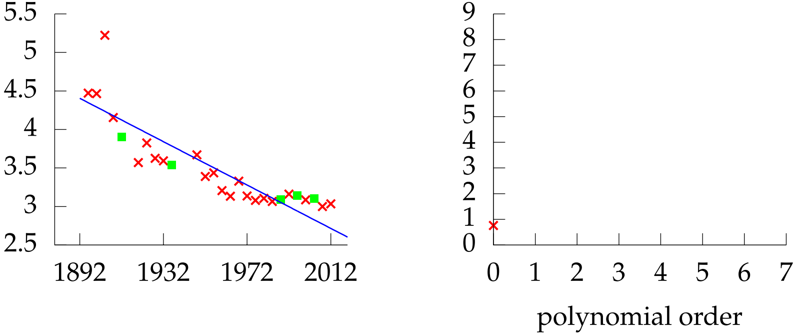
Cross Validation Error





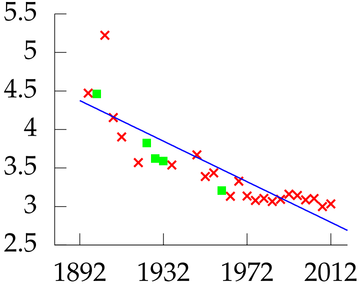
Polynomial order 0, training error -3.2644, leave one out error 0.045811.

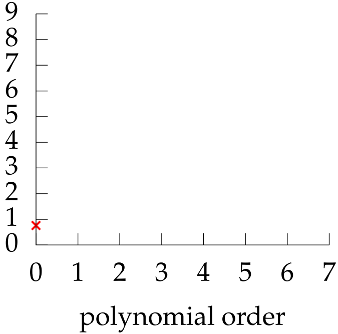
Cross Validation Error



Polynomial order 1, training error -18.873, leave one out error -0.15413.

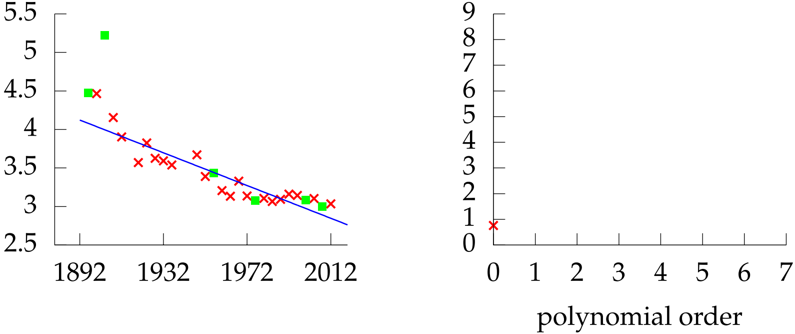
Cross Validation Error





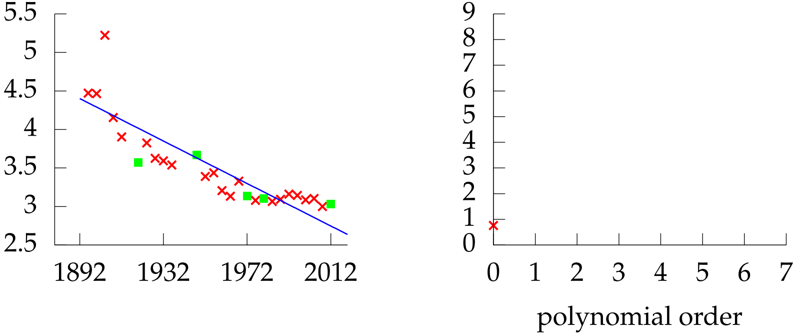
Polynomial order 1, training error -18.873, leave one out error -0.15413.

Cross Validation Error



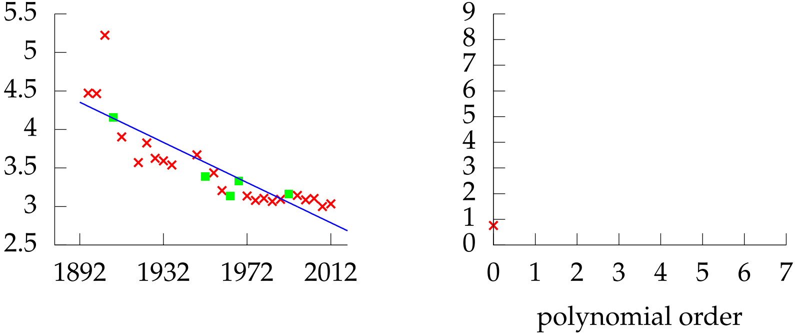
Polynomial order 1, training error -18.873, leave one out error -0.15413.

Cross Validation Error



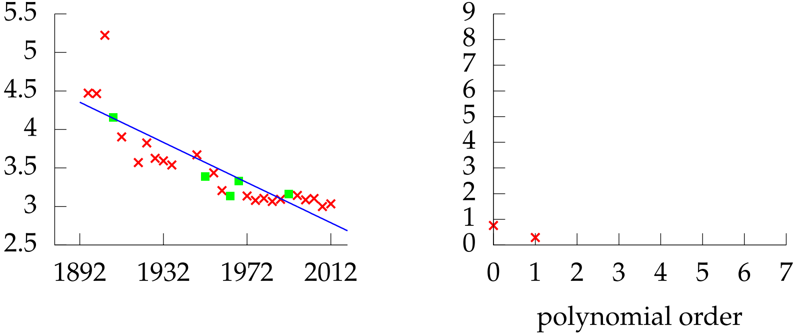
Polynomial order 1, training error -18.873, leave one out error -0.15413.

Cross Validation Error



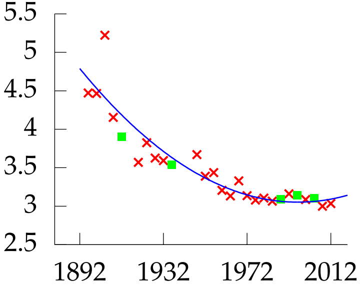
Polynomial order 1, training error -18.873, leave one out error -0.15413.

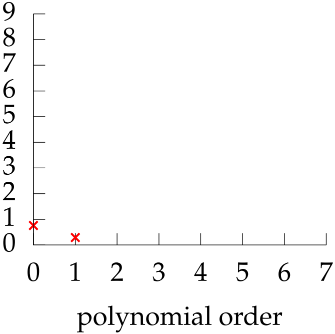
Cross Validation Error



Polynomial order 1, training error -18.873, leave one out error -0.15413.

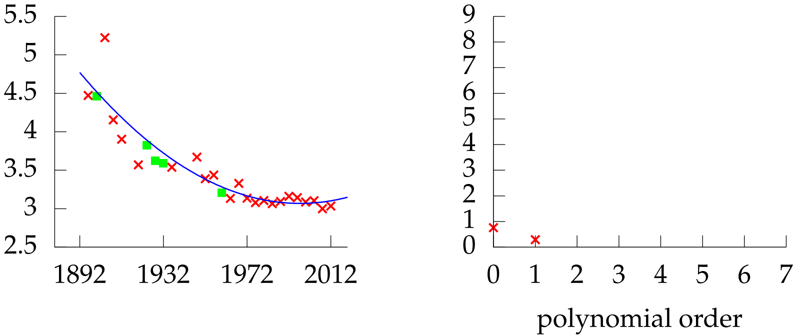
Cross Validation Error





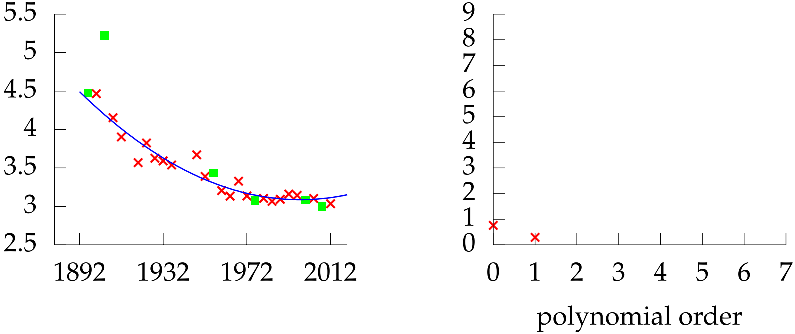
Polynomial order 2, training error -25.177, leave one out error 0.34669.

Cross Validation Error



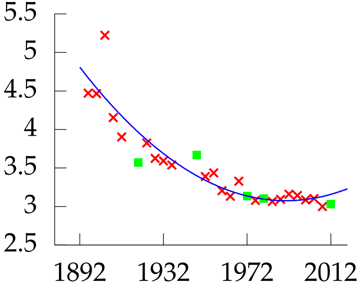
Polynomial order 2, training error -25.177, leave one out error 0.34669.

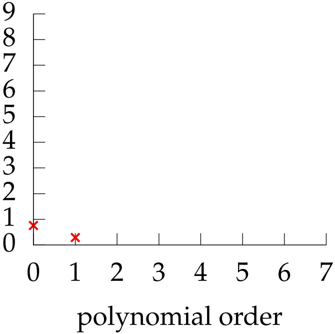
Cross Validation Error



Polynomial order 2, training error -25.177, leave one out error 0.34669.

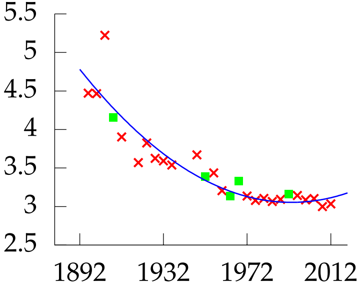
Cross Validation Error

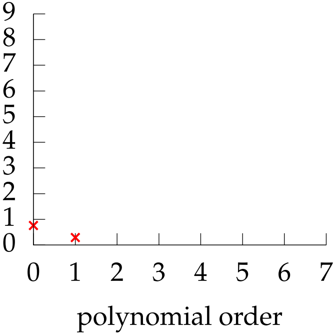




Polynomial order 2, training error -25.177, leave one out error 0.34669.

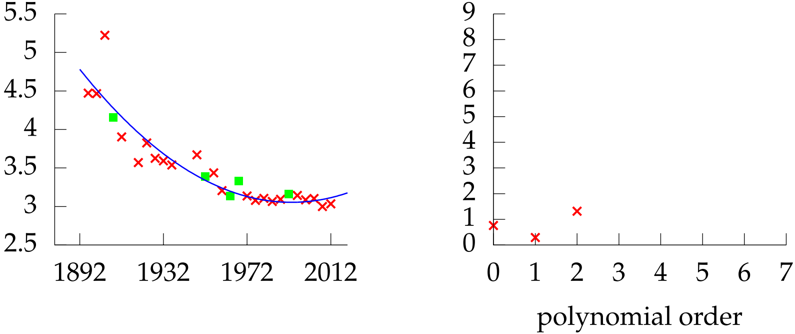
Cross Validation Error





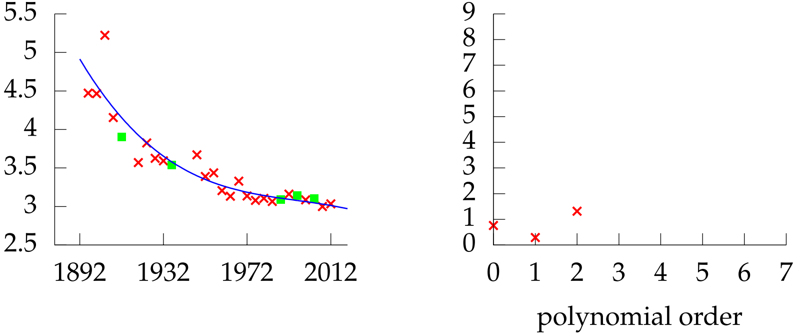
Polynomial order 2, training error -25.177, leave one out error 0.34669.

Cross Validation Error



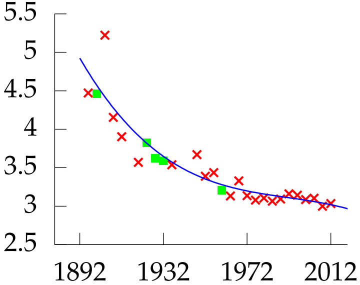
Polynomial order 2, training error -25.177, leave one out error 0.34669.

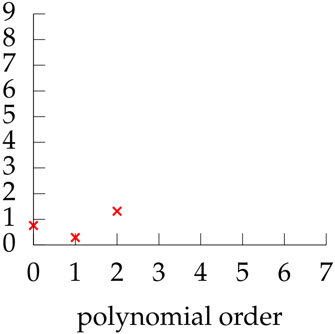
Cross Validation Error



Polynomial order 3, training error -25.777, leave one out error 0.51621.

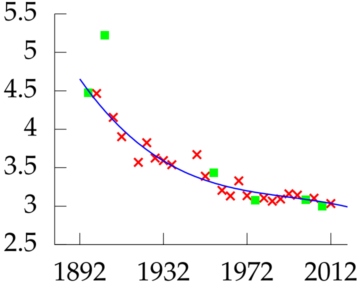
Cross Validation Error

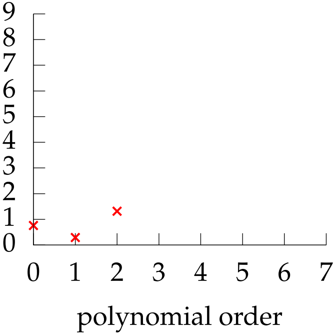




Polynomial order 3, training error -25.777, leave one out error 0.51621.

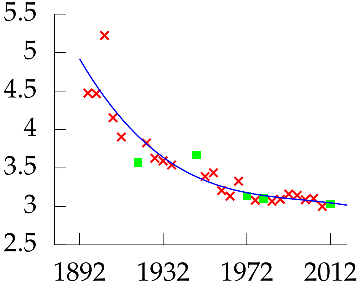
Cross Validation Error

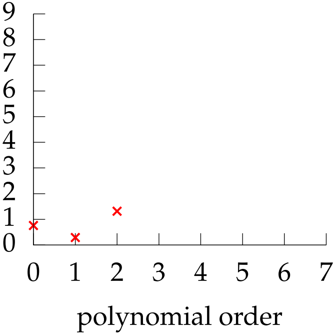




Polynomial order 3, training error -25.777, leave one out error 0.51621.

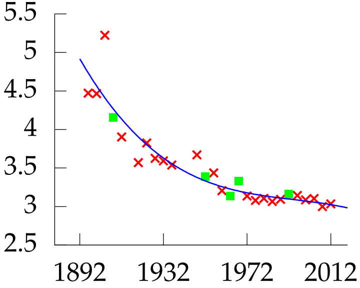
Cross Validation Error

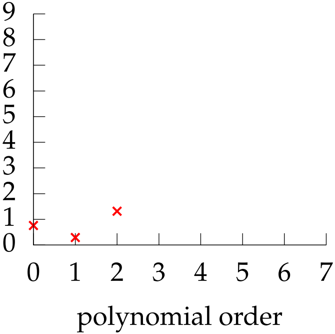




Polynomial order 3, training error -25.777, leave one out error 0.51621.

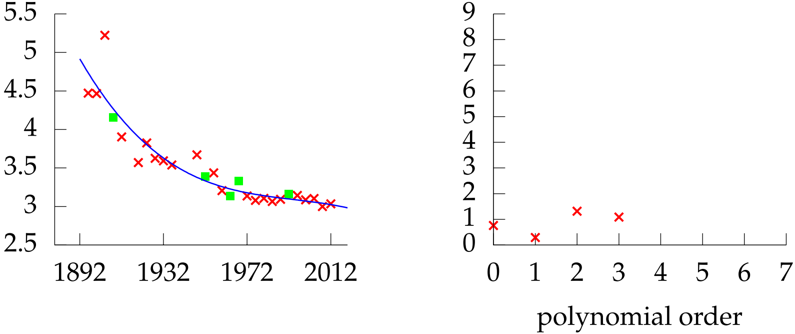
Cross Validation Error





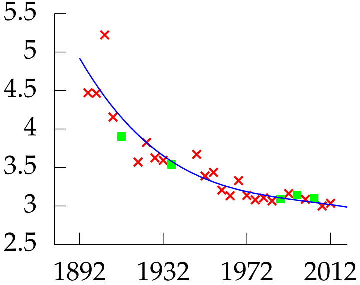
Polynomial order 3, training error -25.777, leave one out error 0.51621.

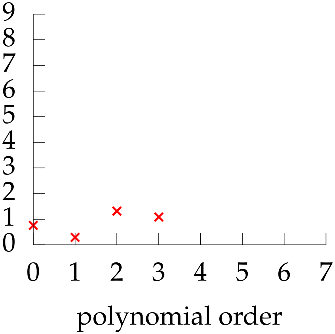
Cross Validation Error



Polynomial order 3, training error -25.777, leave one out error 0.51621.

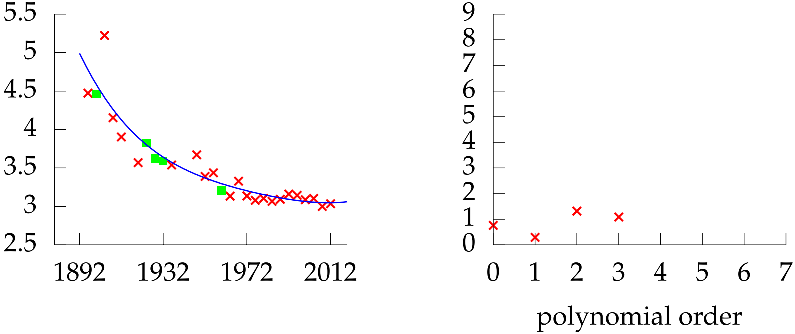
Cross Validation Error





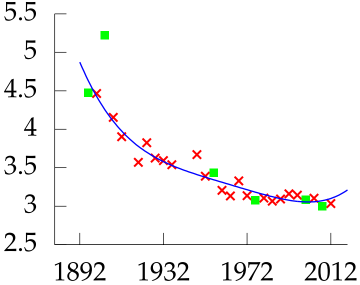
Polynomial order 4, training error -26.048, leave one out error 0.84844.

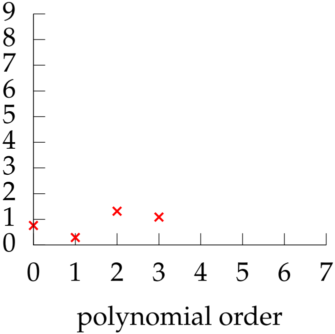
Cross Validation Error



Polynomial order 4, training error -26.048, leave one out error 0.84844.

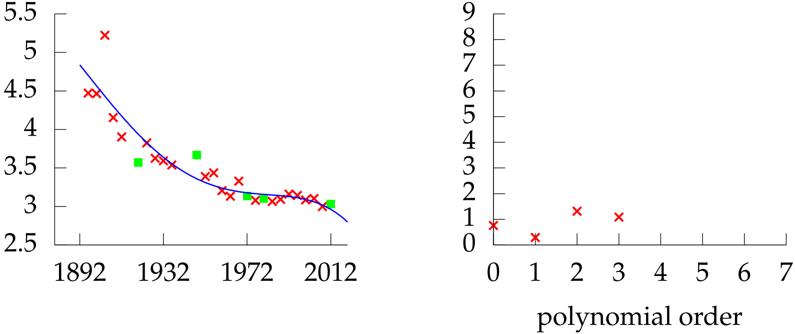
Cross Validation Error





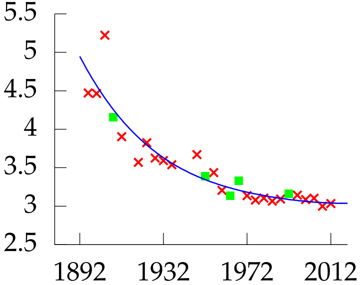
Polynomial order 4, training error -26.048, leave one out error 0.84844.

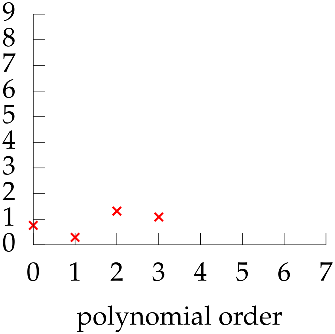
Cross Validation Error



Polynomial order 4, training error -26.048, leave one out error 0.84844.

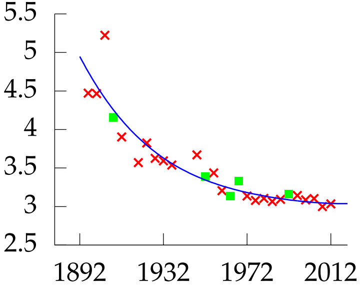
Cross Validation Error

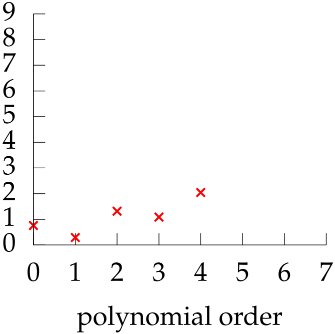




Polynomial order 4, training error -26.048, leave one out error 0.84844.

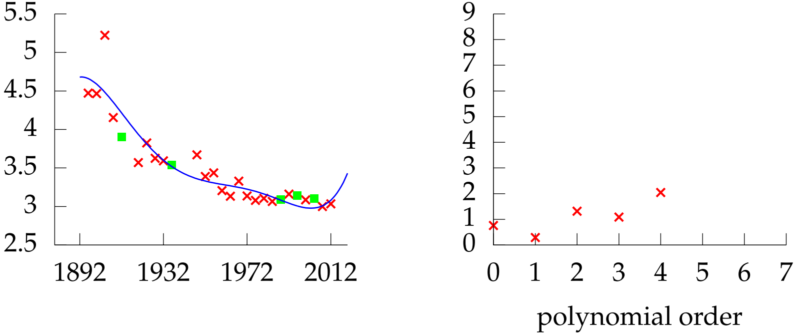
Cross Validation Error





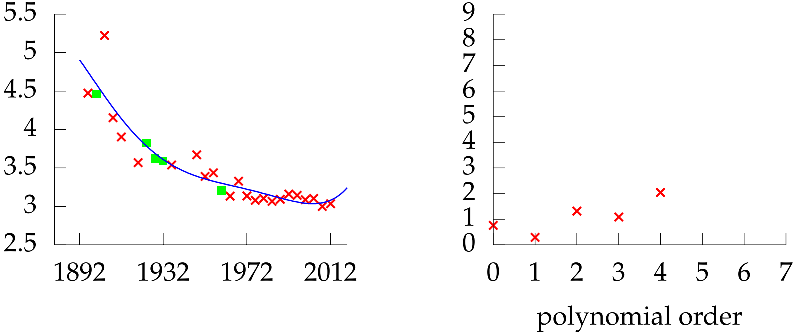
Polynomial order 4, training error -26.048, leave one out error 0.84844.

Cross Validation Error



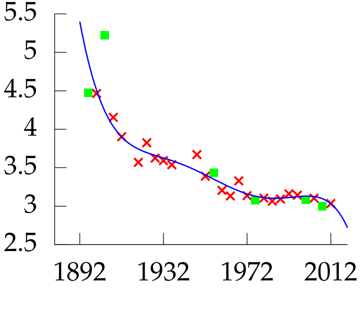
Polynomial order 5, training error -26.892, leave one out error 1.48.

Cross Validation Error



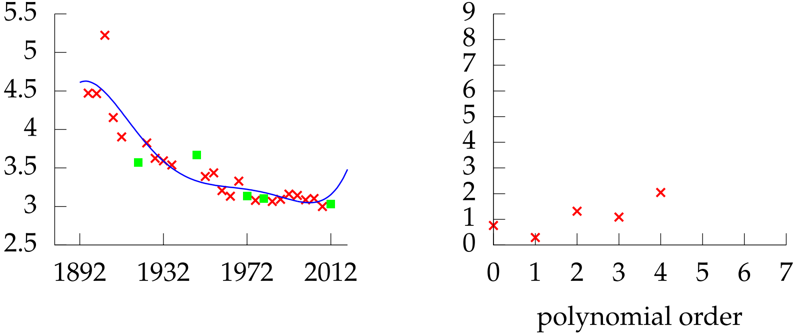
Polynomial order 5, training error -26.892, leave one out error 1.48.

Cross Validation Error



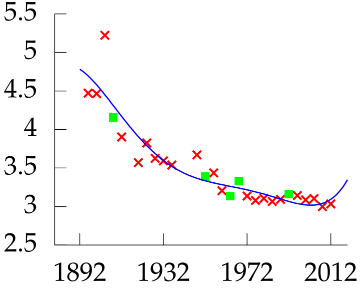
Polynomial order 5, training error -26.892, leave one out error 1.48.

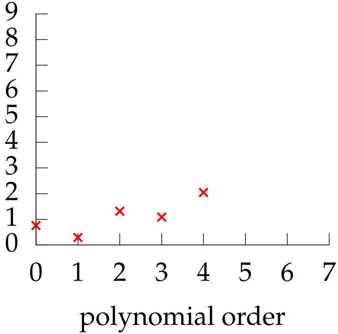
Cross Validation Error



Polynomial order 5, training error -26.892, leave one out error 1.48.

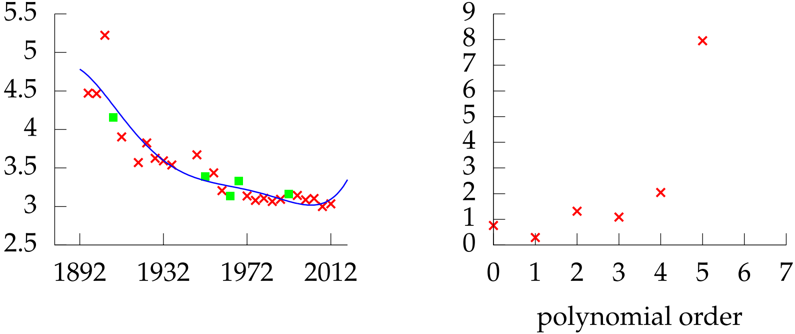
Cross Validation Error





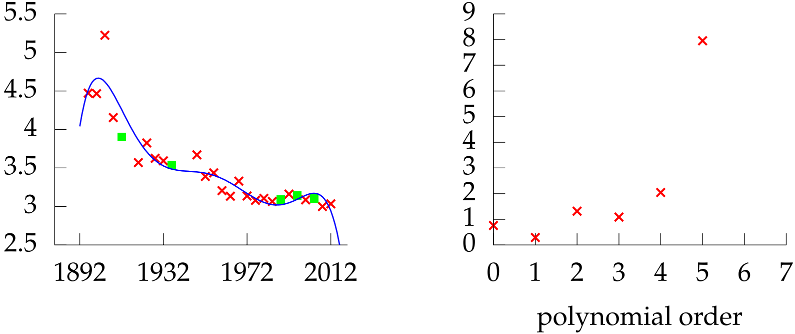
Polynomial order 5, training error -26.892, leave one out error 1.48.

Cross Validation Error



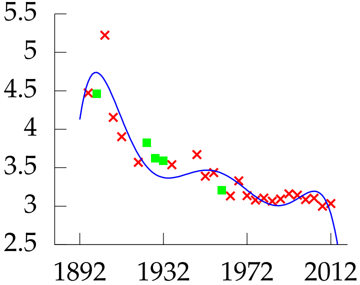
Polynomial order 5, training error -26.892, leave one out error 1.48.

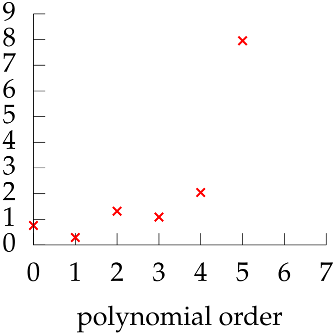
Cross Validation Error



Polynomial order 6, training error -29.395, leave one out error 1.5047.

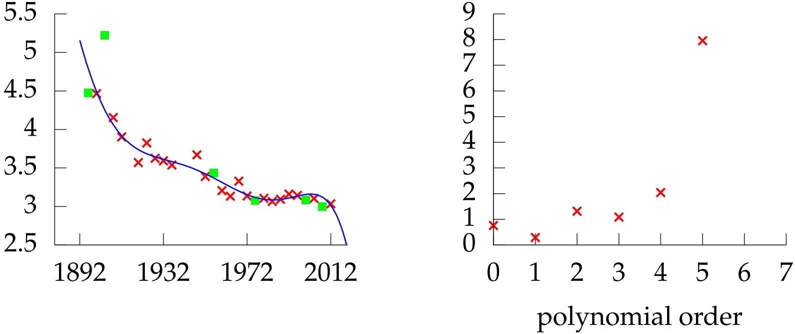
Cross Validation Error





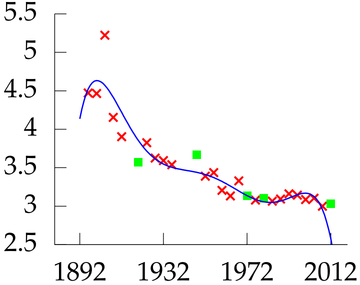
Polynomial order 6, training error -29.395, leave one out error 1.5047.

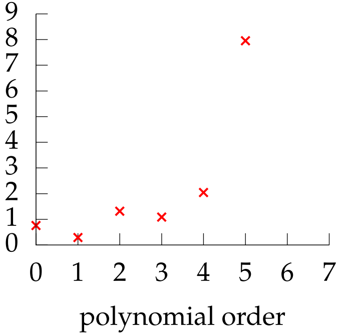
Cross Validation Error



Polynomial order 6, training error -29.395, leave one out error 1.5047.

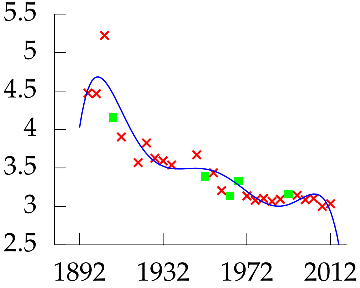
Cross Validation Error

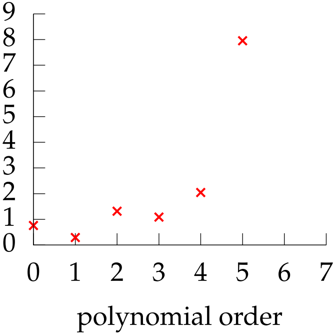




Polynomial order 6, training error -29.395, leave one out error 1.5047.

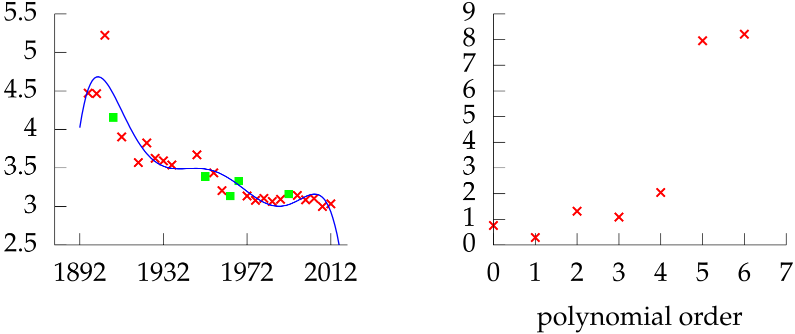
Cross Validation Error





Polynomial order 6, training error -29.395, leave one out error 1.5047.

Cross Validation Error



Polynomial order 6, training error -29.395, leave one out error 1.5047.

Reading

- Section 1.5 of Rogers and Girolami.

Outline

Basis Functions

Fitting Basis Functions

Generalization

Review: Overdetermined Systems

Underdetermined Systems

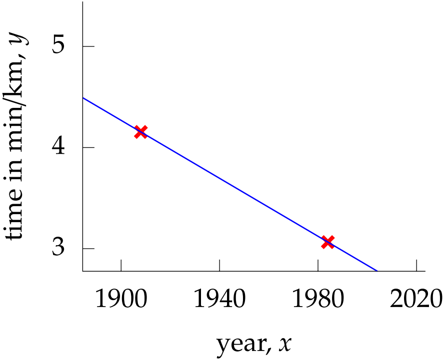
Bayesian Perspective

Bayesian Regression

Two Simultaneous Equations

A system of two simultaneous equations with two unknowns.

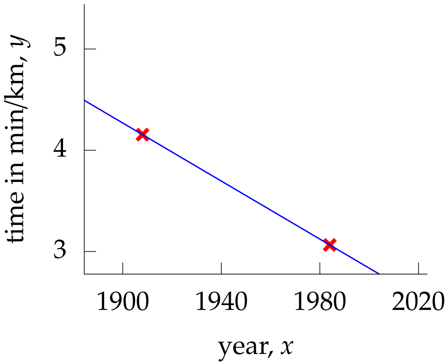
$$y_1 = mx_1 + c$$
$$y_2 = mx_2 + c$$



Two Simultaneous Equations

A system of two simultaneous equations with two unknowns.

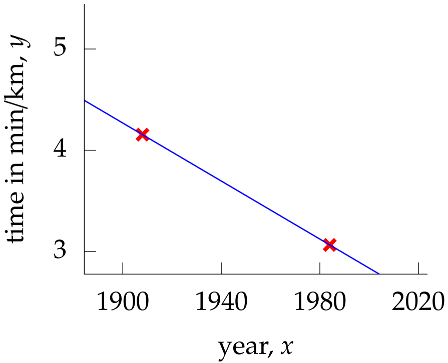
$$y_1 - y_2 = m(x_1 - x_2)$$



Two Simultaneous Equations

A system of two simultaneous equations with two unknowns.

$$\frac{y_1 - y_2}{x_1 - x_2} = m$$

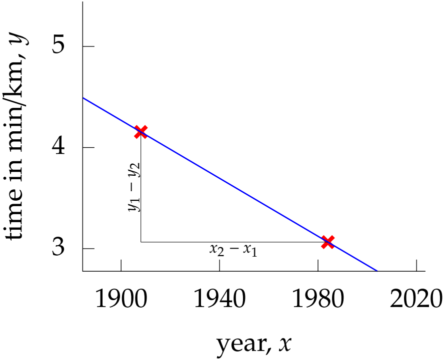


Two Simultaneous Equations

A system of two simultaneous equations with two unknowns.

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

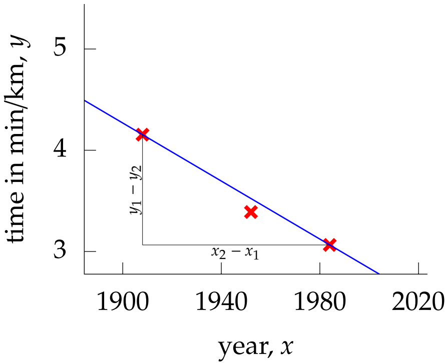
$$c = y_1 - m x_1$$



Two Simultaneous Equations

How do we deal with three simultaneous equations with only two unknowns?

$$y_1 = mx_1 + c$$
$$y_2 = mx_2 + c$$
$$y_3 = mx_3 + c$$



Overdetermined System

- With two unknowns and two observations:

$$y_1 = mx_1 + c$$
$$y_2 = mx_2 + c$$

Overdetermined System

- With two unknowns and two observations:

$$y_1 = mx_1 + c$$
$$y_2 = mx_2 + c$$

- Additional observation leads to *overdetermined* system.

$$y_3 = mx_3 + c$$

Overdetermined System

- With two unknowns and two observations:

$$y_1 = mx_1 + c$$
$$y_2 = mx_2 + c$$

- Additional observation leads to *overdetermined* system.

$$y_3 = mx_3 + c$$

- This problem is solved through a noise model $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$

$$y_1 = mx_1 + c + \epsilon_1$$
$$y_2 = mx_2 + c + \epsilon_2$$
$$y_3 = mx_3 + c + \epsilon_3$$

Noise Models

- We aren't modeling entire system.
- Noise model gives mismatch between model and data.
- Gaussian model justified by appeal to central limit theorem.
- Other models also possible (Student-t for heavy tails).
- Maximum likelihood with Gaussian noise leads to *least squares*.

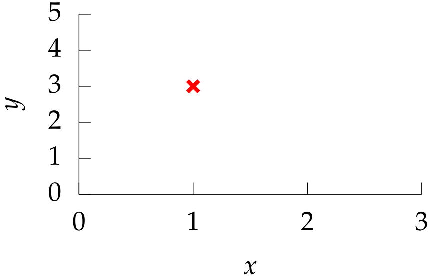
Outline

- Basis Functions
- Fitting Basis Functions
- Generalization
- Review: Overdetermined Systems
- **Underdetermined Systems**
- Bayesian Perspective
- Bayesian Regression

Underdetermined System

What about two unknowns and *one* observation?

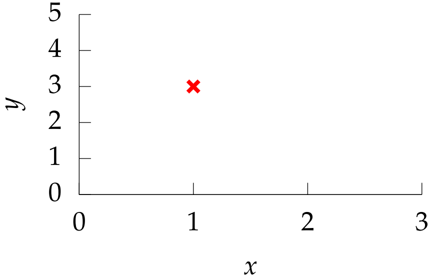
$$y_1 = mx_1 + c$$



Underdetermined System

Can compute m given c.

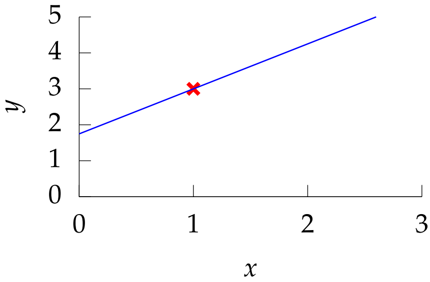
$$m = \frac{y_1 - c}{x}$$



Underdetermined System

Can compute m given c.

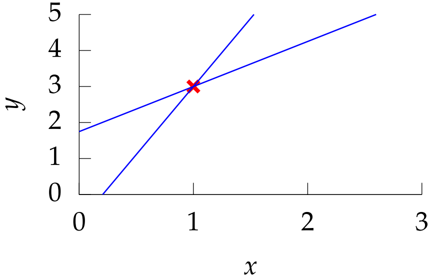
$$c = 1.75 \Longrightarrow m = 1.25$$



Underdetermined System

Can compute m given c.

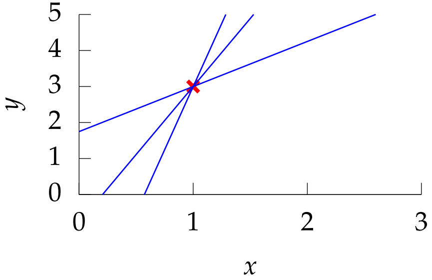
$$c = -0.777 \implies m = 3.78$$



Underdetermined System

Can compute m given c.

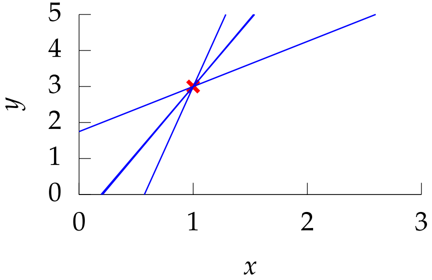
$$c = -4.01 \Longrightarrow m = 7.01$$



Underdetermined System

Can compute m given c.

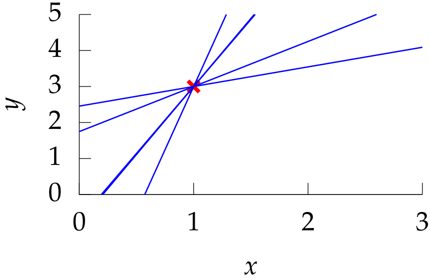
$$c = -0.718 \implies m = 3.72$$



Underdetermined System

Can compute m given c.

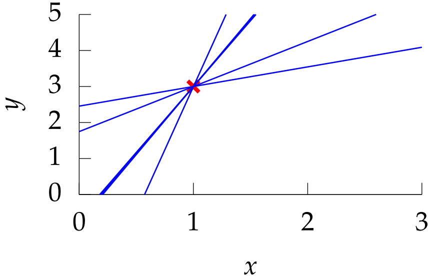
$$c = 2.45 \Longrightarrow m = 0.545$$



Underdetermined System

Can compute m given c.

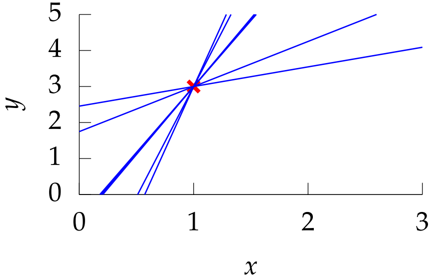
$$c = -0.657 \implies m = 3.66$$



Underdetermined System

Can compute m given c.

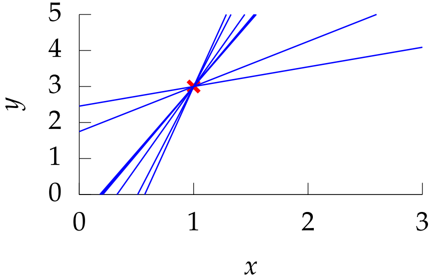
$$c = -3.13 \implies m = 6.13$$



Underdetermined System

Can compute m given c.

$$c = -1.47 \Longrightarrow m = 4.47$$

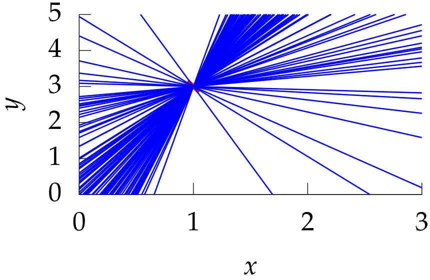


Underdetermined System

Can compute m given c. Assume

$$c \sim \mathcal{N}(0, 4),$$

we find a distribution of solutions.



Different Types of Uncertainty

- The first type of uncertainty we are assuming is *aleatoric* uncertainty.
- The second type of uncertainty we are assuming is *epistemic* uncertainty.

Aleatoric Uncertainty

- This is uncertainty we couldn't know even if we wanted to. e.g. the result of a football match before it's played.
- Where a sheet of paper might land on the floor.

Outline

Bayesian Approach

- Likelihood for the regression example has the form

$$p(\mathbf{y}|\mathbf{w}, \sigma^2) = \prod_{i=1}^{N} \mathcal{N}\left(y_i|\mathbf{w}^\top \boldsymbol{\phi}_i, \sigma^2\right).$$

- Suggestion was to maximize this likelihood with respect to $\mathbf{w}$.
- This can be done with gradient based optimization of the log likelihood.
- Alternative approach: integration across $\mathbf{w}$.
- Consider expected value of likelihood under a range of potential $\mathbf{w}$s.
- This is known as the *Bayesian* approach.

Note on the Term Bayesian

- We will use Bayes' rule to invert probabilities in the Bayesian approach.
  - Bayesian is not named after Bayes' rule (v. common confusion).
  - The term Bayesian refers to the treatment of the parameters as stochastic variables.
  - This approach was proposed by Laplace (1774) and Bayes (1763) independently.
  - For early statisticians this was very controversial (Fisher et al).

Bernoulli Distribution

Bernoulli Distribution

- Jacob Bernoulli described this distribution in terms of an ‘urn’.

Bernoulli Distribution

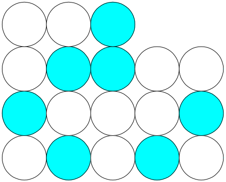
- Jacob Bernoulli described this distribution in terms of an 'urn'.
- Write as a function

$$P(Y = y) = \pi^y (1 - \pi)^{1-y}$$

Bernoulli Distribution

- Jacob Bernoulli described this distribution in terms of an ‘urn’.
- Write as a function

$$P(Y = y) = \pi^y (1 - \pi)^{1-y}$$

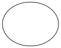


Bernoulli Distribution Revisited

- Thomas Bayes considered a ball landing uniformly across a table.

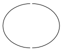
Bernoulli Distribution Revisited

- Thomas Bayes considered a ball landing uniformly across a table.



Bernoulli Distribution Revisited

- Thomas Bayes considered a ball landing uniformly across a table.
- And another ball landing on the left or right (Bayes, 1763, page 385).

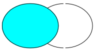


Bernoulli Distribution Revisited

- Thomas Bayes considered a ball landing uniformly across a table.
- And another ball landing on the left or right (Bayes, 1763, page 385).

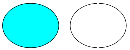
Bernoulli Distribution Revisited

- Thomas Bayes considered a ball landing uniformly across a table.
- And another ball landing on the left or right (Bayes, 1763, page 385).
- The position of the first ball gives the parameter π.



Bernoulli Distribution Revisited

- Thomas Bayes considered a ball landing uniformly across a table.
- And another ball landing on the left or right (Bayes, 1763, page 385).
- The position of the first ball gives the parameter π.
- That 'parameter' is *itself* a random variable.

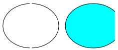


Bernoulli Distribution Revisited

- Thomas Bayes considered a ball landing uniformly across a table.
- And another ball landing on the left or right (Bayes, 1763, page 385).
- The position of the first ball gives the parameter π.
- That 'parameter' is *itself* a random variable.
- This treatment of a parameter, π, as a random variable that was/is considered controversial.

Bernoulli Distribution Revisited

- Thomas Bayes considered a ball landing uniformly across a table.
- And another ball landing on the left or right (Bayes, 1763, page 385).
- The position of the first ball gives the parameter π.
- That 'parameter' is *itself* a random variable.
- This treatment of a parameter, π, as a random variable that was/is considered controversial.

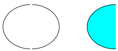


Bernoulli Distribution Revisited

- Thomas Bayes considered a ball landing uniformly across a table.
- And another ball landing on the left or right (Bayes, 1763, page 385).
- The position of the first ball gives the parameter π.
- That 'parameter' is *itself* a random variable.
- This treatment of a parameter, π, as a random variable that was/is considered controversial.

Bernoulli Distribution Revisited

- Thomas Bayes considered a ball landing uniformly across a table.
- And another ball landing on the left or right (Bayes, 1763, page 385).
- The position of the first ball gives the parameter π.
- That 'parameter' is *itself* a random variable.
- This treatment of a parameter, π, as a random variable that was/is considered controversial.



Bayesian Controversy

- Bayesian controversy relates to treating *epistemic* uncertainty as *aleatoric* uncertainty.
- Another analogy:
  - Before a football match the uncertainty about the result is *aleatoric*.
  - If I watch a recorded match *without* knowing the result the uncertainty is *epistemic*.

Simple Bayesian Inference

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}}$$

- Four components:
  1. Prior distribution: represents belief about parameter values before seeing data.
  2. Likelihood: gives relation between parameters and data.
  3. Posterior distribution: represents updated belief about parameters after data is observed.
  4. Marginal likelihood: represents assessment of the quality of the model. Can be compared with other models (likelihood/prior combinations). Ratios of marginal likelihoods are known as Bayes factors.

Outline

Prior Distribution

- Bayesian inference requires a prior on the parameters.
- The prior represents your belief *before* you see the data of the likely value of the parameters.
- For linear regression, consider a Gaussian prior on the intercept:

$$c \sim \mathcal{N}(0, \alpha_1)$$

Posterior Distribution

- Posterior distribution is found by combining the prior with the likelihood.
- Posterior distribution is your belief *after* you see the data of the likely value of the parameters.
- The posterior is found through **Bayes' Rule**

$$p(c|y) = \frac{p(y|c)p(c)}{p(y)}$$

Bayes Update

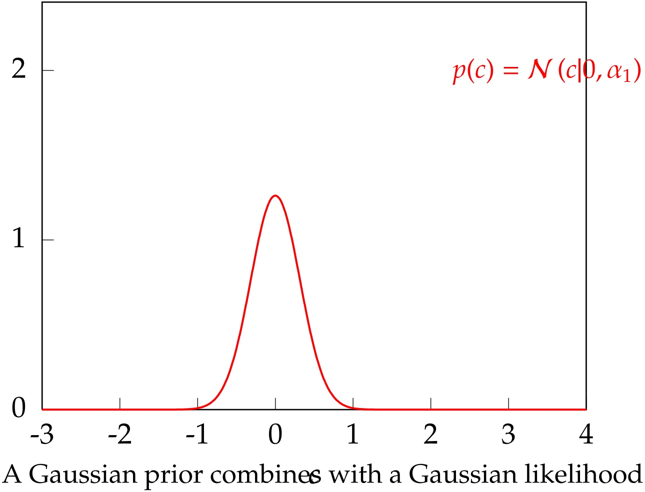


Figure: A Gaussian prior combines with a Gaussian likelihood for a Gaussian posterior.

Bayes Update

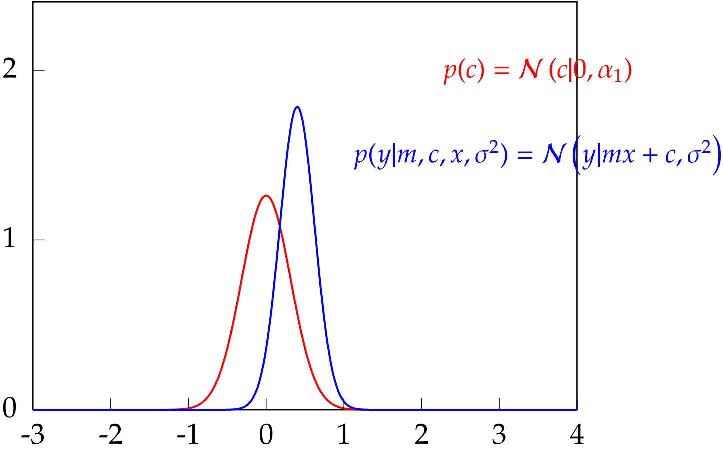


Figure: A Gaussian prior combines with a Gaussian likelihood for a Gaussian posterior.

Bayes Update

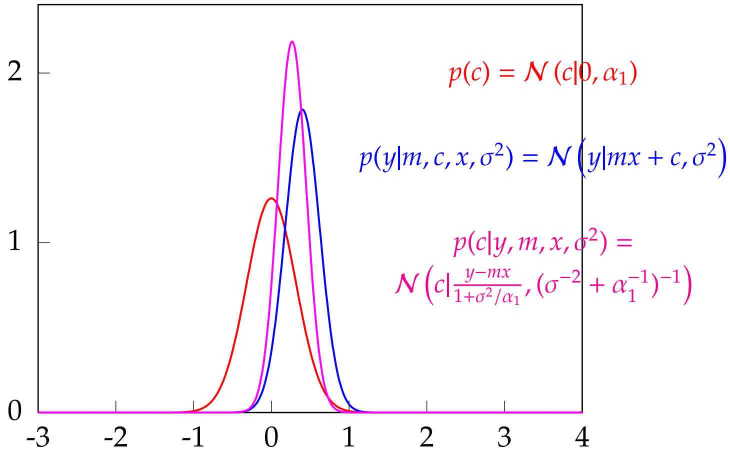


Figure: A Gaussian prior combines with a Gaussian likelihood for a Gaussian posterior.

Stages to Derivation of the Posterior

- Multiply likelihood by prior
  - they are "exponentiated quadratics", the answer is always also an exponentiated quadratic because $\exp(a^2)\exp(b^2) = \exp(a^2 + b^2)$.
- Complete the square to get the resulting density in the form of a Gaussian.
- Recognise the mean and (co)variance of the Gaussian. This is the estimate of the posterior.

Main Trick

$$p(c) = \frac{1}{\sqrt{2\pi\alpha_1}} \exp\left(-\frac{1}{2\alpha_1}c^2\right)$$

$$p(\mathbf{y}|\mathbf{x}, c, m, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - mx_i - c)^2\right)$$

Main Trick

$$p(c) = \frac{1}{\sqrt{2\pi\alpha_1}} \exp\left(-\frac{1}{2\alpha_1}c^2\right)$$

$$p(\mathbf{y}|\mathbf{x}, c, m, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - mx_i - c)^2\right)$$

$$p(c|\mathbf{y}, \mathbf{x}, m, \sigma^2) = \frac{p(\mathbf{y}|\mathbf{x}, c, m, \sigma^2)p(c)}{p(\mathbf{y}|\mathbf{x}, m, \sigma^2)}$$

Main Trick

$$p(c) = \frac{1}{\sqrt{2\pi\alpha_1}} \exp\left(-\frac{1}{2\alpha_1}c^2\right)$$

$$p(\mathbf{y}|\mathbf{x}, c, m, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - mx_i - c)^2\right)$$

$$p(c|\mathbf{y}, \mathbf{x}, m, \sigma^2) = \frac{p(\mathbf{y}|\mathbf{x}, c, m, \sigma^2)p(c)}{\int p(\mathbf{y}|\mathbf{x}, c, m, \sigma^2)p(c)\mathrm{d}c}$$

Main Trick

$$p(c) = \frac{1}{\sqrt{2\pi\alpha_1}} \exp\left(-\frac{1}{2\alpha_1}c^2\right)$$

$$p(\mathbf{y}|\mathbf{x}, c, m, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - mx_i - c)^2\right)$$

$$p(c|\mathbf{y}, \mathbf{x}, m, \sigma^2) \propto p(\mathbf{y}|\mathbf{x}, c, m, \sigma^2)p(c)$$

$$\begin{aligned}\log p(c|\mathbf{y},\mathbf{x},m,\sigma^2) &= -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - c - mx_i)^2 - \frac{1}{2\alpha_1}c^2 + \text{const} \\ &= -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - mx_i)^2 - \left(\frac{N}{2\sigma^2} + \frac{1}{2\alpha_1}\right)c^2 \\ &\quad + c\frac{\sum_{i=1}^{N}(y_i - mx_i)}{\sigma^2},\end{aligned}$$

complete the square of the quadratic form to obtain

$$\log p(c|\mathbf{y},\mathbf{x},m,\sigma^2) = -\frac{1}{2\tau^2}(c-\mu)^2 + \text{const},$$

where $\tau^2 = \left(N\sigma^{-2} + \alpha_1^{-1}\right)^{-1}$ and $\mu = \frac{\tau^2}{\sigma^2}\sum_{n=1}^{N}(y_i - mx_i)$.

The Joint Density

- Really want to know the *joint* posterior density over the parameters *c and m*.
- Could now integrate out over m, but it's easier to consider the multivariate case.

Aleatoric Uncertainty

- This is uncertainty we couldn't know even if we wanted to. e.g. the result of a football match before it's played.
- Where a sheet of paper might land on the floor.

Epistemic Uncertainty

- This is uncertainty we could in principal know the answer too. We just haven't observed enough yet, e.g. the result of a football match *after* it's played.
- What colour socks your lecturer is wearing.

Reading

- Bishop Section 1.2.3 (pg 21–24).
- Bishop Section 1.2.6 (start from just past eq 1.64 pg 30-32).
- Rogers and Girolami use an example of a coin toss for introducing Bayesian inference Chapter 3, Sections 3.1-3.4 (pg 95-117). Although you also need the beta density which we haven't yet discussed. This is also the example that Laplace used.

Reading Summary

- Basis Functions
  - Section 1.4 of Rogers and Girolami.
  - Chapter 1, pg 1-6 of Bishop.
  - Chapter 3, Section 3.1 of Bishop up to pg 143.
- Generalization
  - Section 1.5 of Rogers and Girolami.
- Bayesian Inference
  - Rogers and Girolami use an example of a coin toss for introducing Bayesian inference Chapter 3, Sections 3.1-3.4 (pg 95-117). Although you also need the beta density which we haven't yet discussed. This is also the example that Laplace used.
  - Bishop Section 1.2.3 (pg 21–24).
  - Bishop Section 1.2.6 (start from just past eq 1.64 pg 30-32).

References I

T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53:370–418, 1763. [DOI].

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006. [Google Books] .

P. S. Laplace. Mémoire sur la probabilité des causes par les évènemens. In *Mémoires de mathèmatique et de physique, presentés à lAcadémie Royale des Sciences, par divers savans, & lù dans ses assemblées 6*, pages 621–656, 1774. Translated in Stigler (1986).

S. Rogers and M. Girolami. *A First Course in Machine Learning*. CRC Press, 2011. [Google Books] .

S. M. Stigler. Laplace's 1774 memoir on inverse probability. *Statistical Science*, 1:359–378, 1986.