Basis Functions

Neil D. Lawrence

GPRS
7th August 2013

Outline

- 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:

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

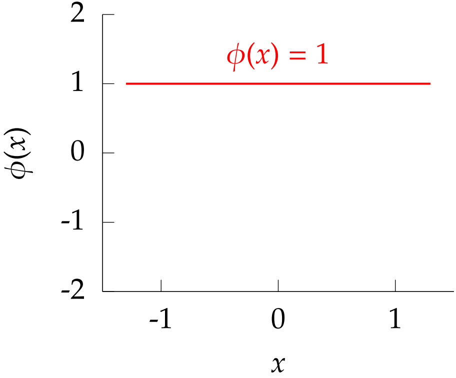


Figure: A quadratic basis.

Quadratic Basis

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

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

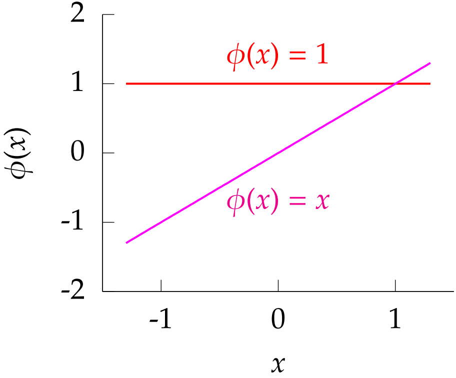


Figure: A quadratic basis.

Quadratic Basis

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

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

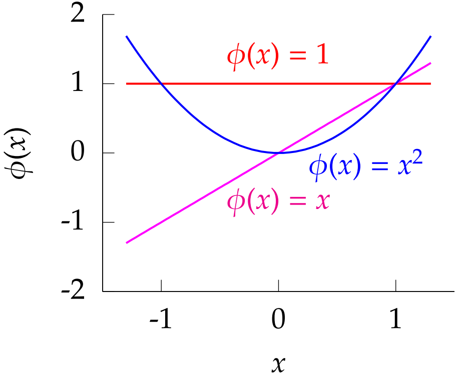


Figure: A quadratic basis.

Functions Derived from Quadratic Basis

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

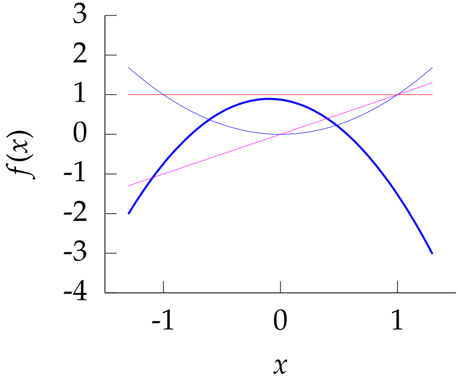


Figure: Function from quadratic basis with weights $w_1 = 0.87466$, $w_2 = -0.38835$, $w_3 = -2.0058$.

Functions Derived from Quadratic Basis

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

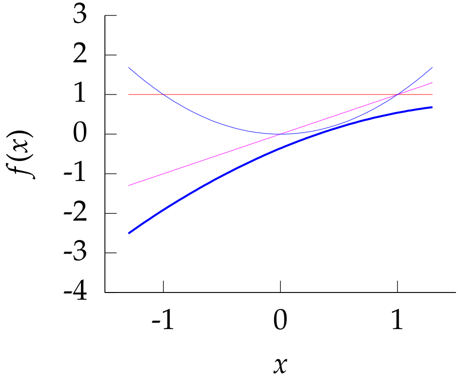


Figure: Function from quadratic basis with weights $w_1 = -0.35908$, $w_2 = 1.2274$, $w_3 = -0.32825$.

Functions Derived from Quadratic Basis

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

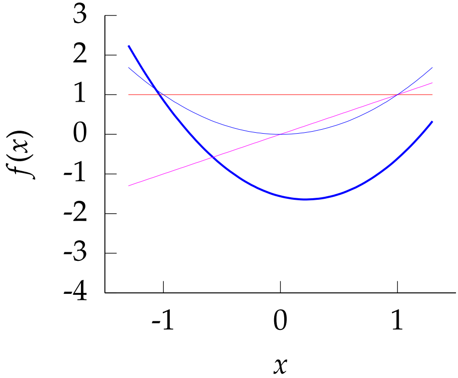


Figure: Function from quadratic basis with weights $w_1 = -1.5638$, $w_2 = -0.73577$, $w_3 = 1.6861$.

Radial Basis Functions

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

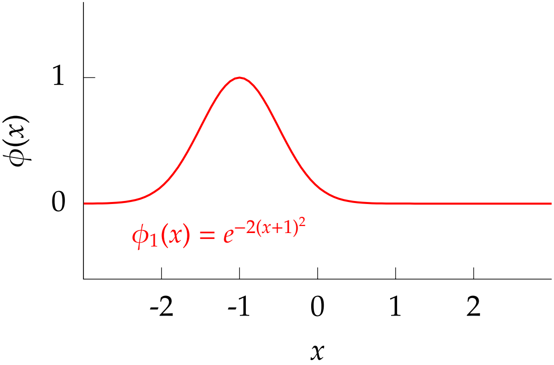


Figure: Radial basis functions.

Radial Basis Functions

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

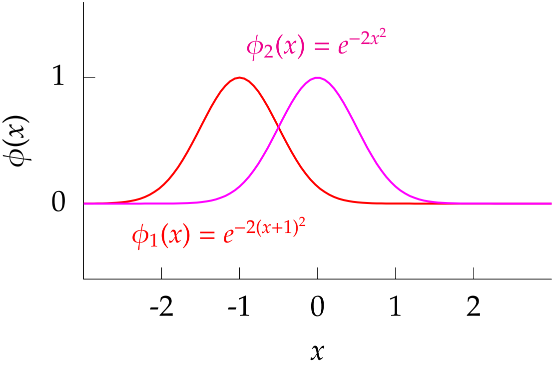


Figure: Radial basis functions.

Radial Basis Functions

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

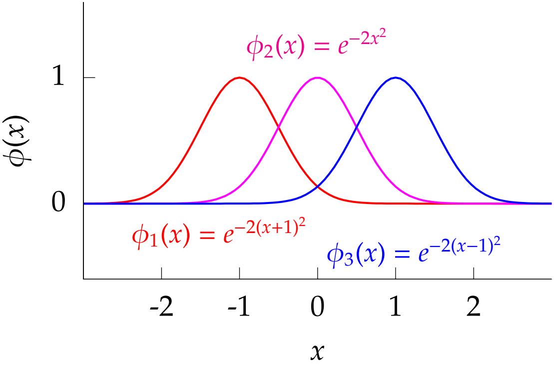


Figure: Radial basis functions.

Functions Derived from Radial Basis

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

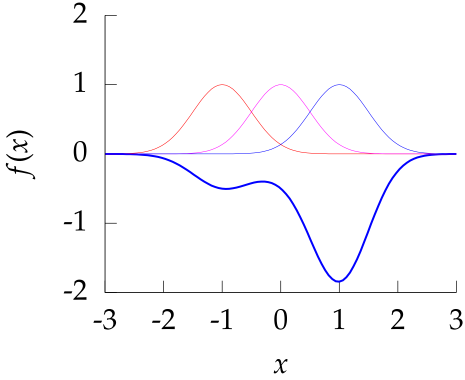


Figure: Function from radial basis with weights $w_1 = -0.47518$, $w_2 = -0.18924$, $w_3 = -1.8183$.

Functions Derived from Radial Basis

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

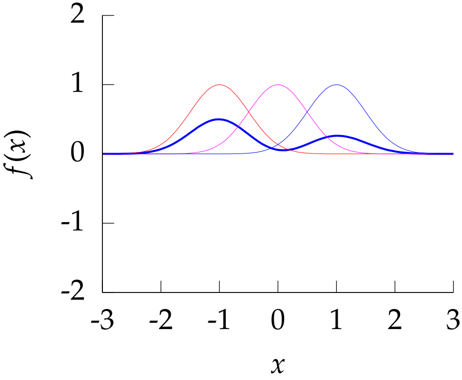


Figure: Function from radial basis with weights $w_1 = 0.50596$, $w_2 = -0.046315$, $w_3 = 0.26813$.

Functions Derived from Radial Basis

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

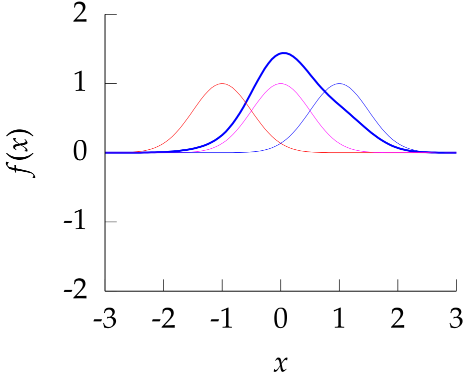


Figure: Function from radial basis with weights $w_1 = 0.07179$, $w_2 = 1.3591$, $w_3 = 0.50604$.

Reading

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

Multi-dimensional Inputs

- Multivariate functions involve more than one input.
- Height might be a function of weight and gender.
- There could be other contributory factors.
- Place these factors in a feature vector $\mathbf{x}_i$.
- Linear function is now defined as

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

Vector Notation

mo

- Write in vector notation,

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

- Can absorb c into $\mathbf{w}$ by assuming extra input x_0 which is always 1.

$$f(\mathbf{x}_i) = \mathbf{w}^\top \mathbf{x}_i$$

Log Likelihood for Multivariate Regression

- The likelihood of a single data point is

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

- Leading to a log likelihood for the data set of

$$L(\mathbf{w}, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi - \frac{\sum_{i=1}^{n} \left(y_i - \mathbf{w}^\top \mathbf{x}_i\right)^2}{2\sigma^2}.$$

- And a corresponding error function of

$$E(\mathbf{w}, \sigma^2) = \frac{n}{2} \log \sigma^2 + \frac{\sum_{i=1}^{n} \left(y_i - \mathbf{w}^\top \mathbf{x}_i\right)^2}{2\sigma^2}.$$

Expand the Brackets

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

Multivariate Derivatives

- We will need some multivariate calculus.
- For now some simple multivariate differentiation:

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

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 $\mathbf{w}$ we obtain

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

Leading to

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

Rewrite in matrix notation:

$$\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top = \mathbf{X}^\top \mathbf{X}$$

$$\sum_{i=1}^{n} \mathbf{x}_i y_i = \mathbf{X}^\top \mathbf{y}$$

Update Equations

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

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

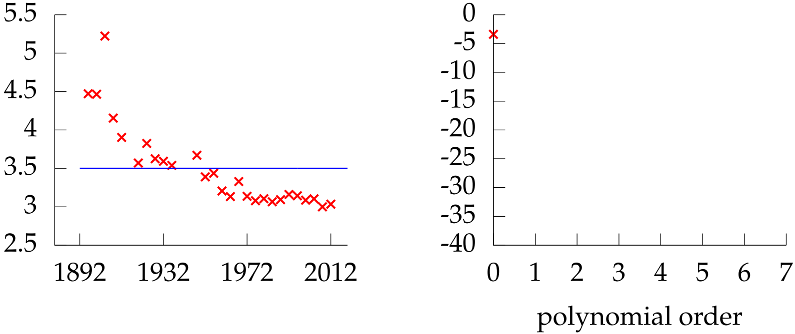
- The equation for ${\sigma^2}^*$ may also be found

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

Reading

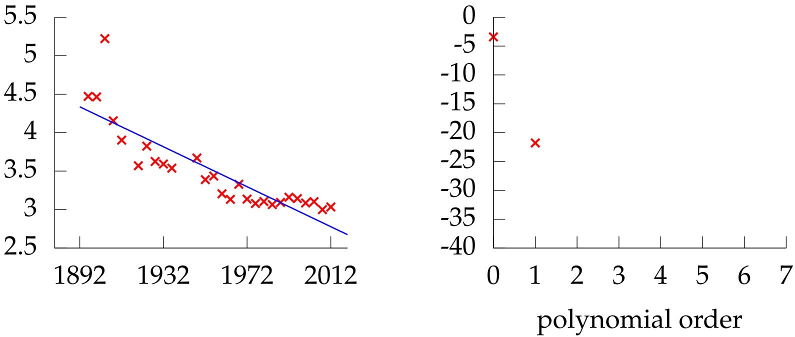
- Section 1.3 of Rogers and Girolami for Matrix & Vector Review.

Polynomial Fits to Olympics Data



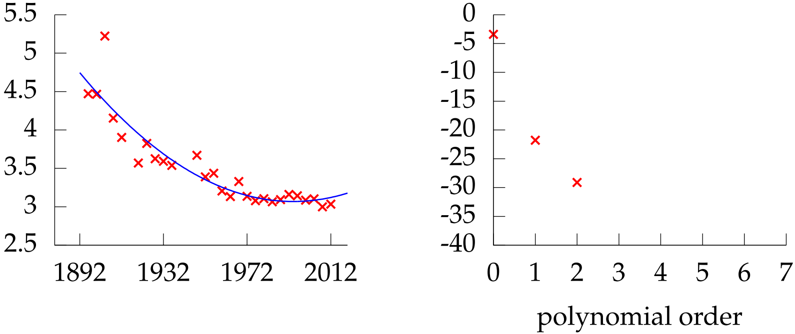
Left: fit to data, *Right*: model error. Polynomial order 0, model error -3.3989, $\sigma^2 = 0.286$, $\sigma = 0.535$.

Polynomial Fits to Olympics Data



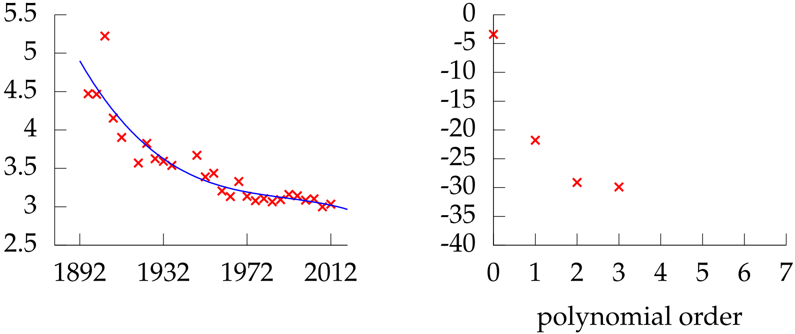
Left: fit to data, *Right*: model error. Polynomial order 1, model error -21.772, $\sigma^2 = 0.0733$, $\sigma = 0.271$.

Polynomial Fits to Olympics Data



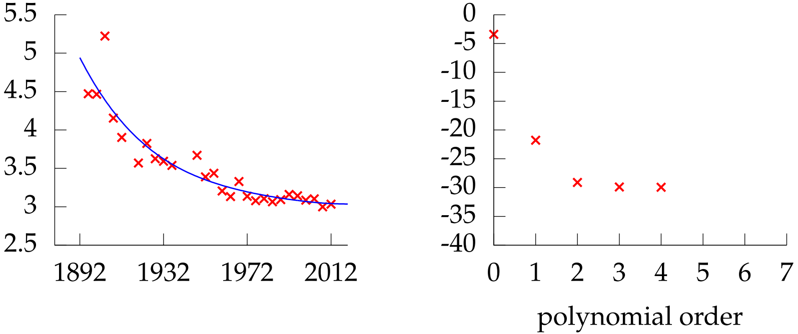
Left: fit to data, *Right*: model error. Polynomial order 2, model error -29.101, $\sigma^2 = 0.0426$, $\sigma = 0.206$.

Polynomial Fits to Olympics Data



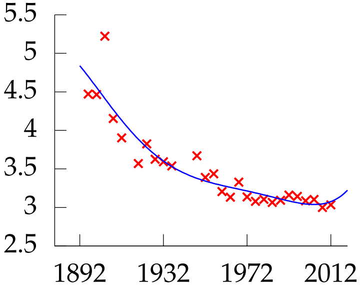
Left: fit to data, *Right*: model error. Polynomial order 3, model error -29.907, $\sigma^2 = 0.0401$, $\sigma = 0.200$.

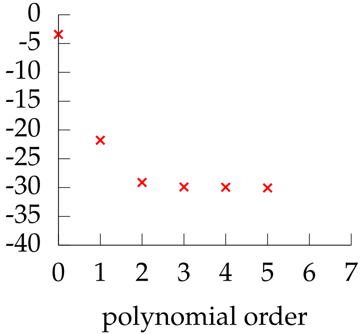
Polynomial Fits to Olympics Data



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

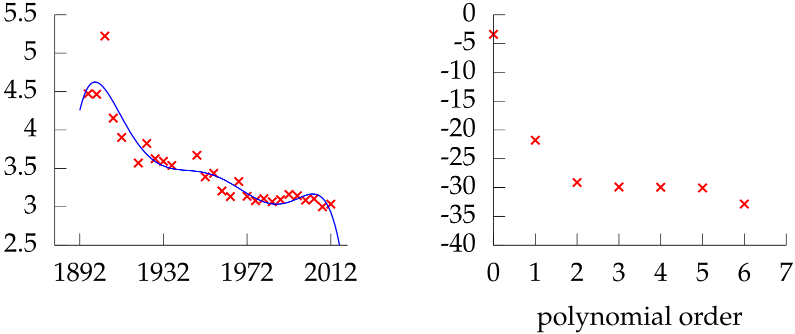
Polynomial Fits to Olympics Data





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, $\sigma^2 = 0.0322$, $\sigma = 0.180$.

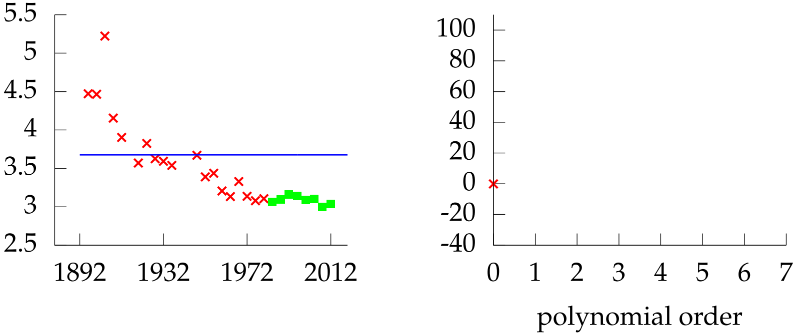
Overfitting

- 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

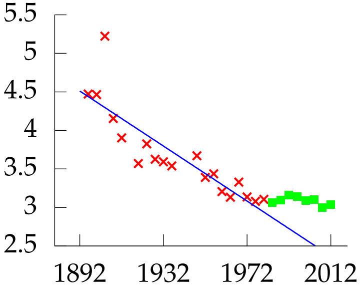
- We call the data used for fitting the model the ‘training set’.
- Data not used for training, but when the model is applied ‘in the field’ is called the ‘test data’.
- Challenge for generalization is to ensure a good performance on test data given only training data.

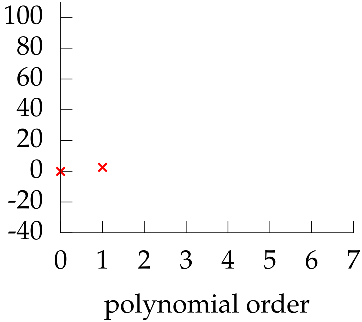
Validation Set



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

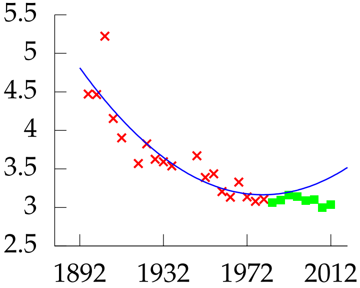
Validation Set

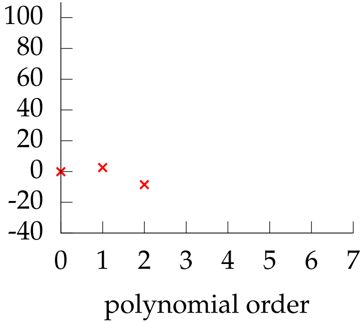




Left: fit to data, *Right*: model error. Polynomial order 1, training error -15.325, validation error 2.5863, $\sigma^2 = 0.0733$, $\sigma = 0.271$.

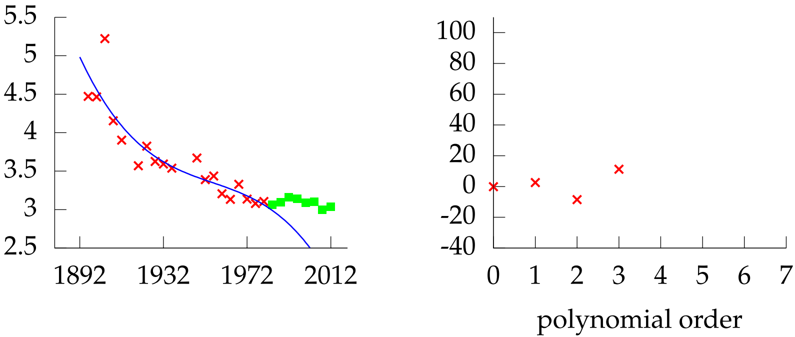
Validation Set





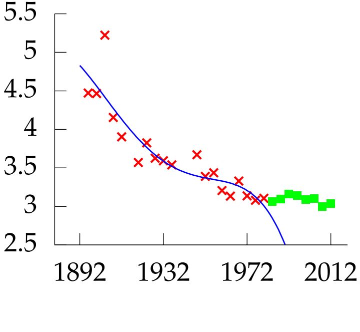
Left: fit to data, *Right*: model error. Polynomial order 2, training error -17.579, validation error -8.4831, $\sigma^2 = 0.0578$, $\sigma = 0.240$.

Validation Set



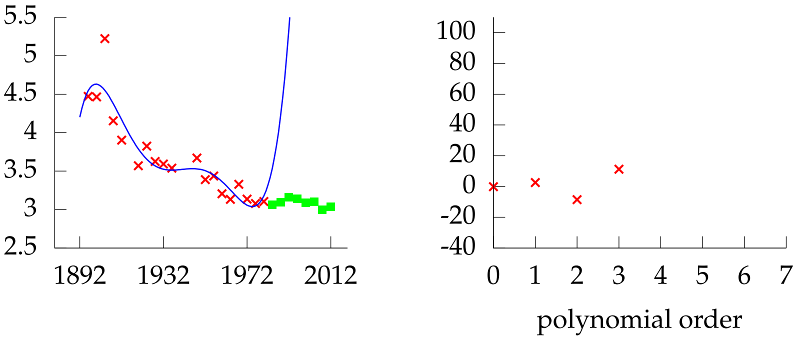
Left: fit to data, *Right*: model error. Polynomial order 3, training error -18.064, validation error 11.27, $\sigma^2 = 0.0549$, $\sigma = 0.234$.

Validation Set



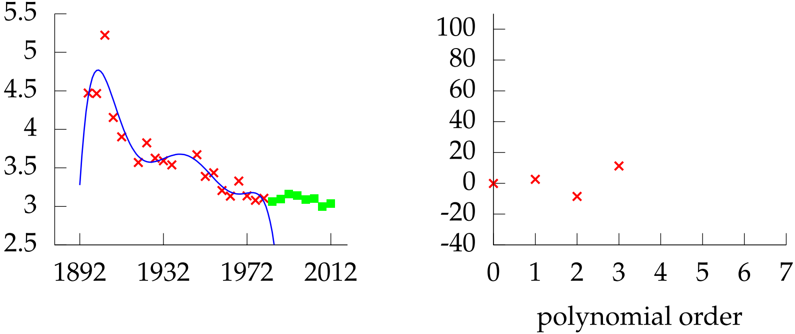
Left: fit to data, *Right*: model error. Polynomial order 4, training error -18.245, validation error 232.92, $\sigma^2 = 0.0539$, $\sigma = 0.232$.

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

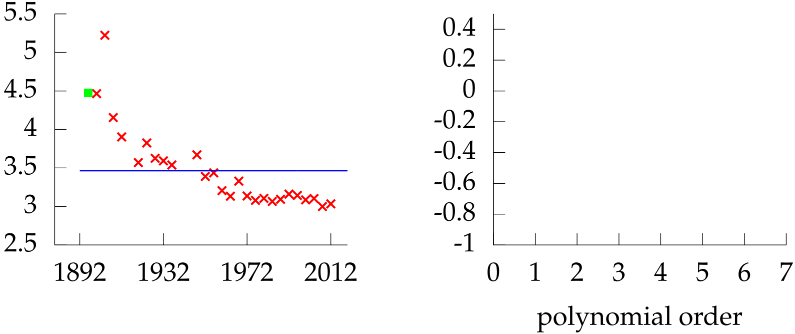


Left: fit to data, *Right*: model error. Polynomial order 6, training error -22.881, validation error 67775, $\sigma^2 = 0.0331$, $\sigma = 0.182$.

Leave One Out Error

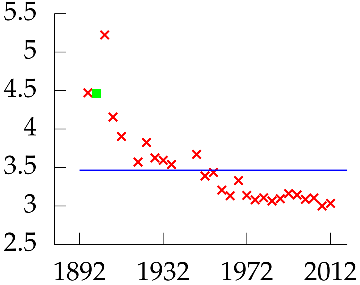
- Take training set and remove one point.
- Train on the remaining data.
- Compute the error on the point you removed (which wasn't in the training data).
- Do this for each point in the training set in turn.
- Average the resulting error. This is the leave one out error.

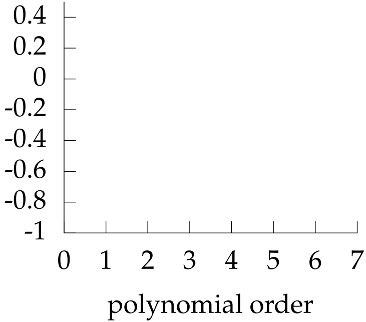
Leave One Out Error



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

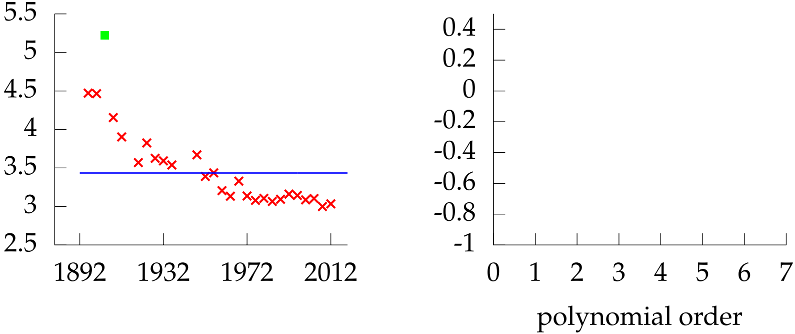
Leave One Out Error





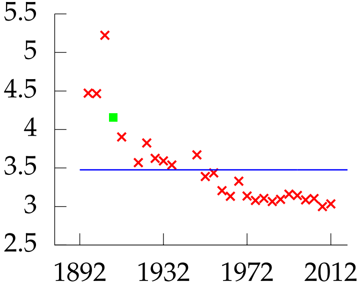
Polynomial order 0, training error -3.346, leave one out error 0.045811.

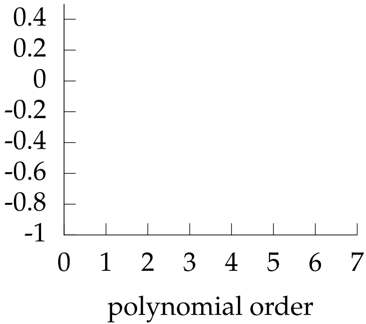
Leave One Out Error



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

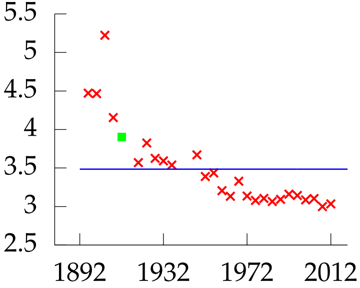
Leave One Out Error

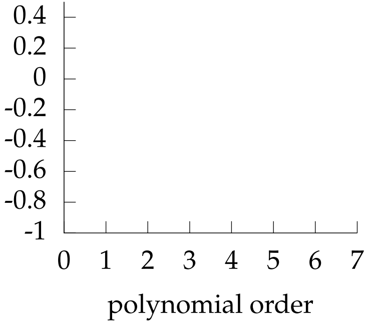




Polynomial order 0, training error -3.346, leave one out error 0.045811.

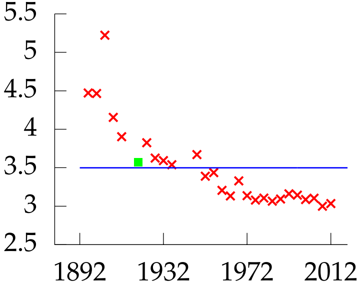
Leave One Out Error

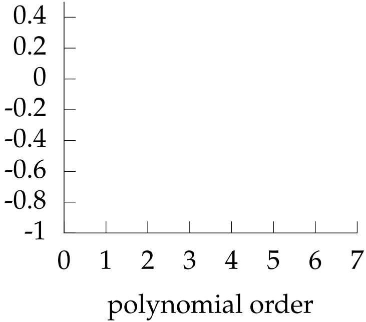




Polynomial order 0, training error -3.346, leave one out error 0.045811.

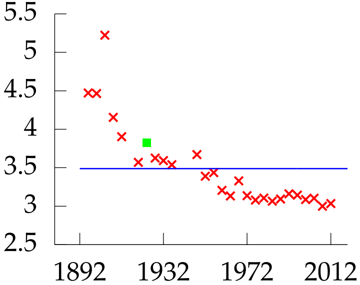
Leave One Out Error

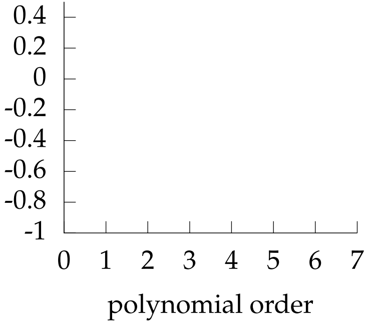




Polynomial order 0, training error -3.346, leave one out error 0.045811.

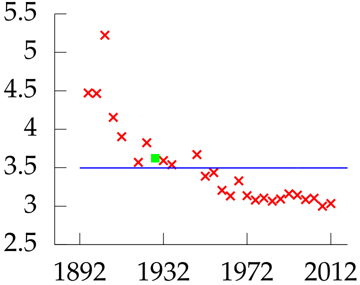
Leave One Out Error

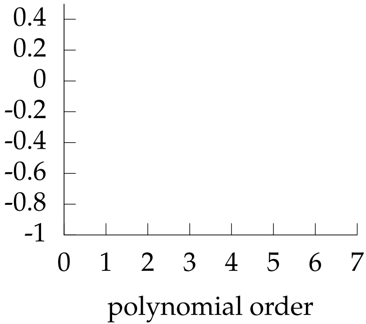




Polynomial order 0, training error -3.346, leave one out error 0.045811.

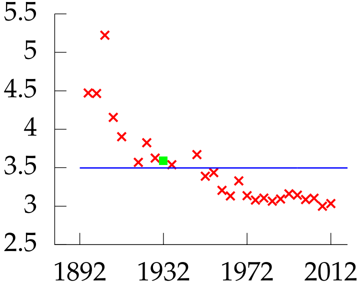
Leave One Out Error

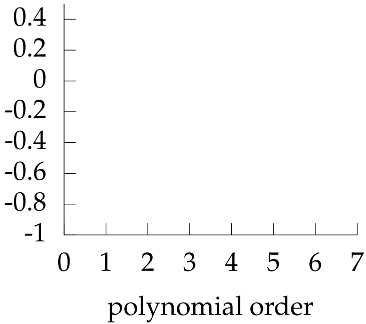




Polynomial order 0, training error -3.346, leave one out error 0.045811.

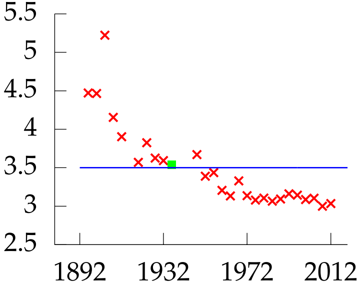
Leave One Out Error

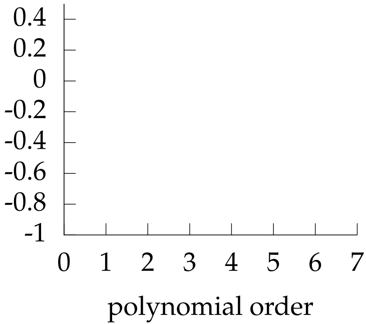




Polynomial order 0, training error -3.346, leave one out error 0.045811.

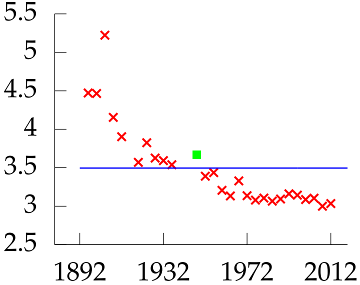
Leave One Out Error

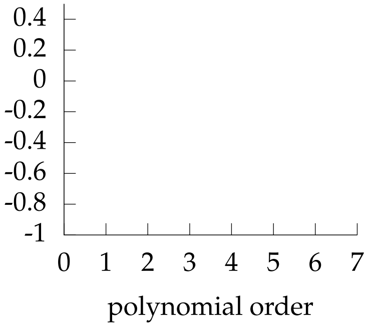




Polynomial order 0, training error -3.346, leave one out error 0.045811.

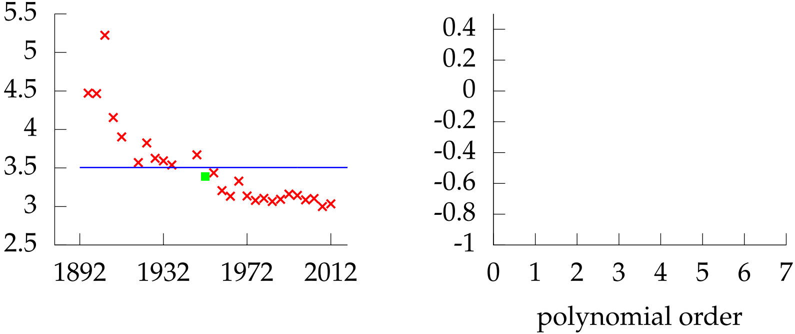
Leave One Out Error





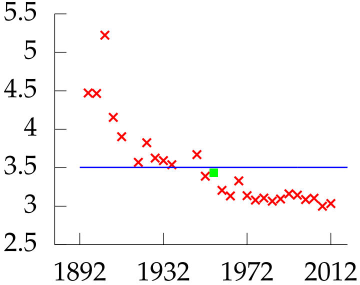
Polynomial order 0, training error -3.346, leave one out error 0.045811.

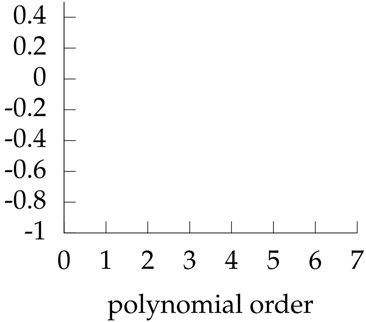
Leave One Out Error



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

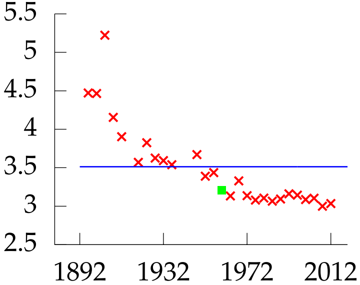
Leave One Out Error

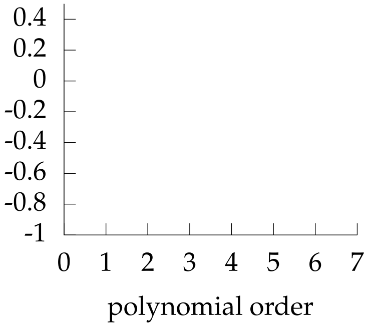




Polynomial order 0, training error -3.346, leave one out error 0.045811.

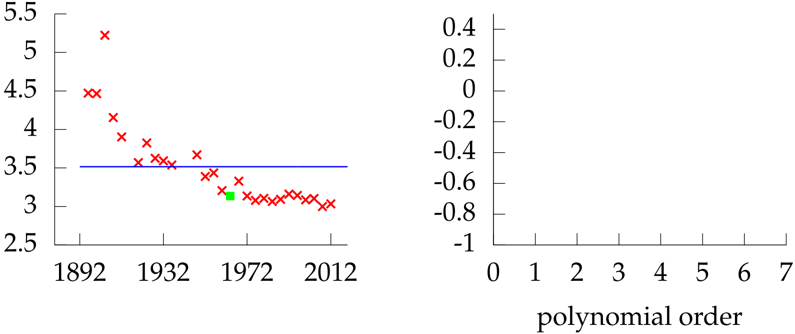
Leave One Out Error





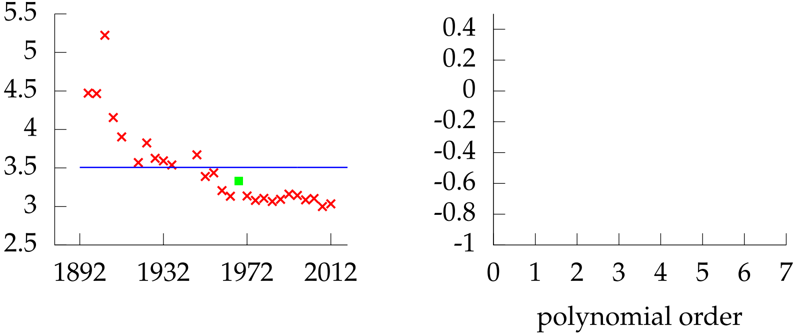
Polynomial order 0, training error -3.346, leave one out error 0.045811.

Leave One Out Error



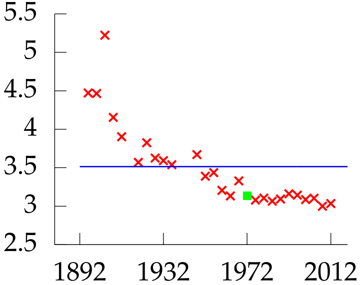
Polynomial order 0, training error -3.346, leave one out error 0.045811.

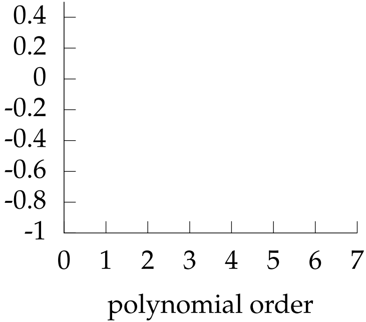
Leave One Out Error



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

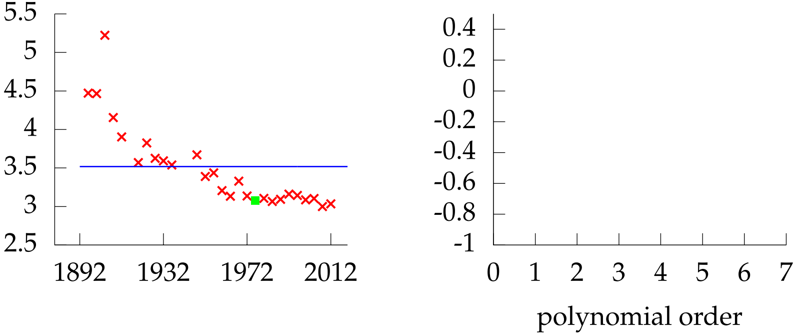
Leave One Out Error





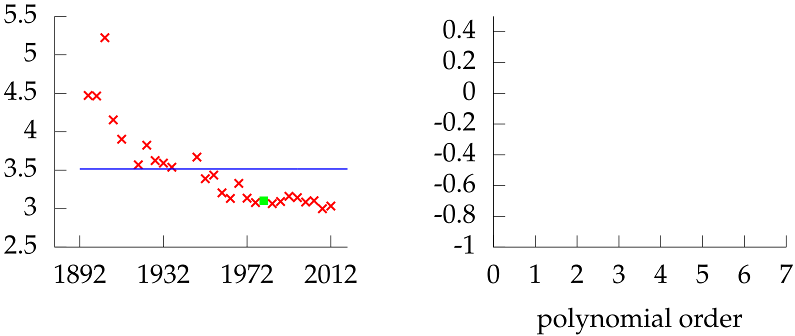
Polynomial order 0, training error -3.346, leave one out error 0.045811.

Leave One Out Error



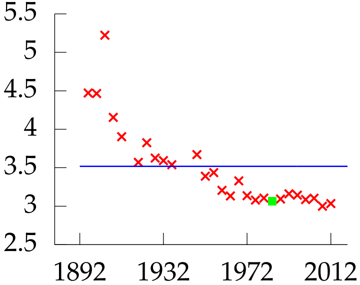
Polynomial order 0, training error -3.346, leave one out error 0.045811.

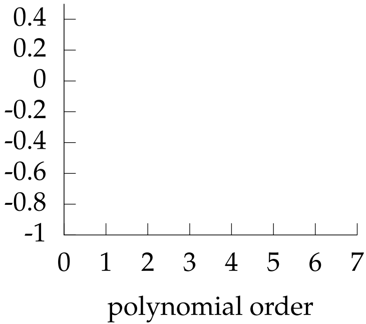
Leave One Out Error



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

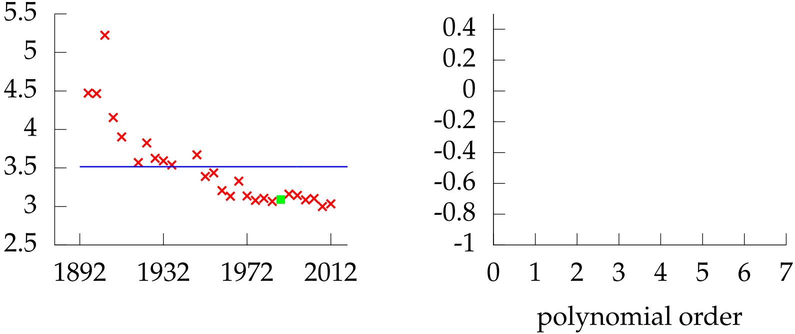
Leave One Out Error





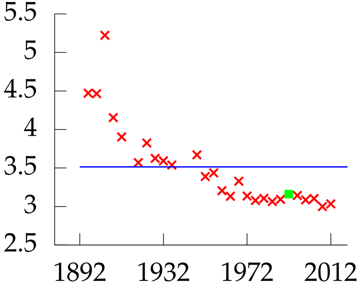
Polynomial order 0, training error -3.346, leave one out error 0.045811.

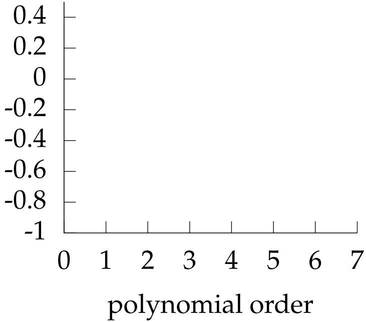
Leave One Out Error



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

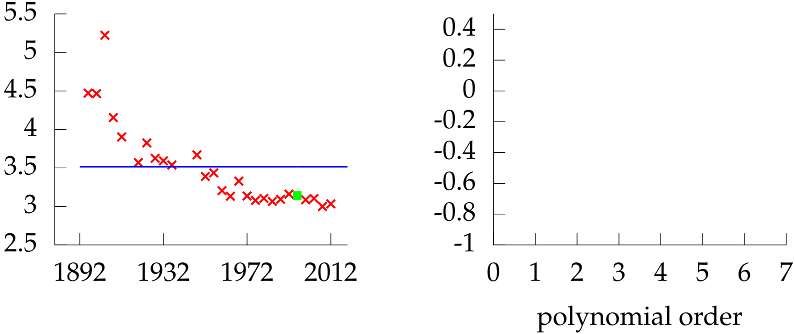
Leave One Out Error





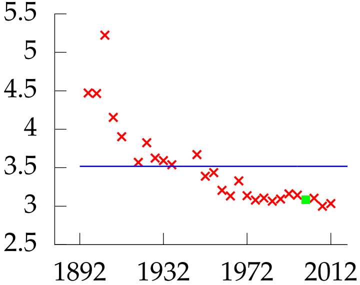
Polynomial order 0, training error -3.346, leave one out error 0.045811.

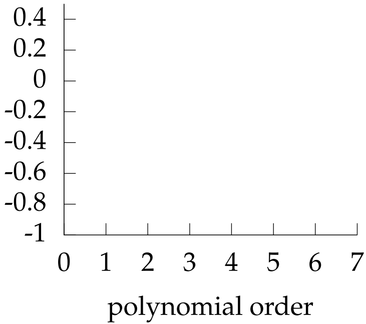
Leave One Out Error



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

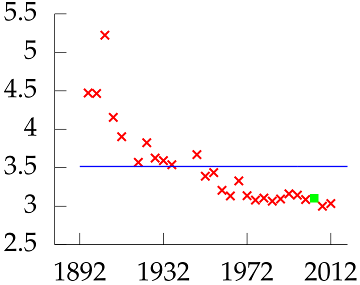
Leave One Out Error

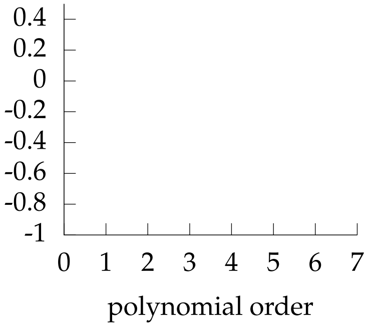




Polynomial order 0, training error -3.346, leave one out error 0.045811.

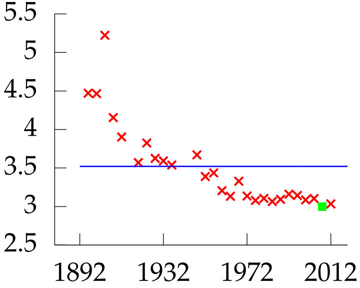
Leave One Out Error

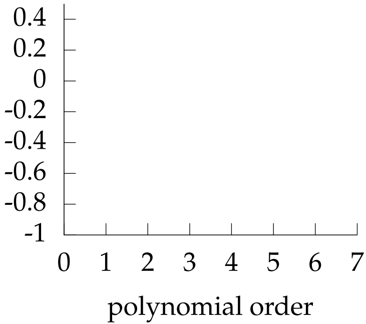




Polynomial order 0, training error -3.346, leave one out error 0.045811.

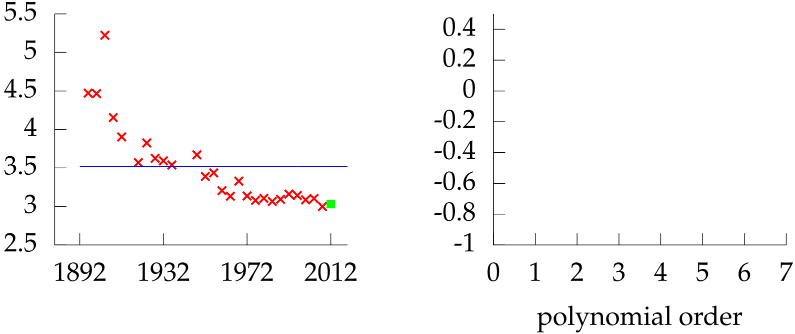
Leave One Out Error





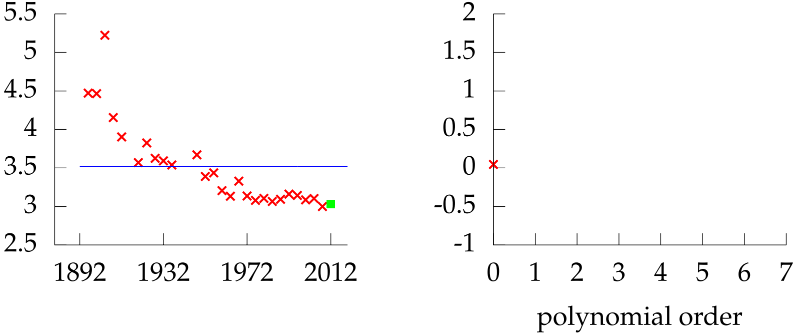
Polynomial order 0, training error -3.346, leave one out error 0.045811.

Leave One Out Error



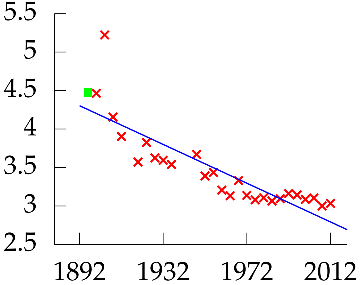
Polynomial order 0, training error -3.346, leave one out error 0.045811.

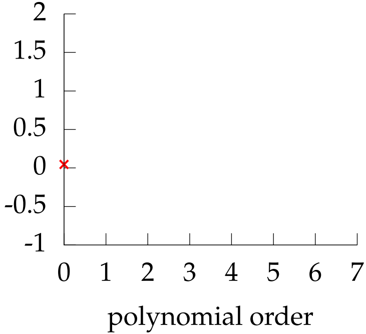
Leave One Out Error



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

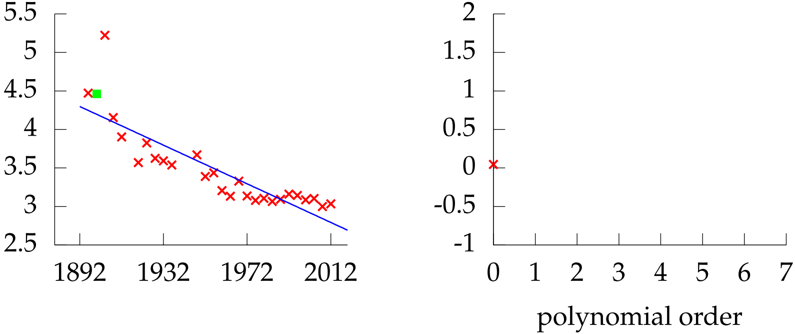
Leave One Out Error





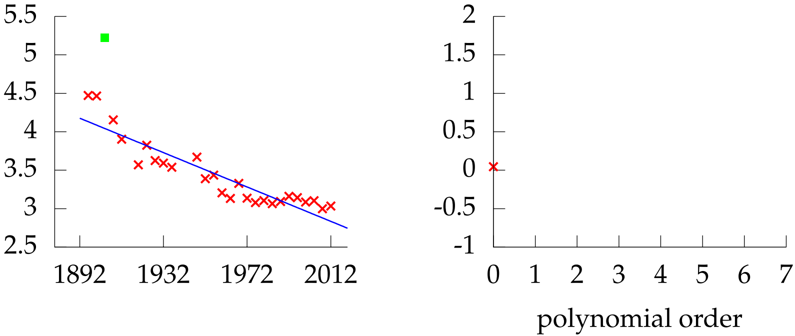
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



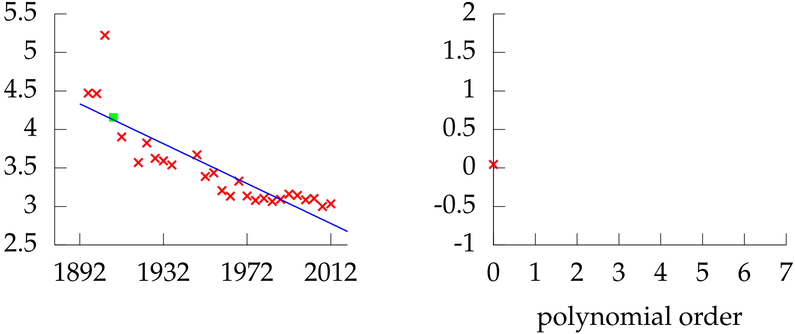
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



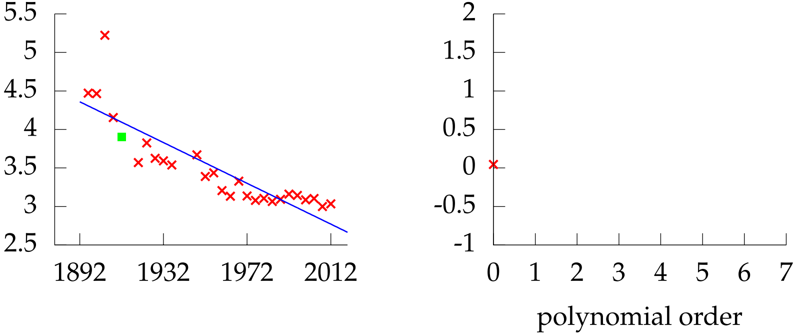
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



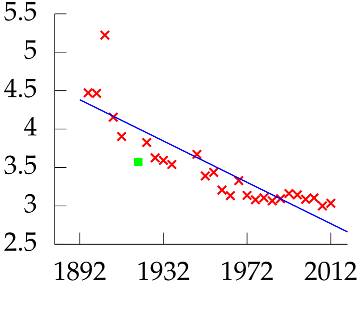
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



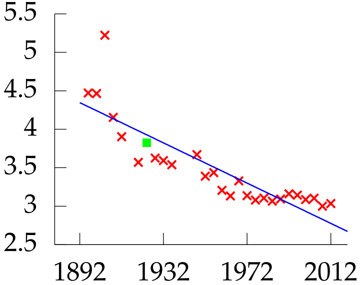
Polynomial order 1, training error -21.183, leave one out error -0.15413.

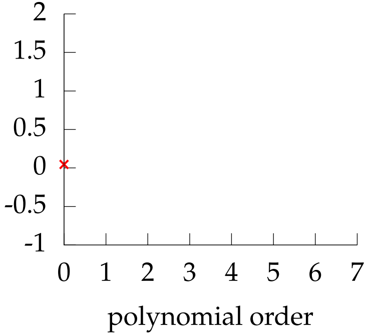
Leave One Out Error



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

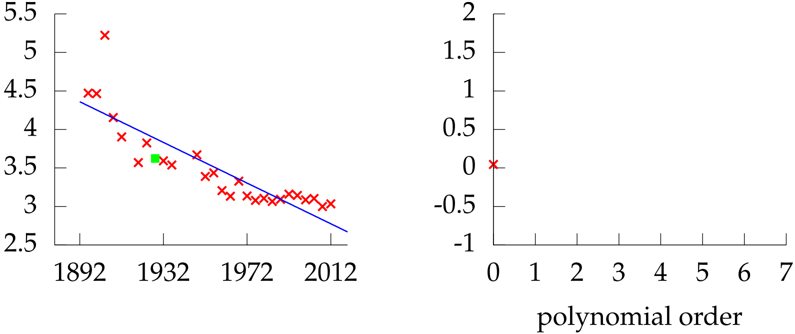
Leave One Out Error





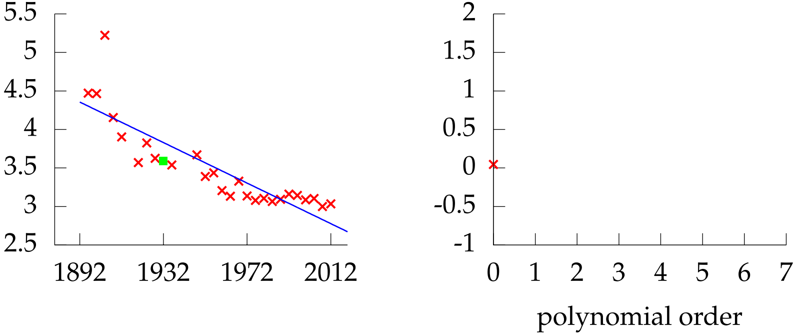
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



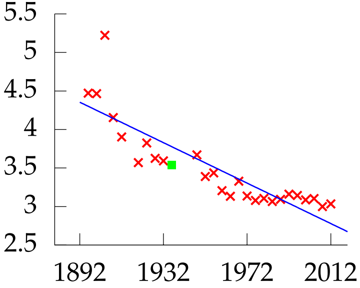
Polynomial order 1, training error -21.183, leave one out error -0.15413.

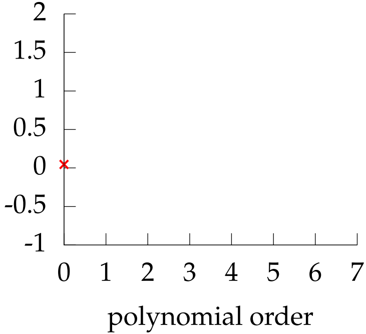
Leave One Out Error



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

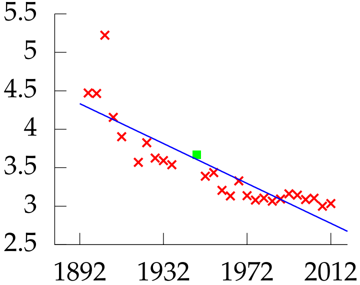
Leave One Out Error

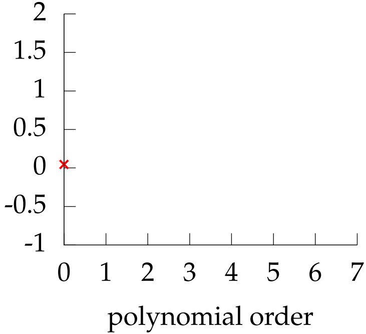




Polynomial order 1, training error -21.183, leave one out error -0.15413.

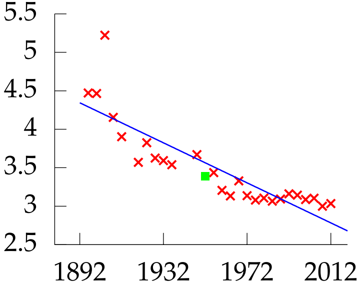
Leave One Out Error

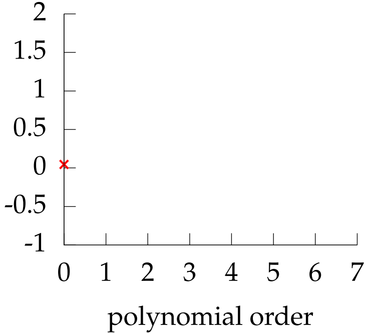




Polynomial order 1, training error -21.183, leave one out error -0.15413.

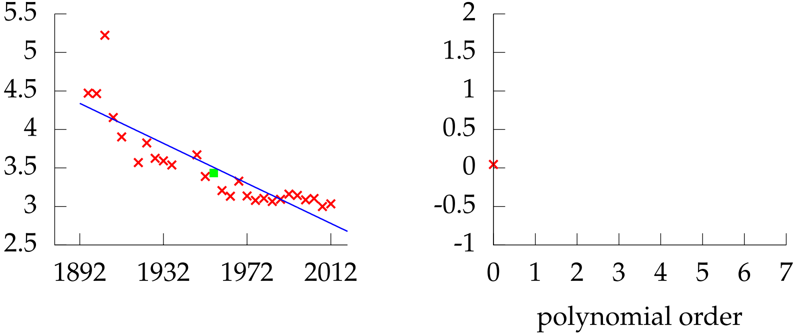
Leave One Out Error





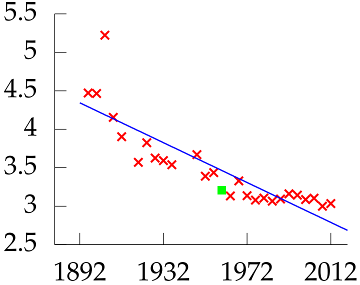
Polynomial order 1, training error -21.183, leave one out error -0.15413.

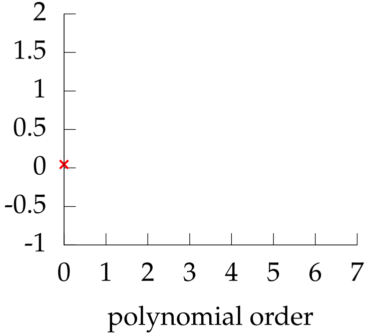
Leave One Out Error



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

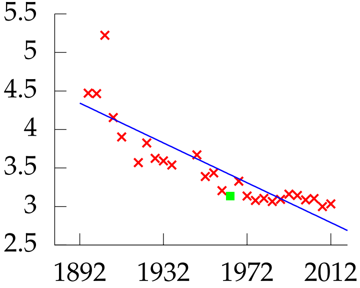
Leave One Out Error

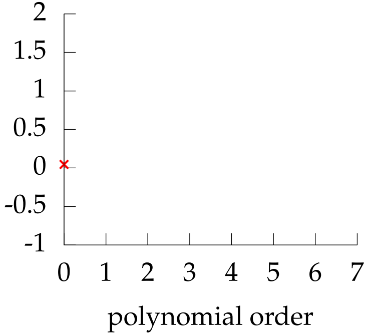




Polynomial order 1, training error -21.183, leave one out error -0.15413.

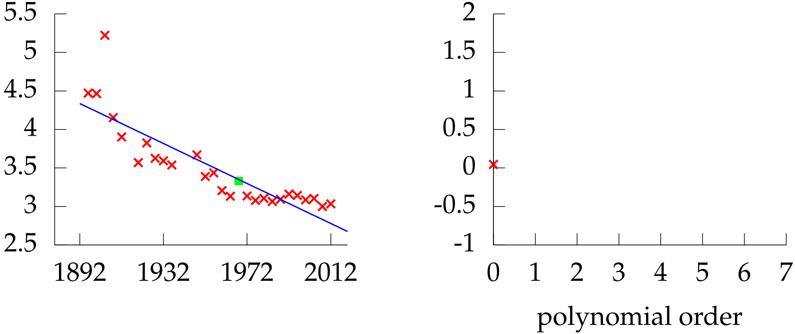
Leave One Out Error





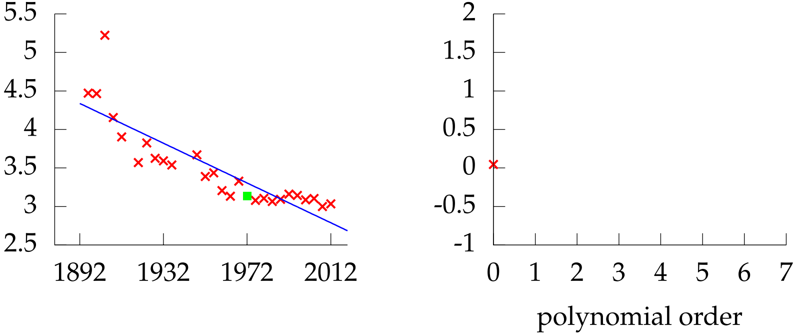
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



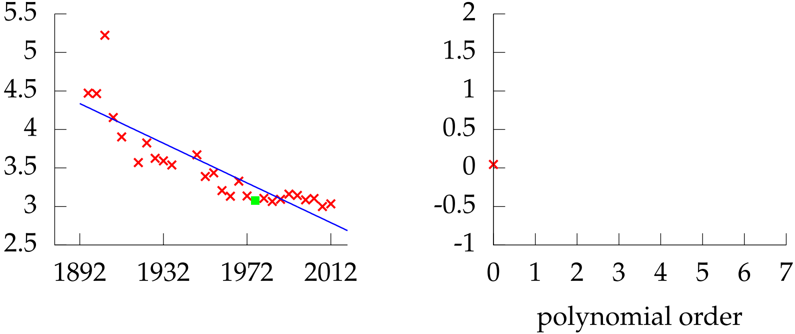
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



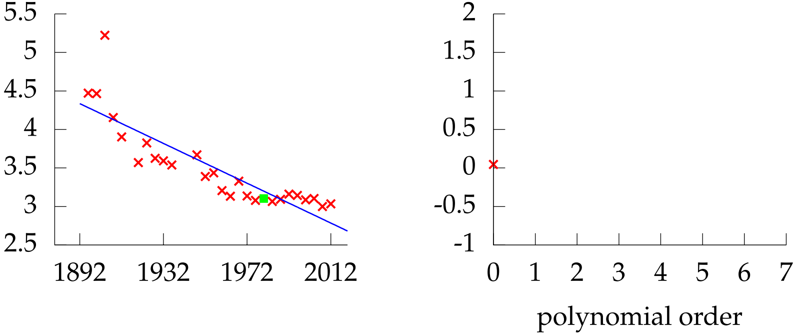
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



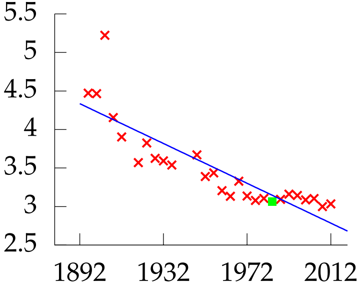
Polynomial order 1, training error -21.183, leave one out error -0.15413.

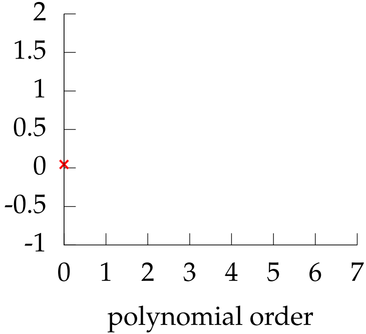
Leave One Out Error



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

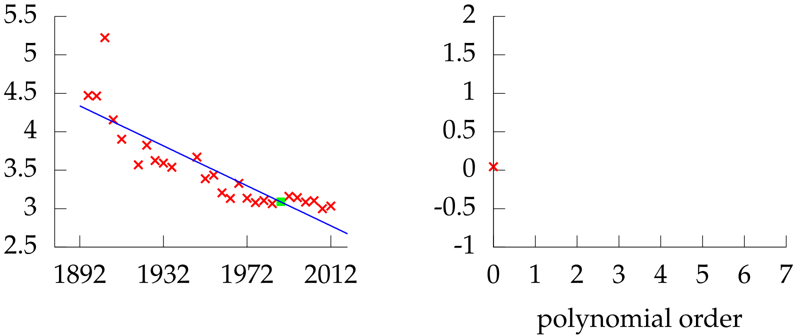
Leave One Out Error





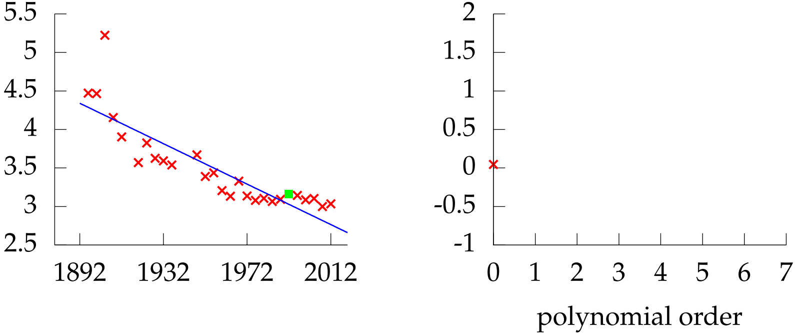
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



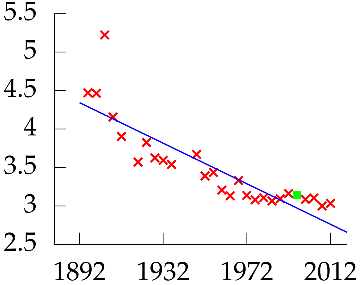
Polynomial order 1, training error -21.183, leave one out error -0.15413.

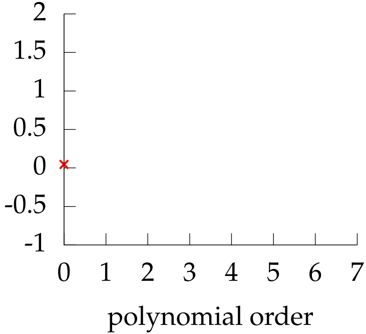
Leave One Out Error



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

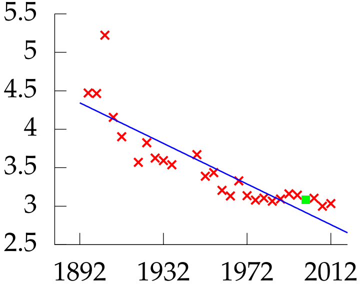
Leave One Out Error

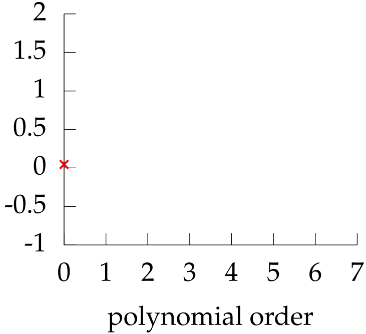




Polynomial order 1, training error -21.183, leave one out error -0.15413.

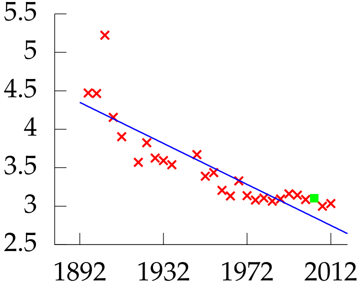
Leave One Out Error

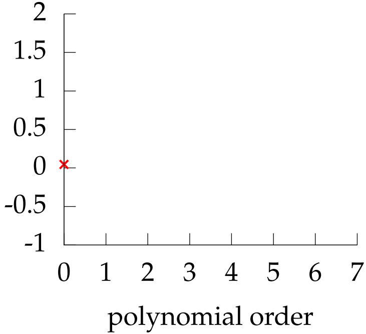




Polynomial order 1, training error -21.183, leave one out error -0.15413.

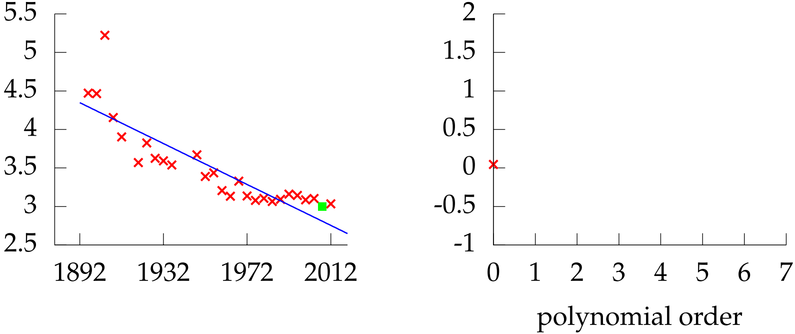
Leave One Out Error





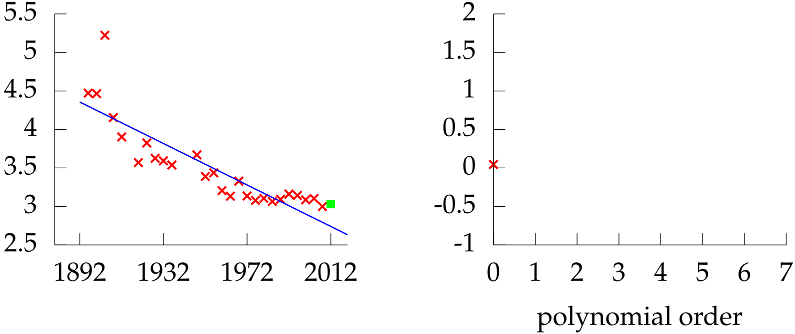
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



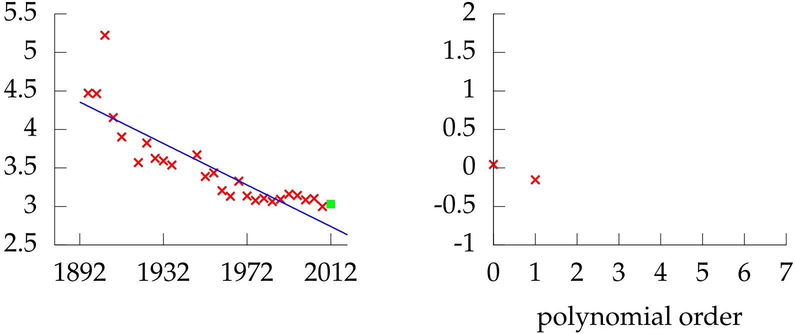
Polynomial order 1, training error -21.183, leave one out error -0.15413.

Leave One Out Error



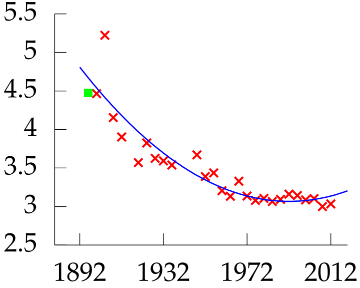
Polynomial order 1, training error -21.183, leave one out error -0.15413.

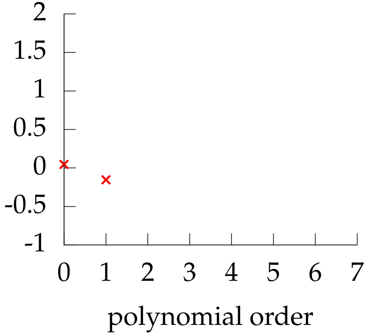
Leave One Out Error



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

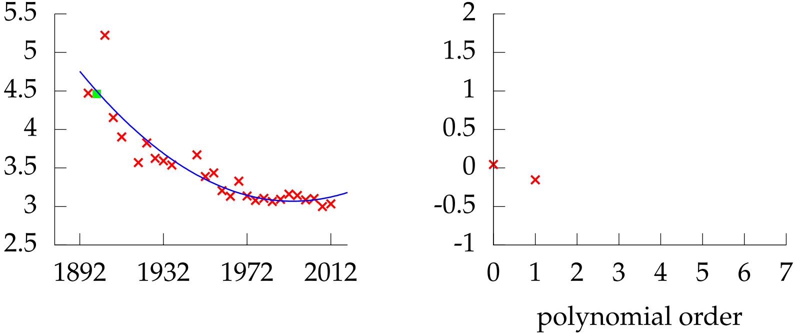
Leave One Out Error





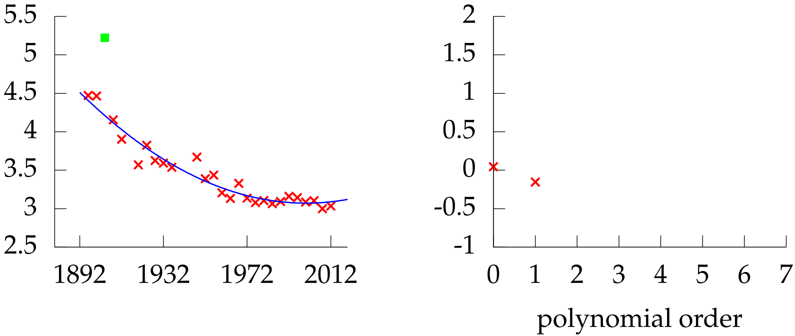
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



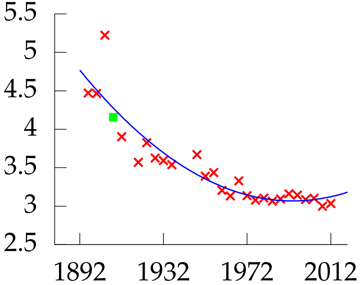
Polynomial order 2, training error -28.403, leave one out error 0.34669.

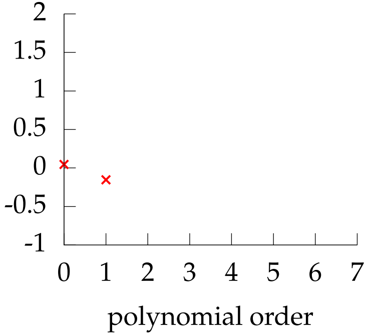
Leave One Out Error



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

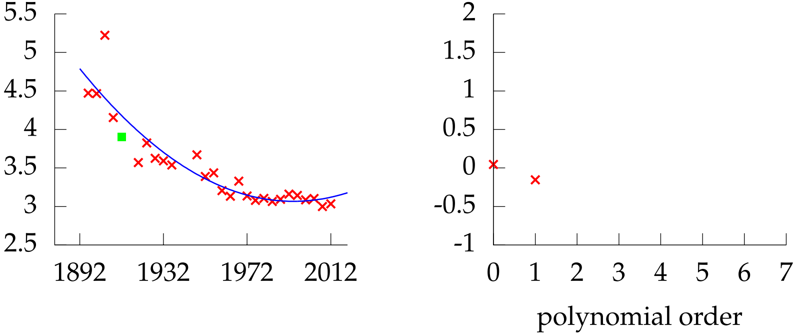
Leave One Out Error





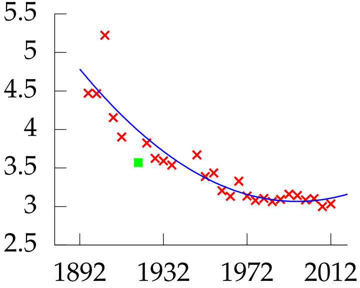
Polynomial order 2, training error -28.403, leave one out error 0.34669.

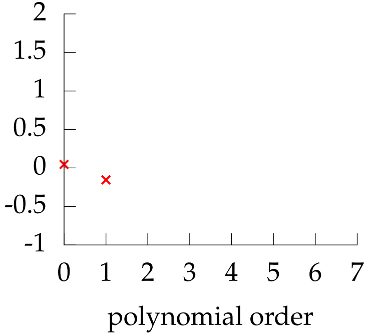
Leave One Out Error



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

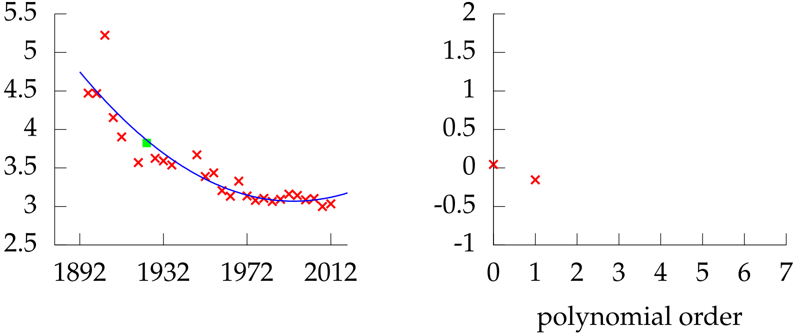
Leave One Out Error





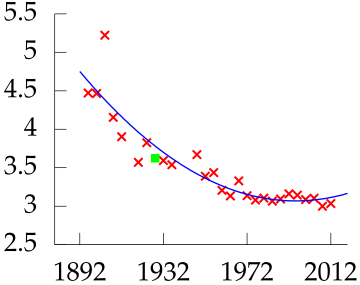
Polynomial order 2, training error -28.403, leave one out error 0.34669.

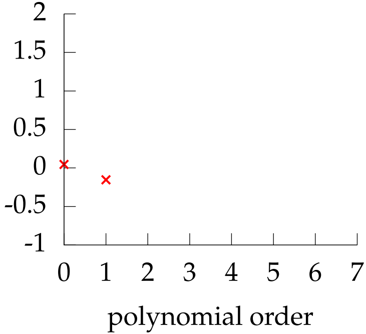
Leave One Out Error



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

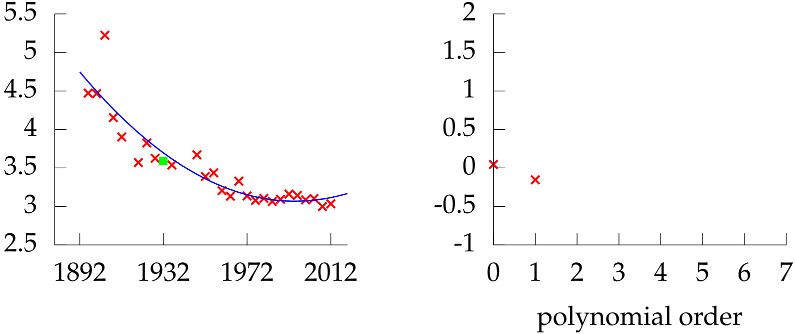
Leave One Out Error





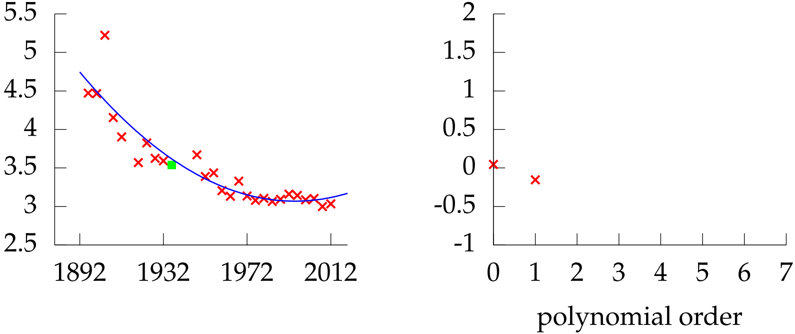
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



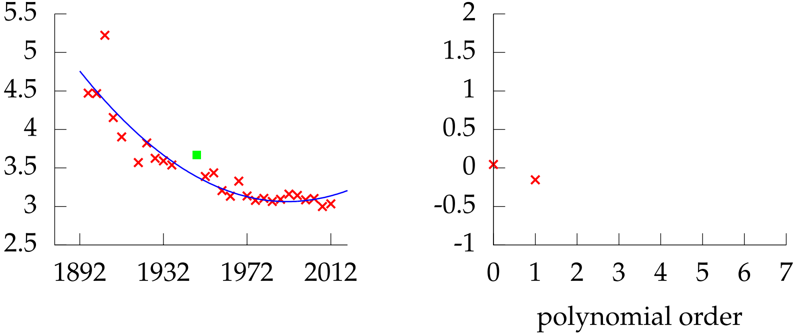
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



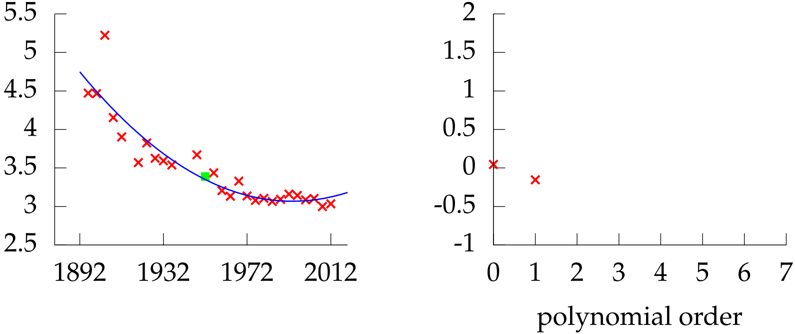
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



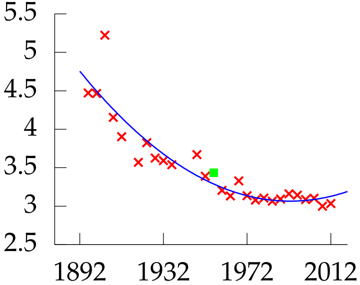
Polynomial order 2, training error -28.403, leave one out error 0.34669.

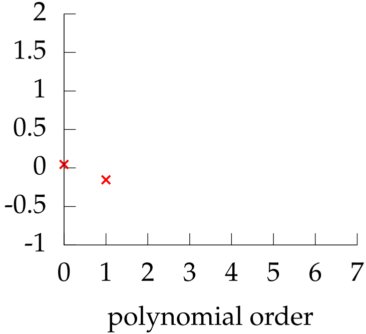
Leave One Out Error



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

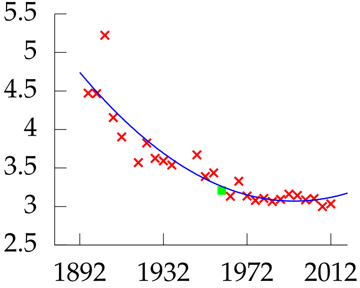
Leave One Out Error

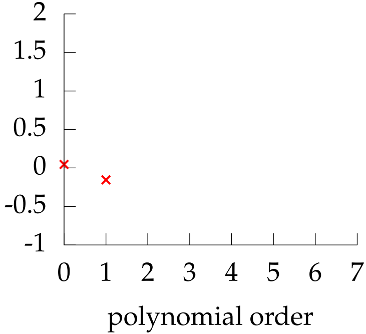




Polynomial order 2, training error -28.403, leave one out error 0.34669.

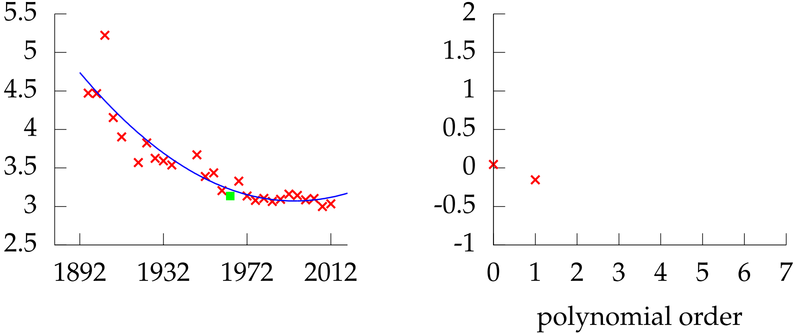
Leave One Out Error





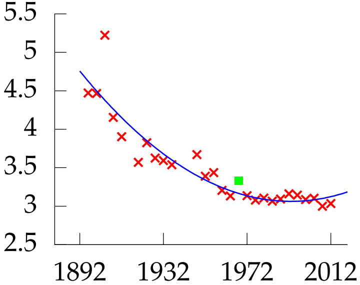
Polynomial order 2, training error -28.403, leave one out error 0.34669.

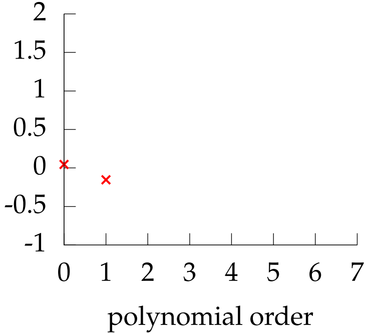
Leave One Out Error



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

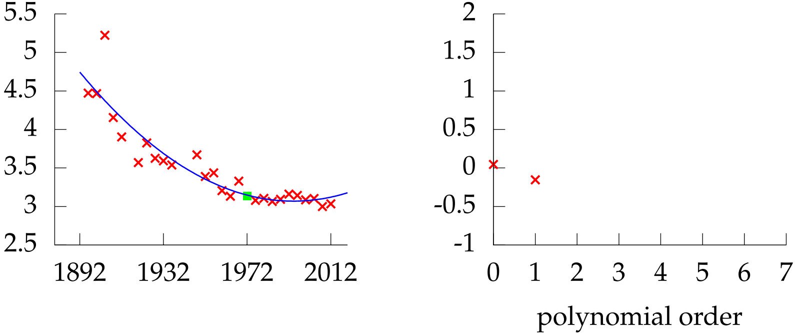
Leave One Out Error





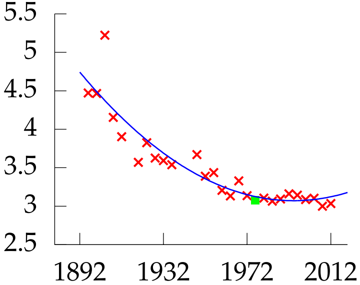
Polynomial order 2, training error -28.403, leave one out error 0.34669.

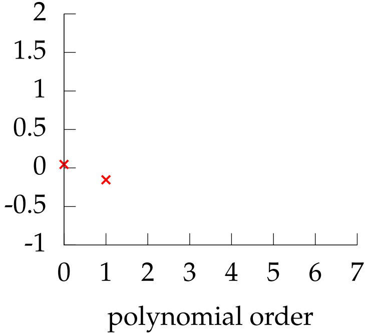
Leave One Out Error



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

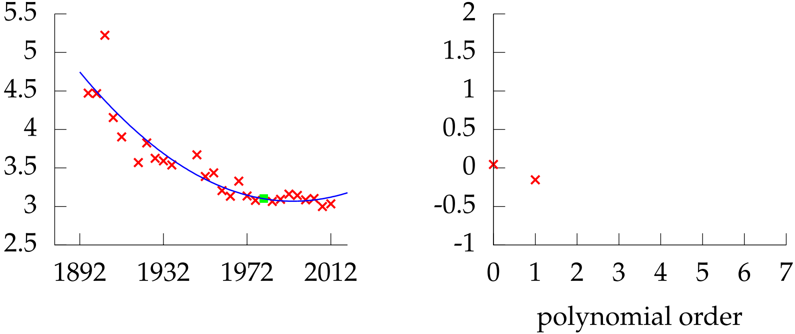
Leave One Out Error





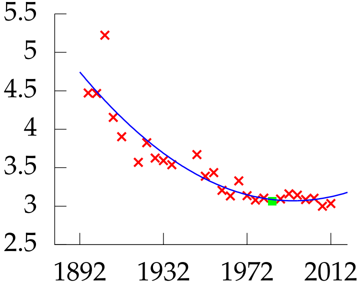
Polynomial order 2, training error -28.403, leave one out error 0.34669.

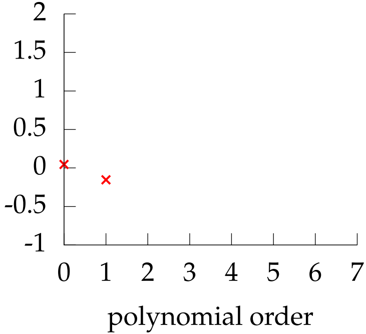
Leave One Out Error



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

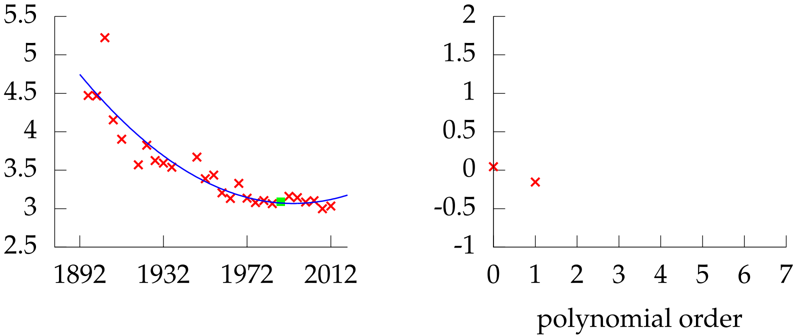
Leave One Out Error





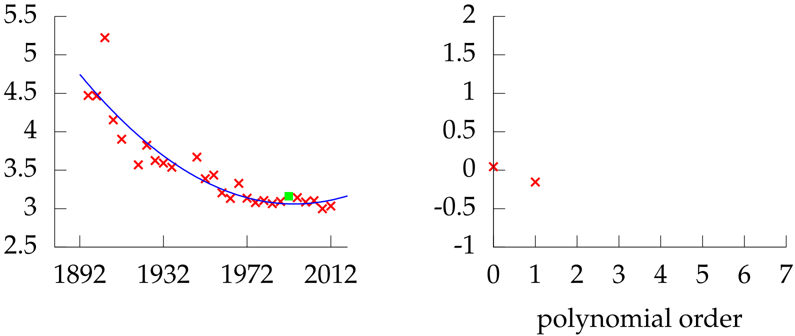
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



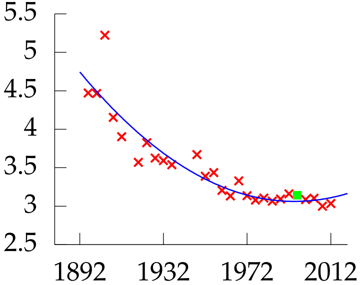
Polynomial order 2, training error -28.403, leave one out error 0.34669.

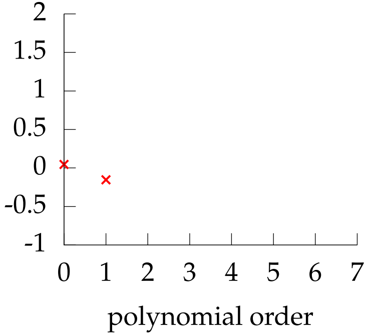
Leave One Out Error



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

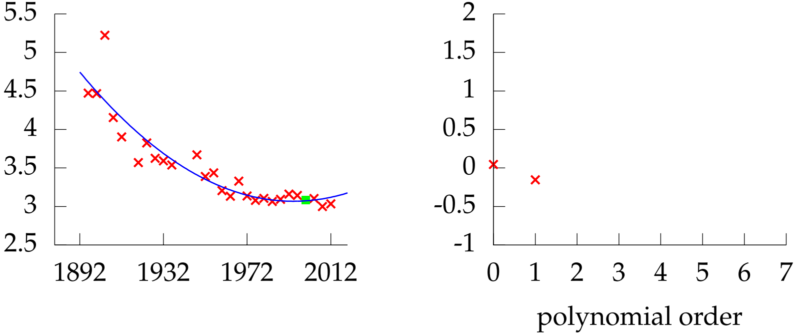
Leave One Out Error





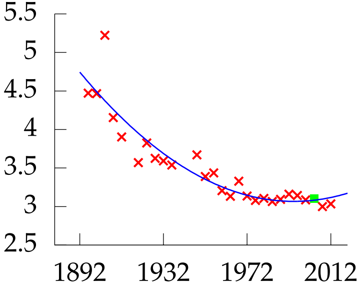
Polynomial order 2, training error -28.403, leave one out error 0.34669.

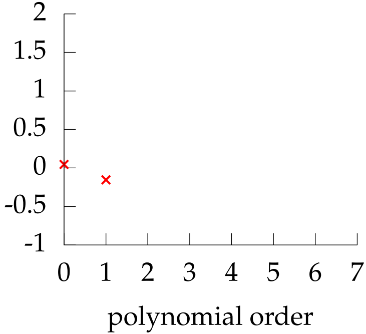
Leave One Out Error



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

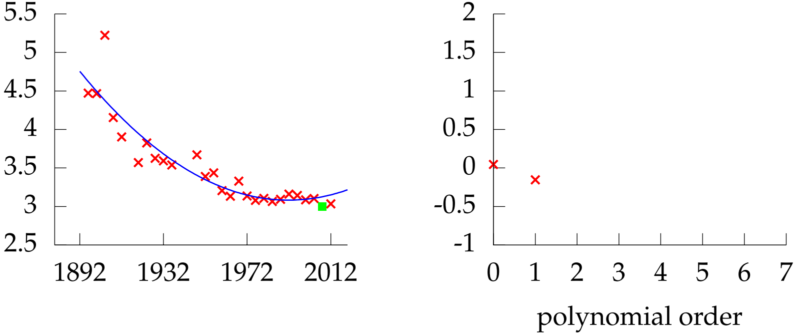
Leave One Out Error





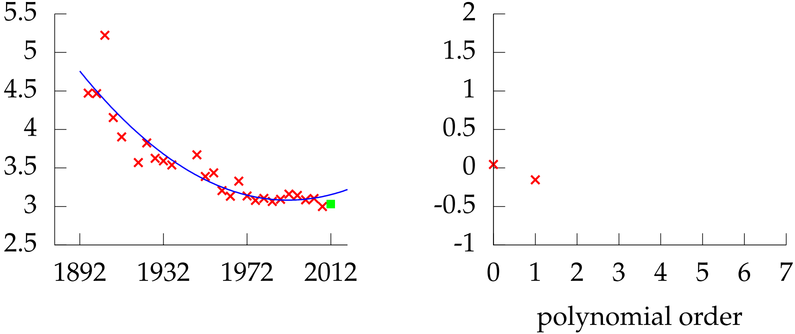
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



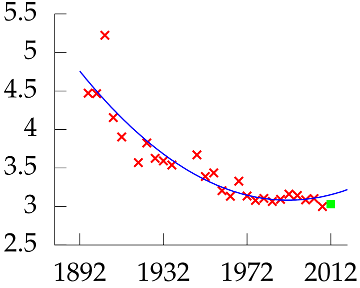
Polynomial order 2, training error -28.403, leave one out error 0.34669.

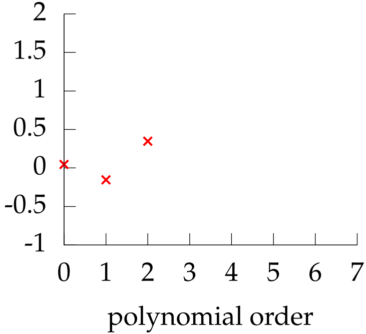
Leave One Out Error



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

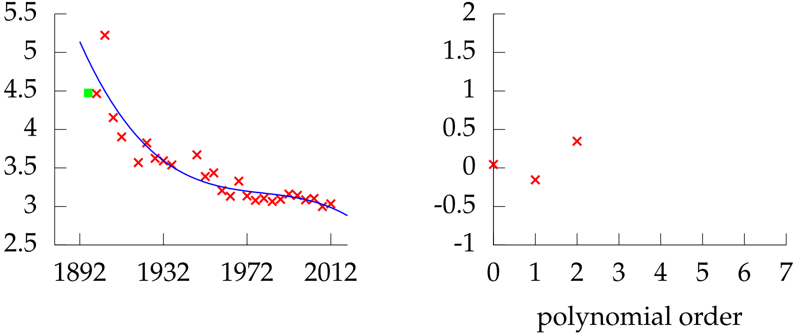
Leave One Out Error





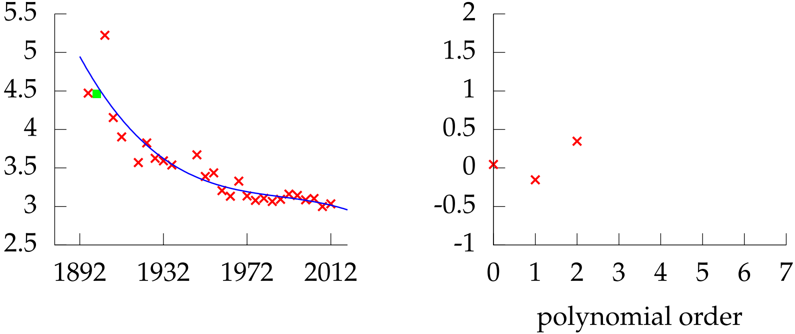
Polynomial order 2, training error -28.403, leave one out error 0.34669.

Leave One Out Error



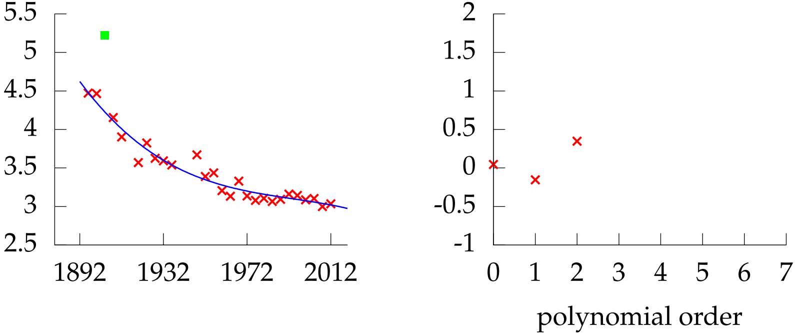
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



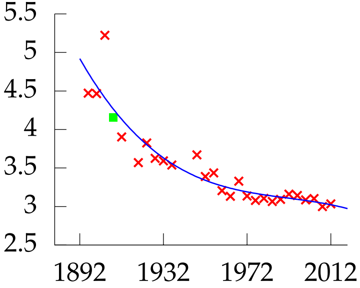
Polynomial order 3, training error -29.223, leave one out error 0.51621.

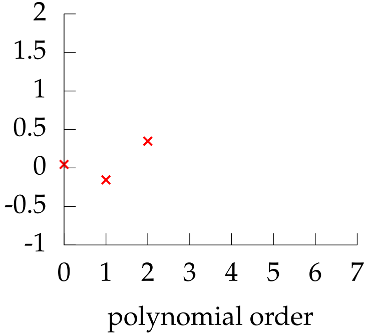
Leave One Out Error



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

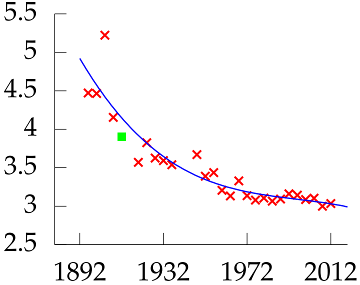
Leave One Out Error

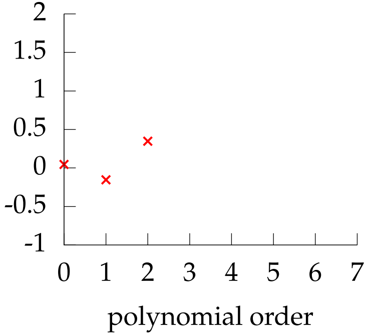




Polynomial order 3, training error -29.223, leave one out error 0.51621.

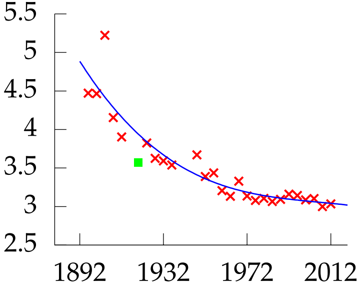
Leave One Out Error

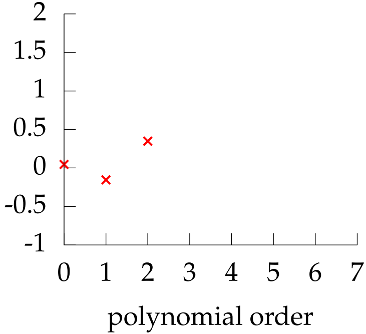




Polynomial order 3, training error -29.223, leave one out error 0.51621.

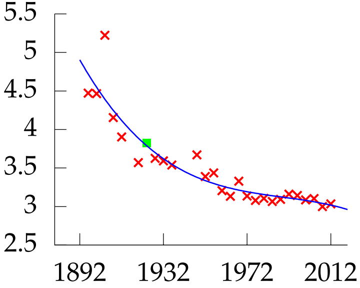
Leave One Out Error

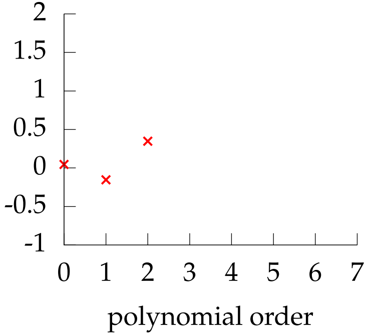




Polynomial order 3, training error -29.223, leave one out error 0.51621.

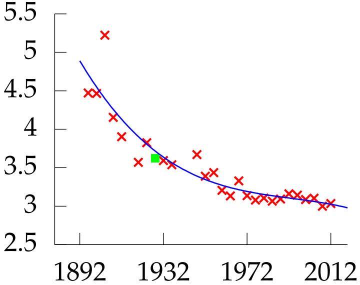
Leave One Out Error

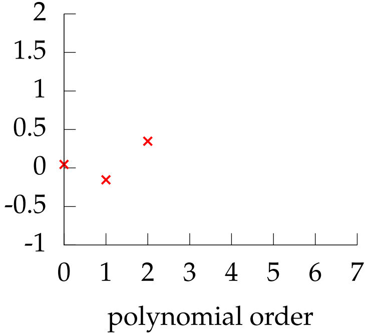




Polynomial order 3, training error -29.223, leave one out error 0.51621.

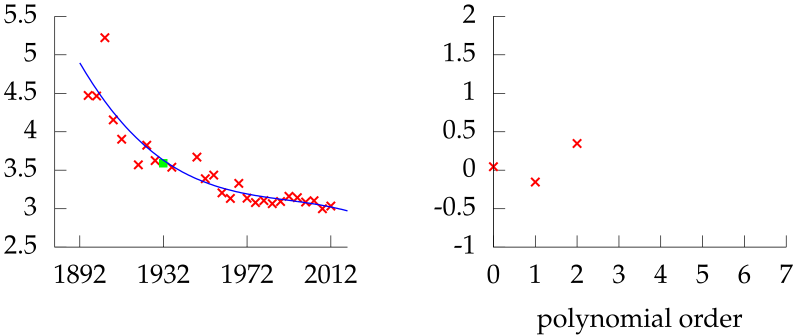
Leave One Out Error





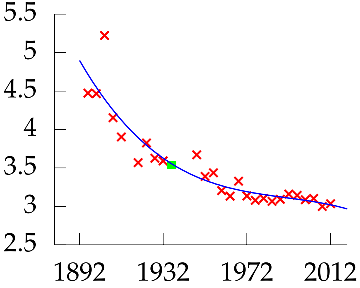
Polynomial order 3, training error -29.223, leave one out error 0.51621.

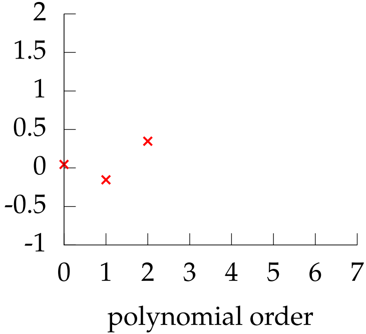
Leave One Out Error



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

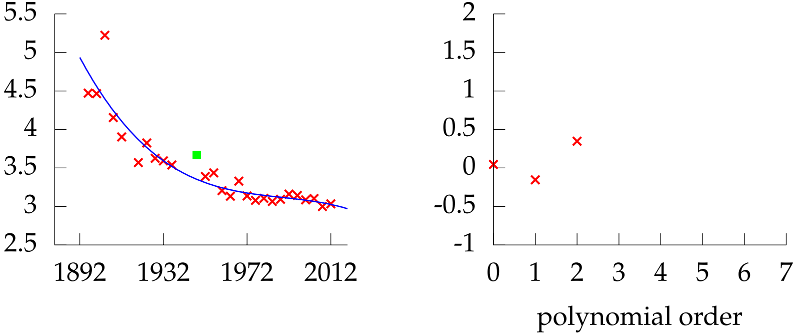
Leave One Out Error





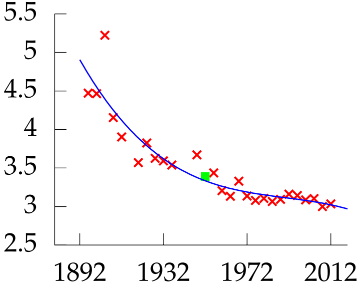
Polynomial order 3, training error -29.223, leave one out error 0.51621.

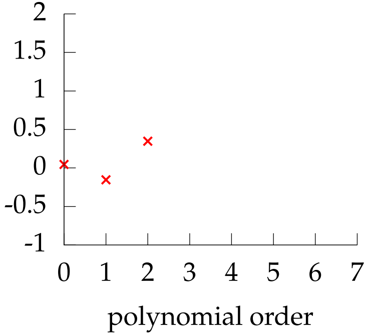
Leave One Out Error



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

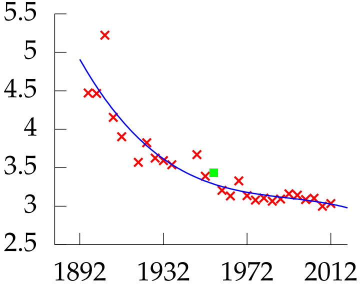
Leave One Out Error

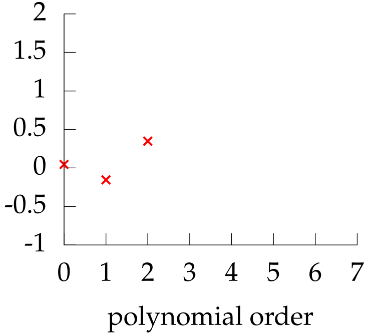




Polynomial order 3, training error -29.223, leave one out error 0.51621.

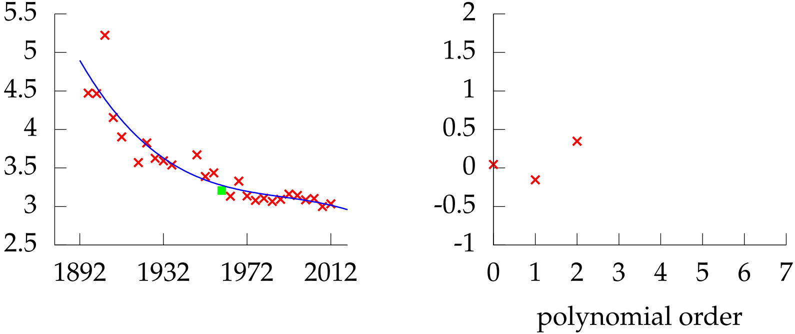
Leave One Out Error





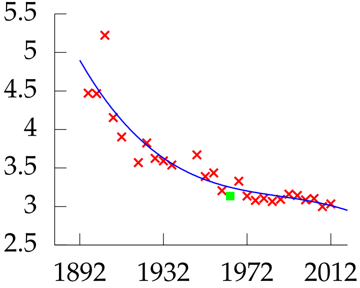
Polynomial order 3, training error -29.223, leave one out error 0.51621.

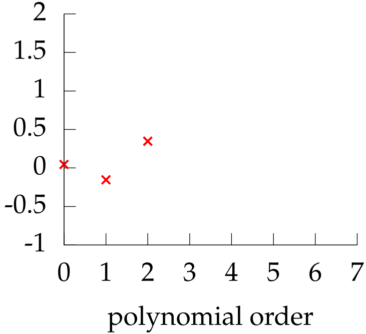
Leave One Out Error



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

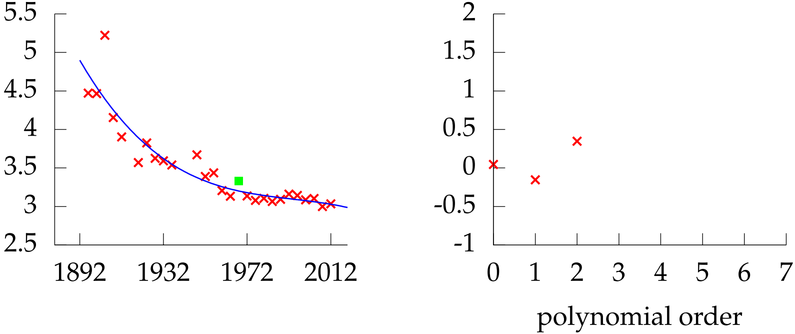
Leave One Out Error





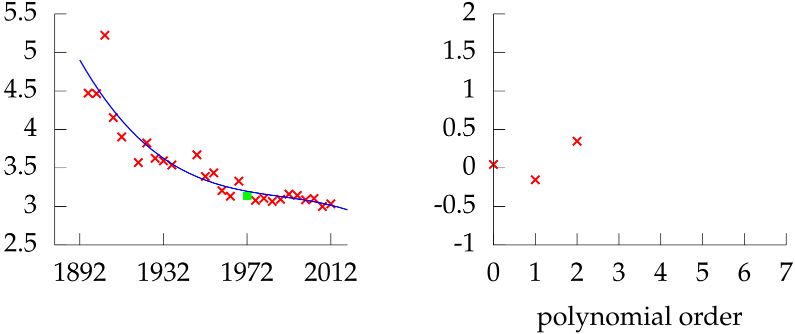
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



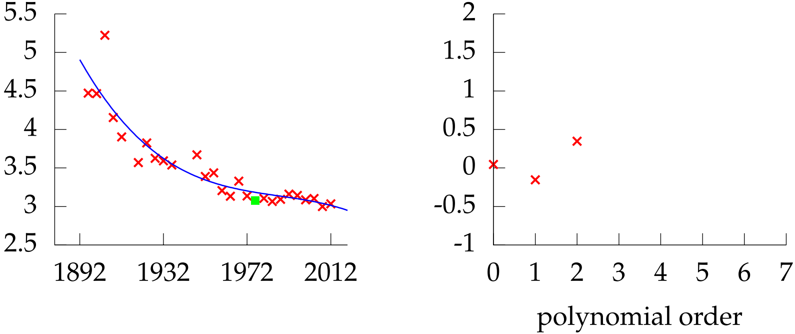
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



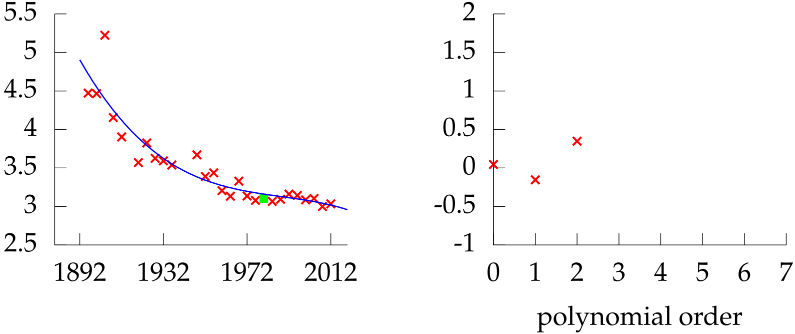
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



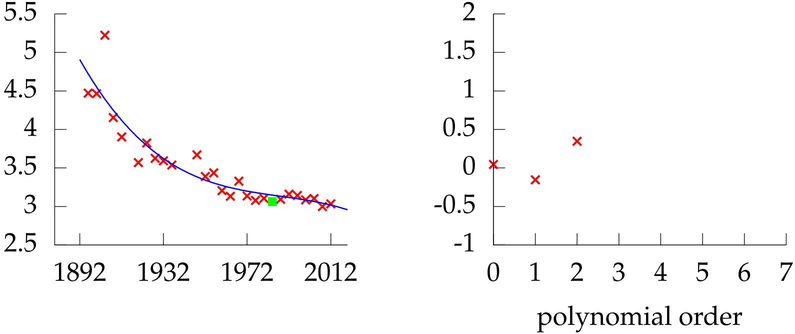
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



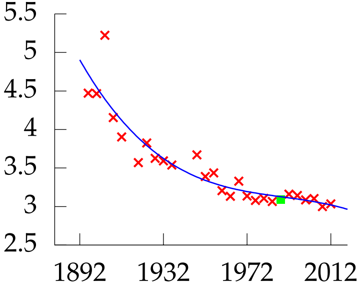
Polynomial order 3, training error -29.223, leave one out error 0.51621.

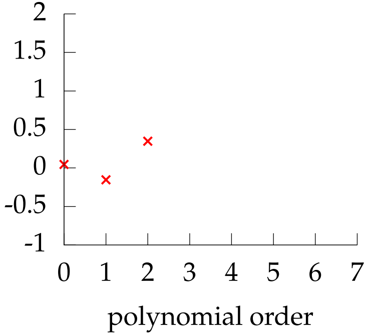
Leave One Out Error



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

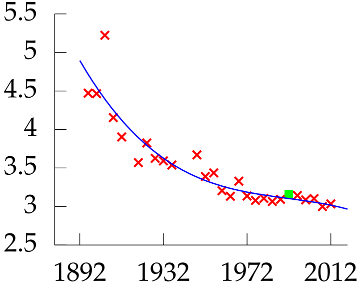
Leave One Out Error

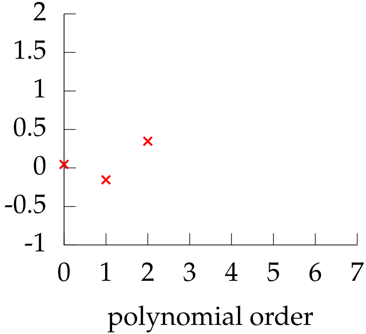




Polynomial order 3, training error -29.223, leave one out error 0.51621.

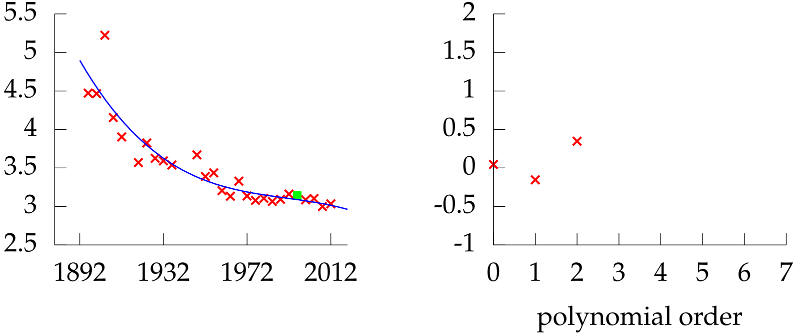
Leave One Out Error





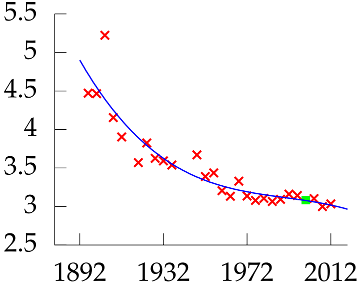
Polynomial order 3, training error -29.223, leave one out error 0.51621.

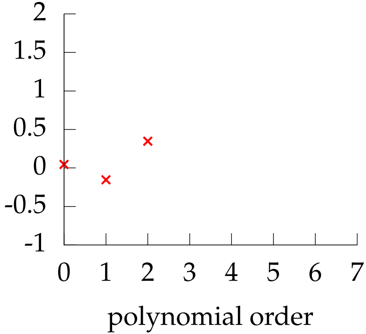
Leave One Out Error



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

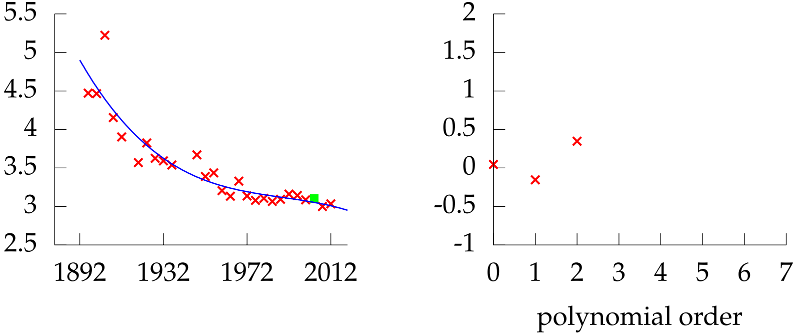
Leave One Out Error





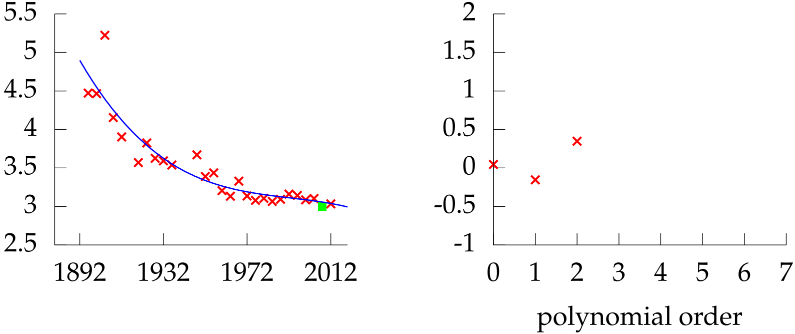
Polynomial order 3, training error -29.223, leave one out error 0.51621.

Leave One Out Error



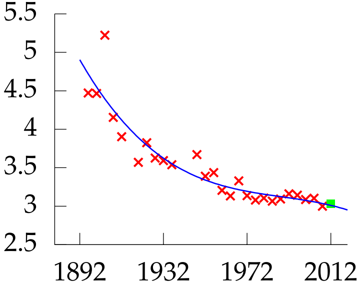
Polynomial order 3, training error -29.223, leave one out error 0.51621.

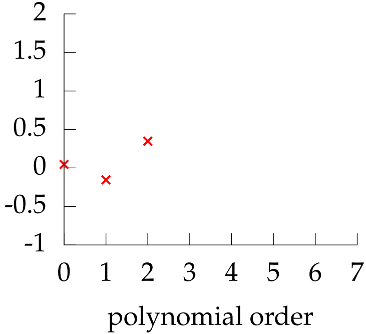
Leave One Out Error



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

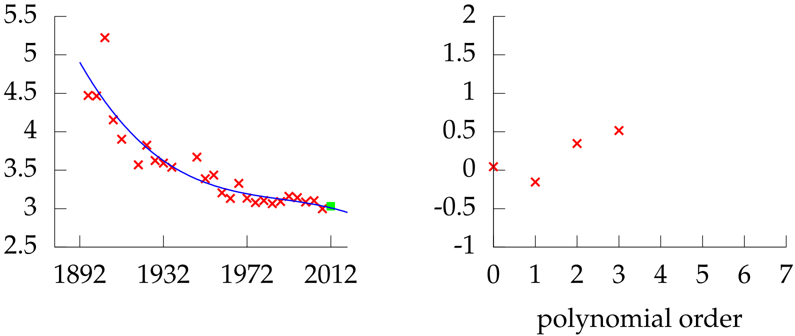
Leave One Out Error





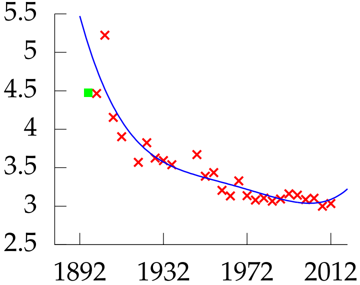
Polynomial order 3, training error -29.223, leave one out error 0.51621.

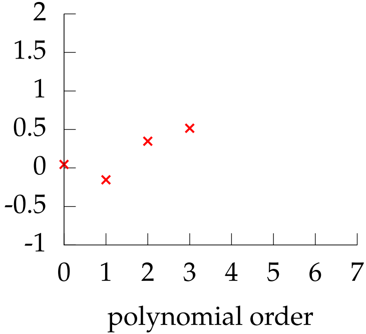
Leave One Out Error



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

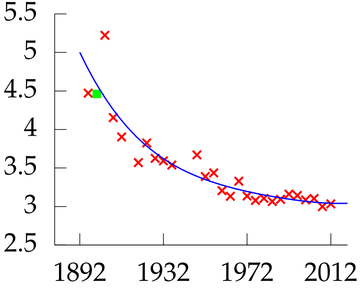
Leave One Out Error

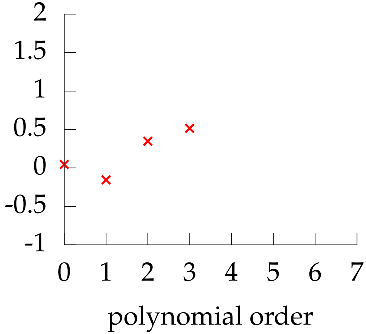




Polynomial order 4, training error -29.324, leave one out error 0.84844.

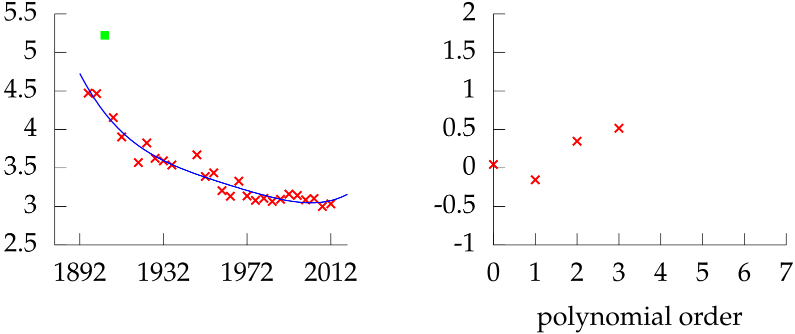
Leave One Out Error





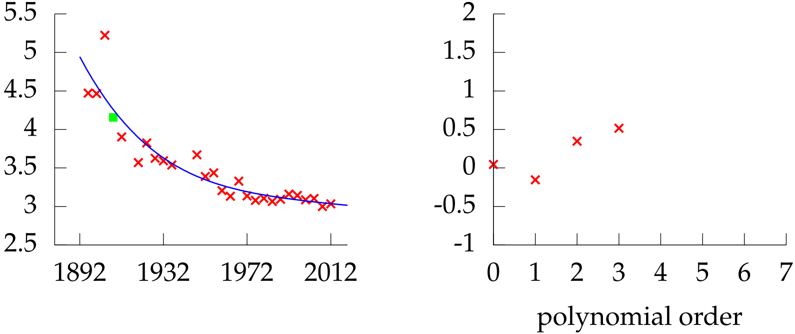
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



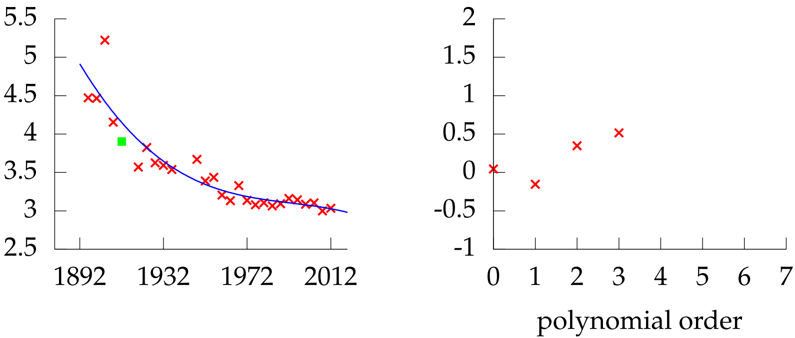
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



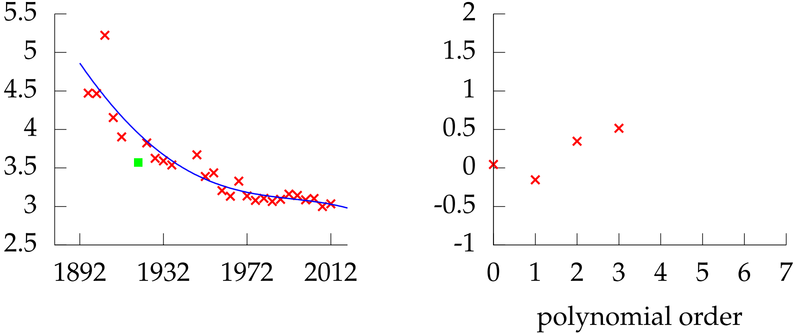
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



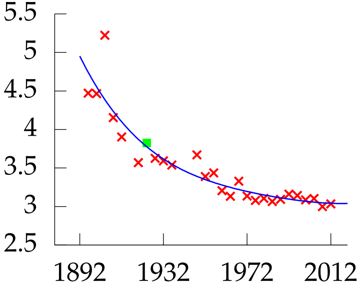
Polynomial order 4, training error -29.324, leave one out error 0.84844.

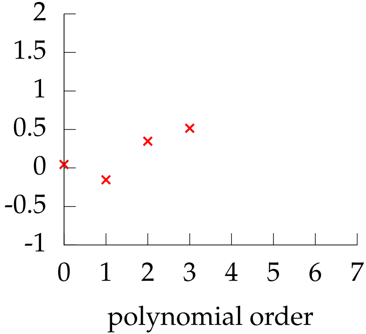
Leave One Out Error



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

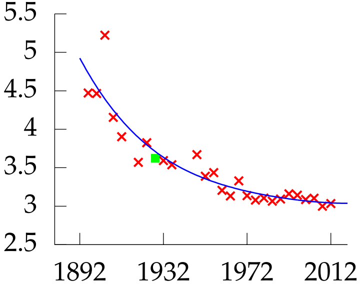
Leave One Out Error

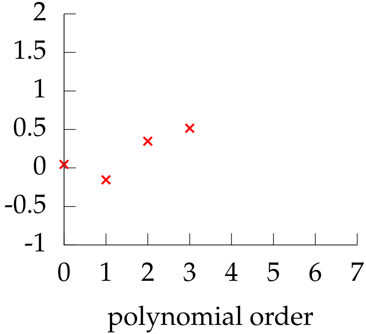




Polynomial order 4, training error -29.324, leave one out error 0.84844.

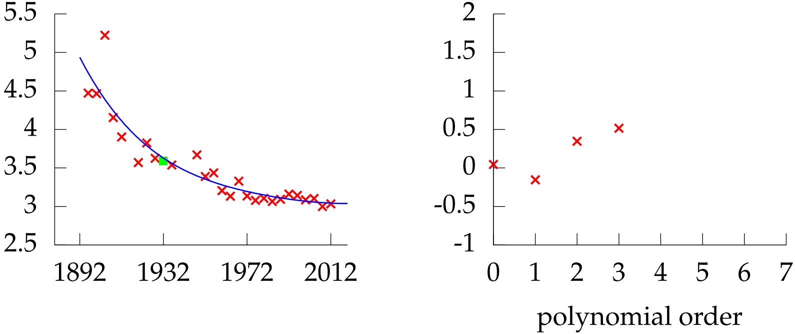
Leave One Out Error





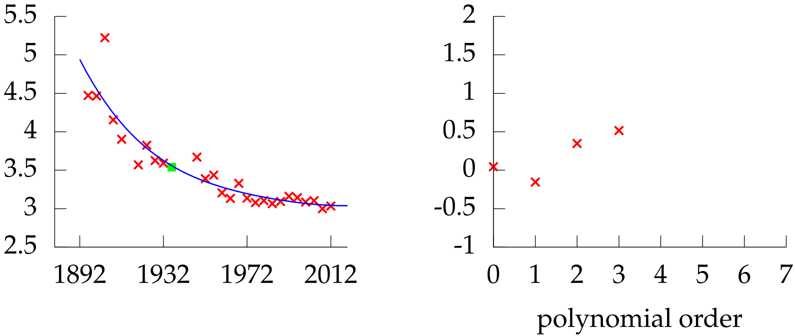
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



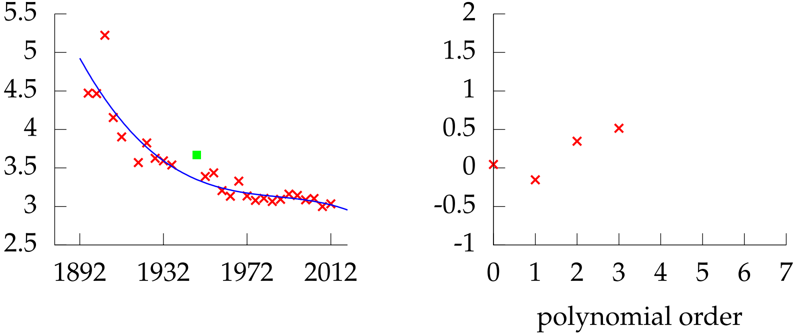
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



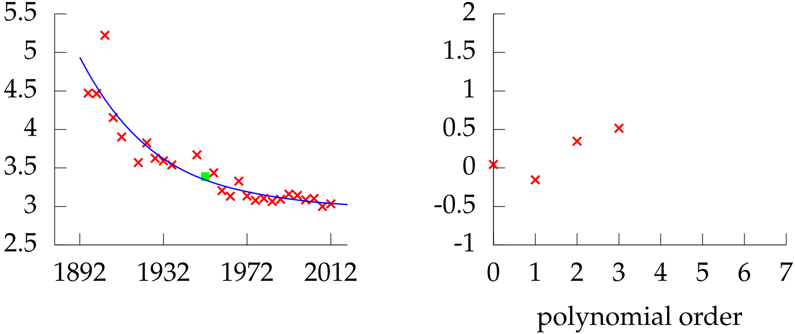
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



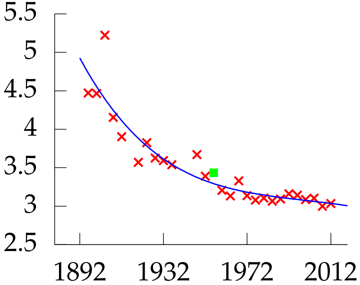
Polynomial order 4, training error -29.324, leave one out error 0.84844.

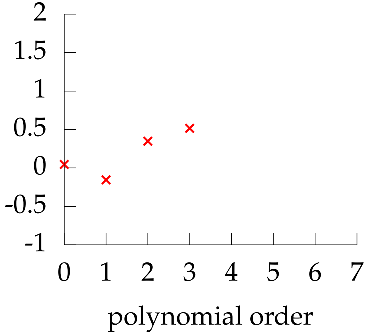
Leave One Out Error



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

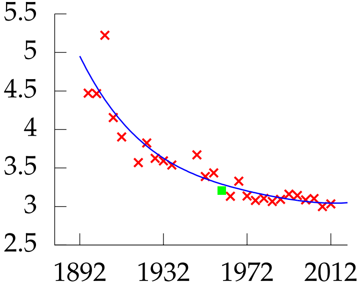
Leave One Out Error

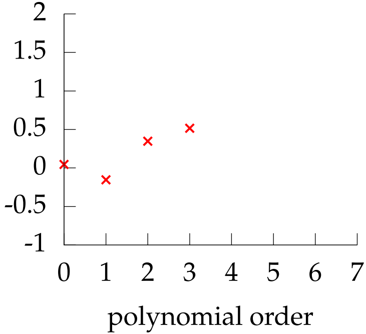




Polynomial order 4, training error -29.324, leave one out error 0.84844.

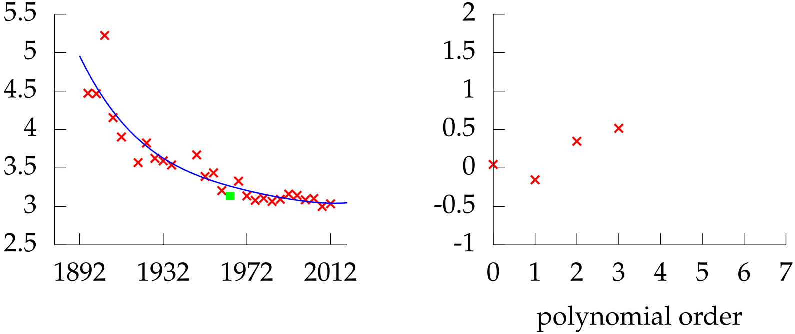
Leave One Out Error





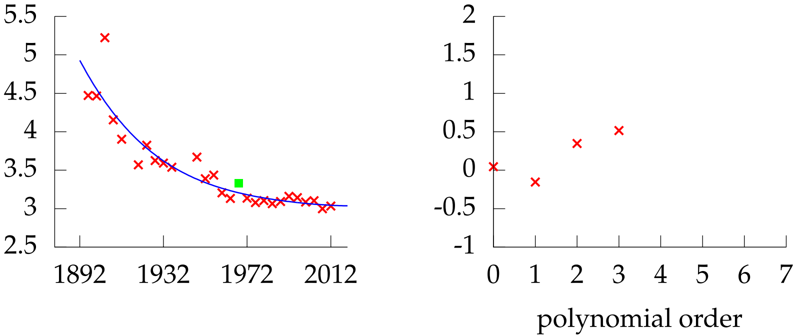
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



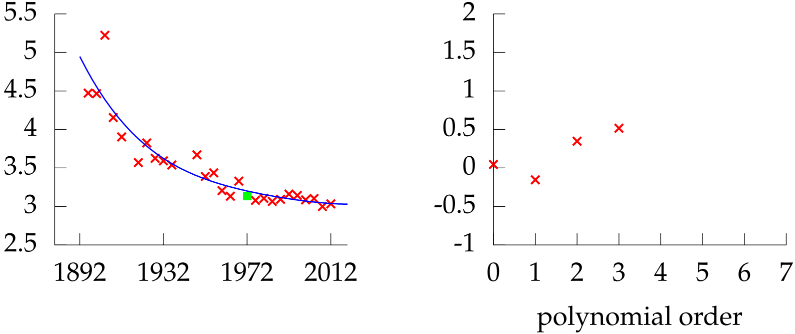
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



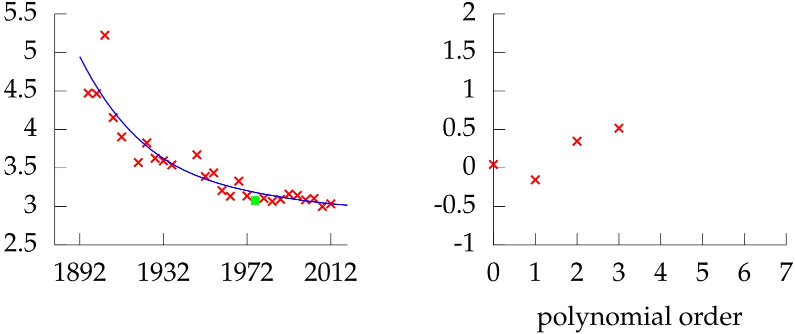
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



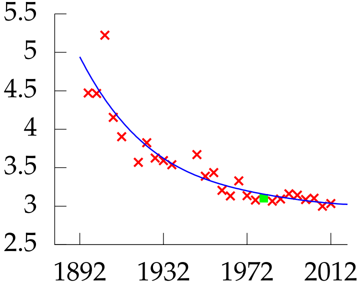
Polynomial order 4, training error -29.324, leave one out error 0.84844.

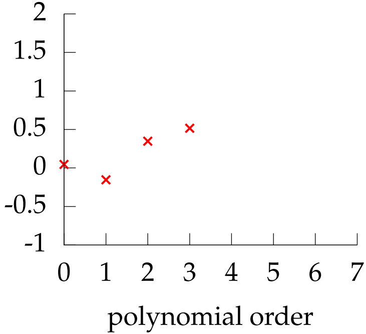
Leave One Out Error



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

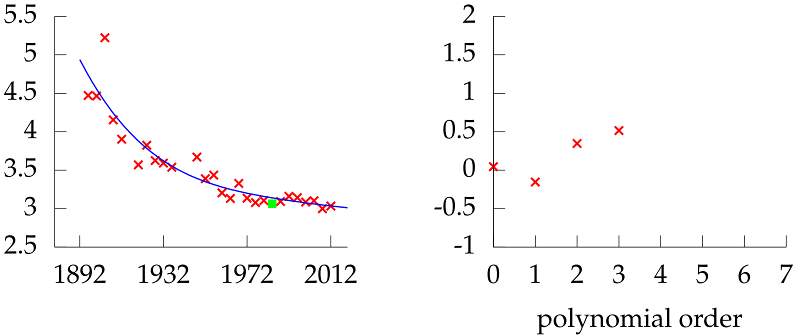
Leave One Out Error





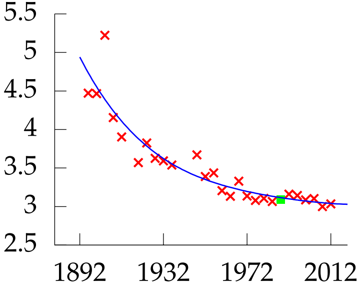
Polynomial order 4, training error -29.324, leave one out error 0.84844.

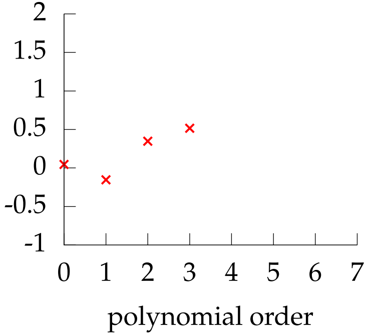
Leave One Out Error



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

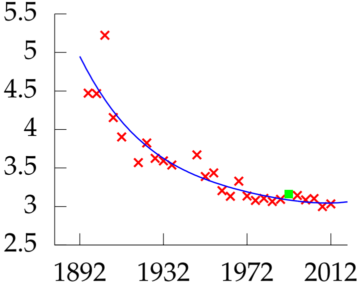
Leave One Out Error

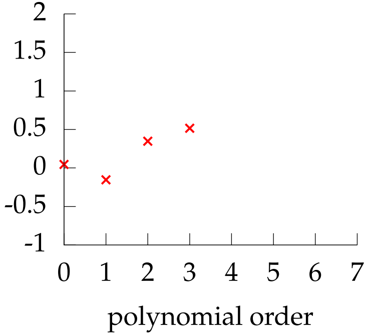




Polynomial order 4, training error -29.324, leave one out error 0.84844.

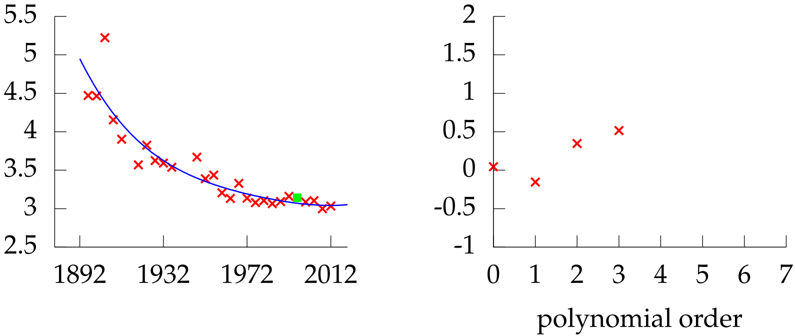
Leave One Out Error





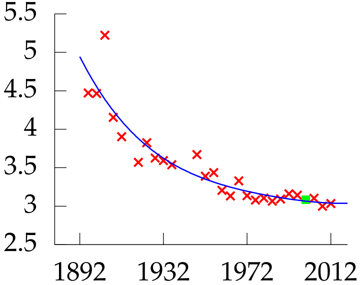
Polynomial order 4, training error -29.324, leave one out error 0.84844.

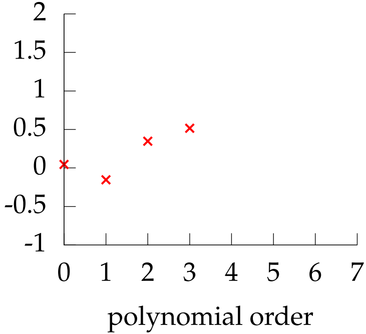
Leave One Out Error



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

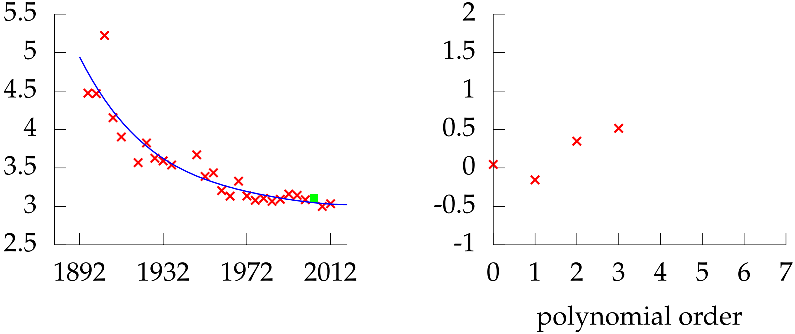
Leave One Out Error





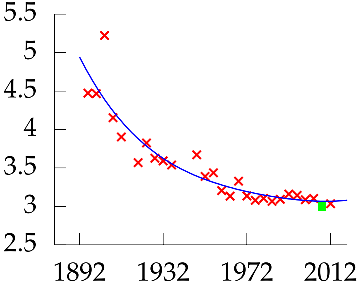
Polynomial order 4, training error -29.324, leave one out error 0.84844.

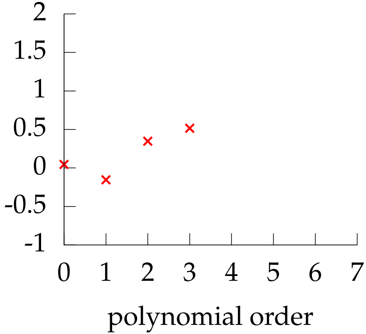
Leave One Out Error



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

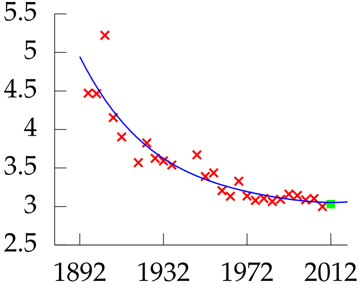
Leave One Out Error

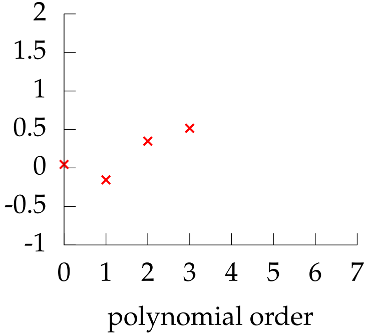




Polynomial order 4, training error -29.324, leave one out error 0.84844.

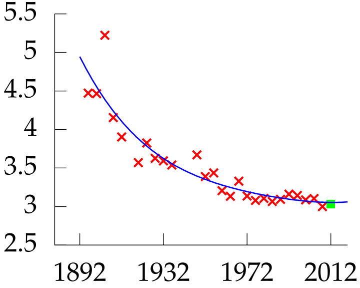
Leave One Out Error

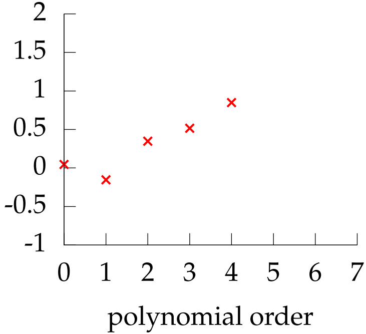




Polynomial order 4, training error -29.324, leave one out error 0.84844.

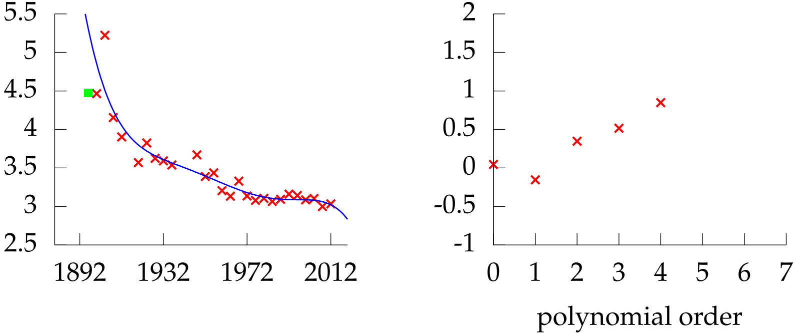
Leave One Out Error





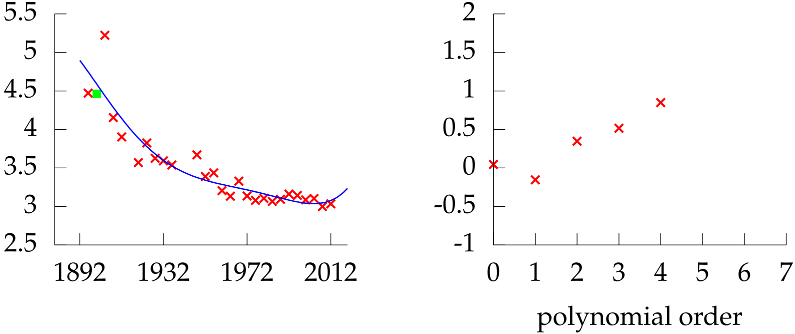
Polynomial order 4, training error -29.324, leave one out error 0.84844.

Leave One Out Error



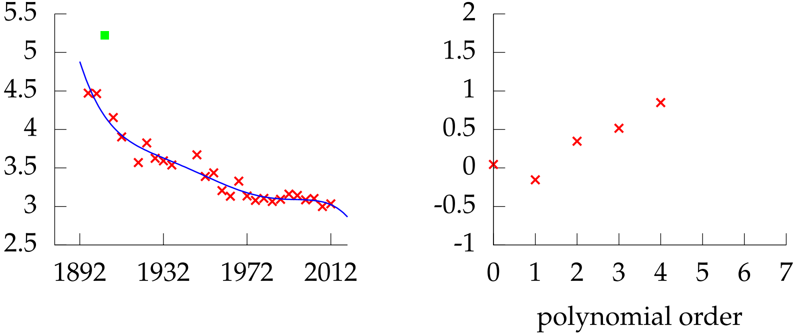
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



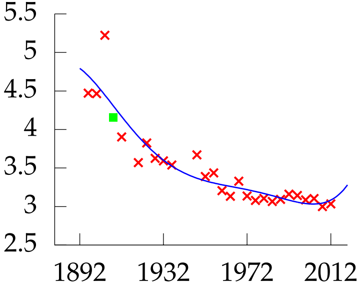
Polynomial order 5, training error -29.524, leave one out error 1.48.

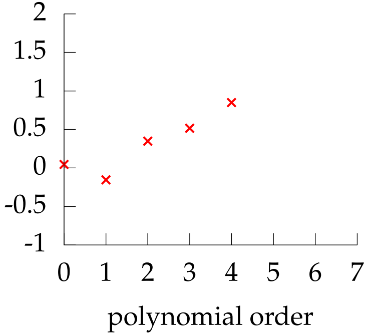
Leave One Out Error



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

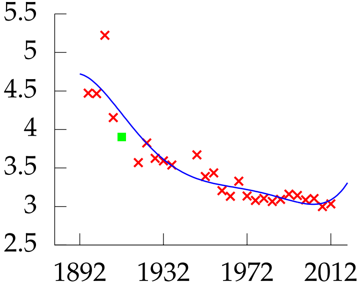
Leave One Out Error

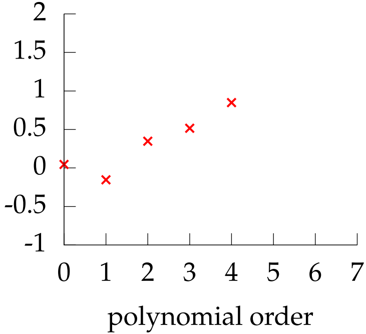




Polynomial order 5, training error -29.524, leave one out error 1.48.

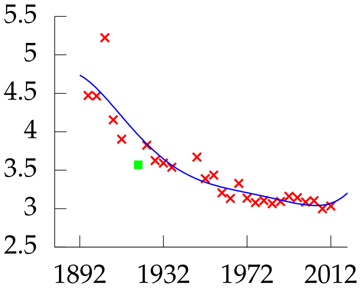
Leave One Out Error

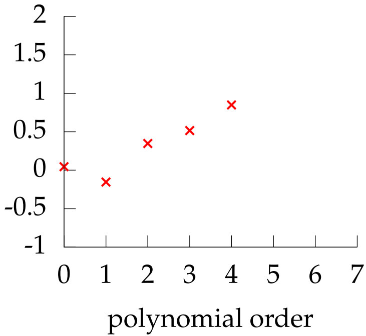




Polynomial order 5, training error -29.524, leave one out error 1.48.

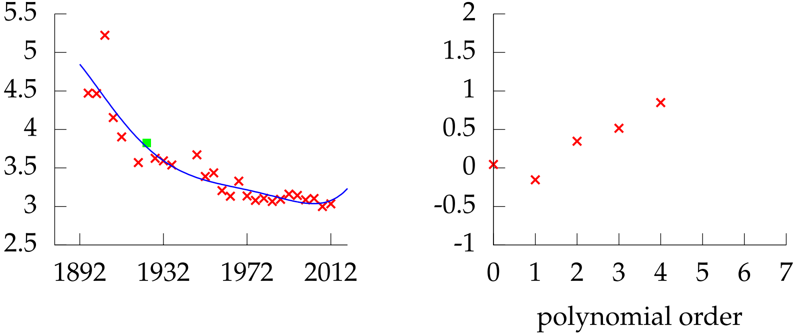
Leave One Out Error





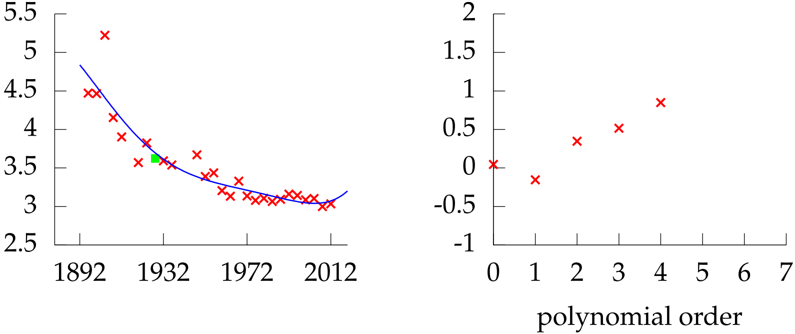
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



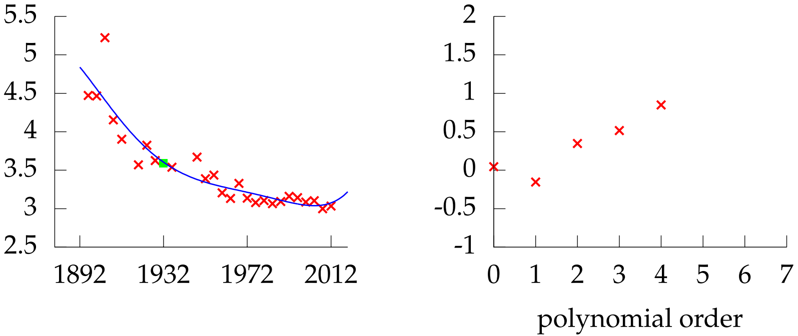
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



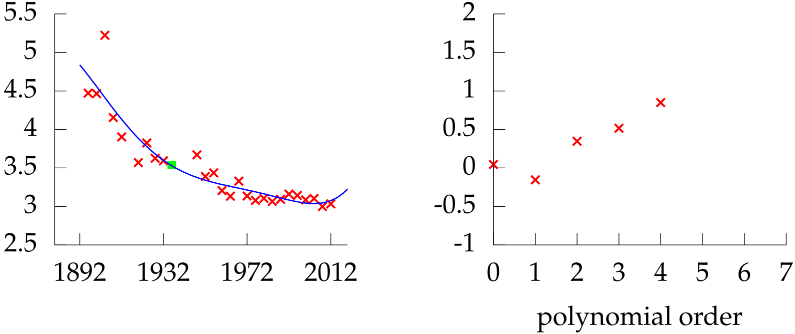
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



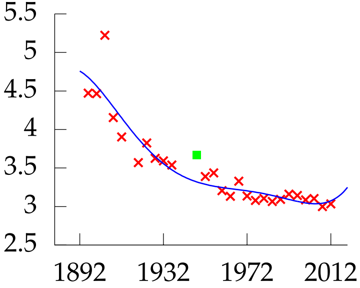
Polynomial order 5, training error -29.524, leave one out error 1.48.

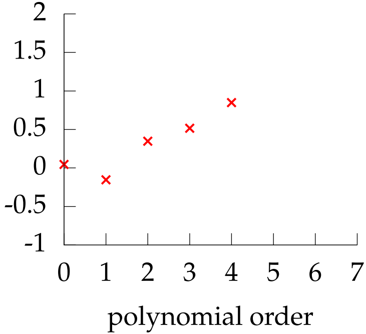
Leave One Out Error



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

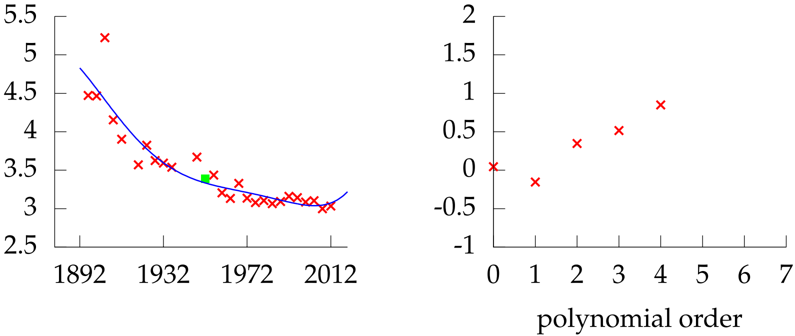
Leave One Out Error





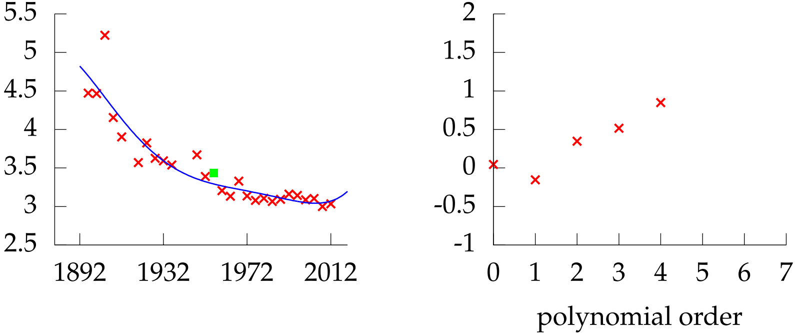
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



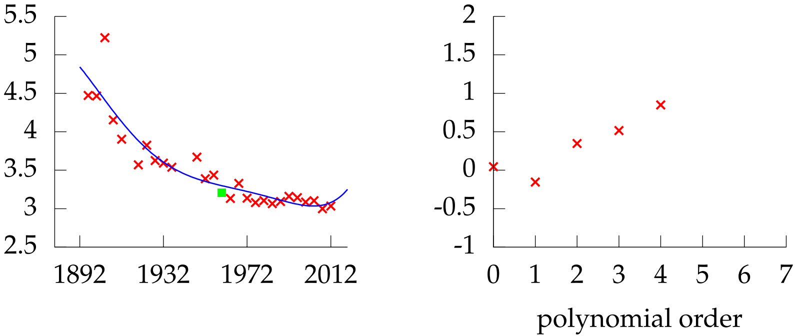
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



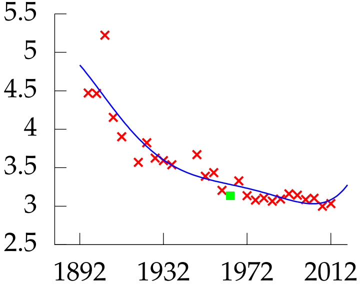
Polynomial order 5, training error -29.524, leave one out error 1.48.

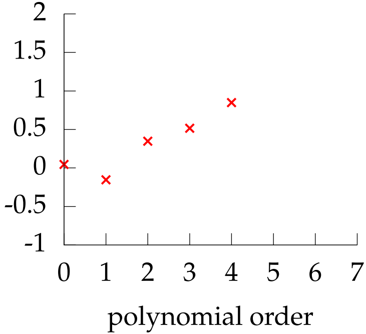
Leave One Out Error



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

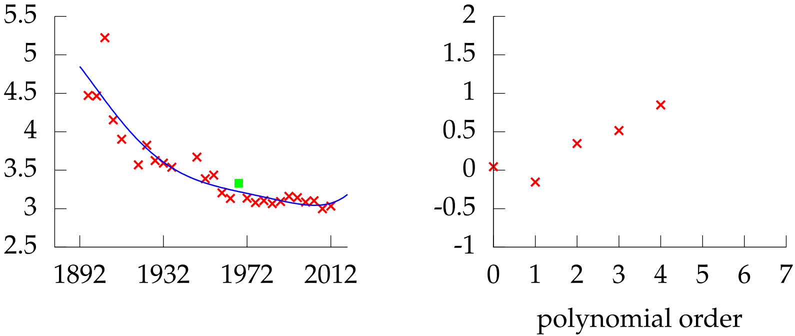
Leave One Out Error





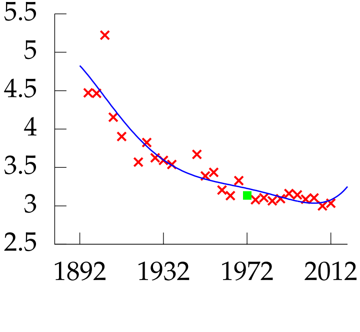
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



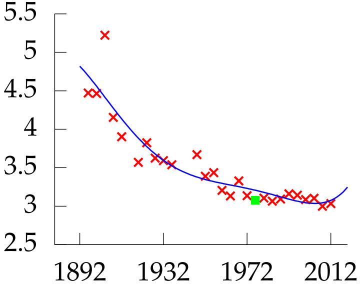
Polynomial order 5, training error -29.524, leave one out error 1.48.

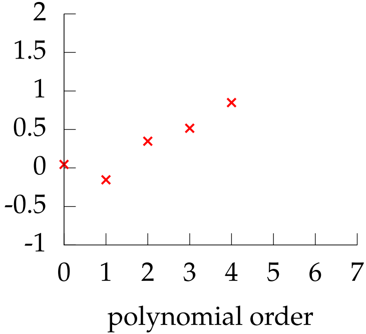
Leave One Out Error



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

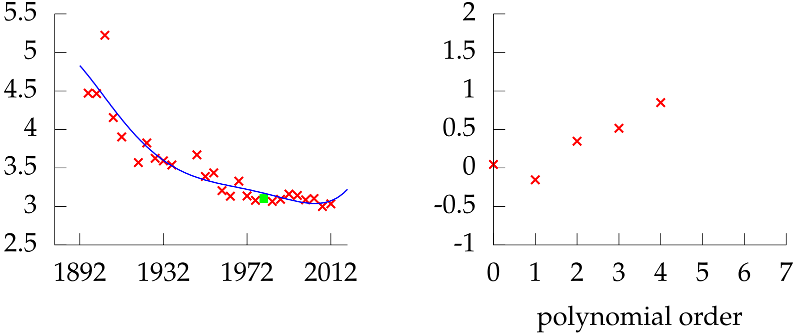
Leave One Out Error





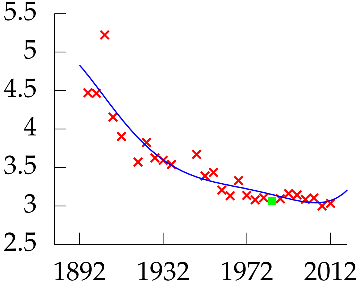
Polynomial order 5, training error -29.524, leave one out error 1.48.

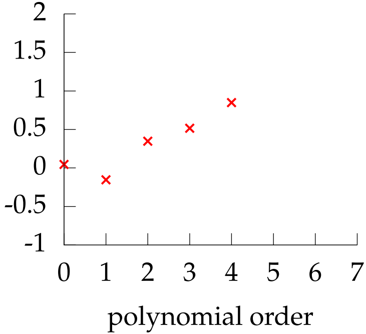
Leave One Out Error



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

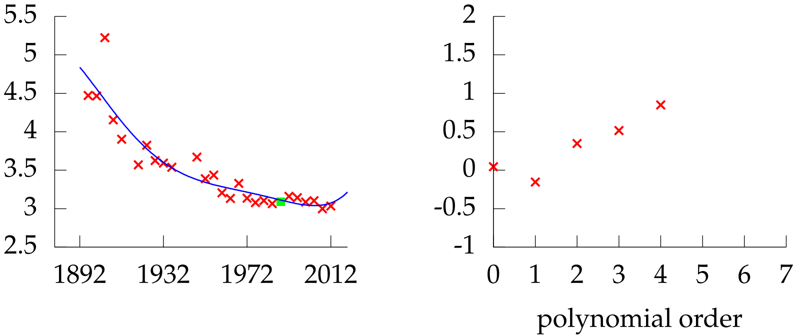
Leave One Out Error





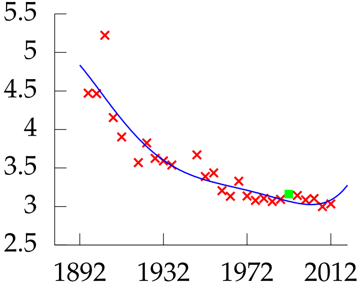
Polynomial order 5, training error -29.524, leave one out error 1.48.

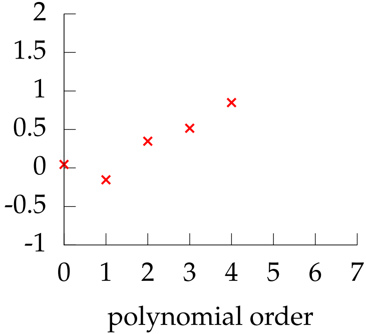
Leave One Out Error



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

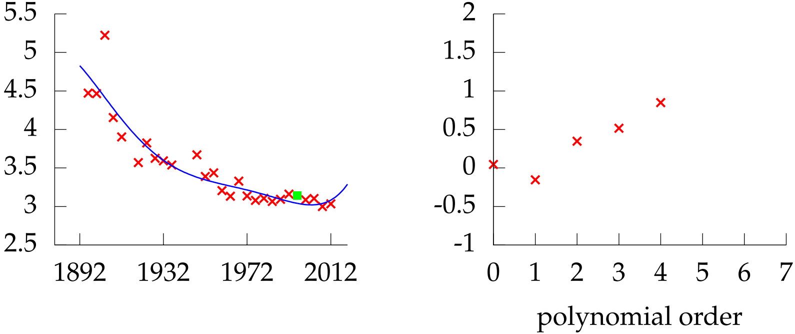
Leave One Out Error





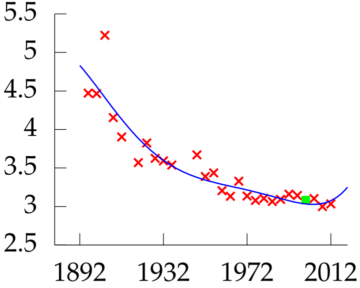
Polynomial order 5, training error -29.524, leave one out error 1.48.

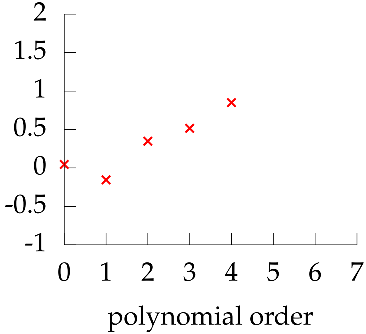
Leave One Out Error



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

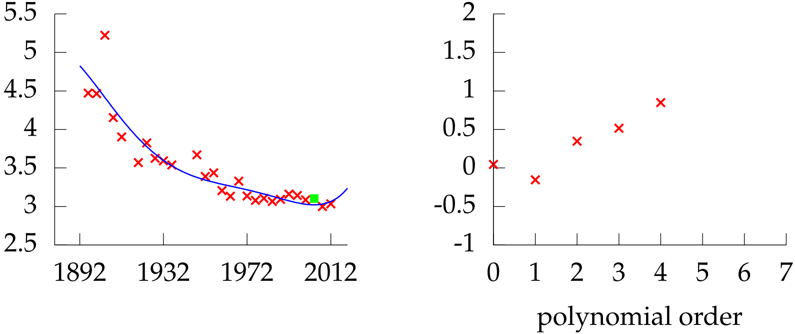
Leave One Out Error





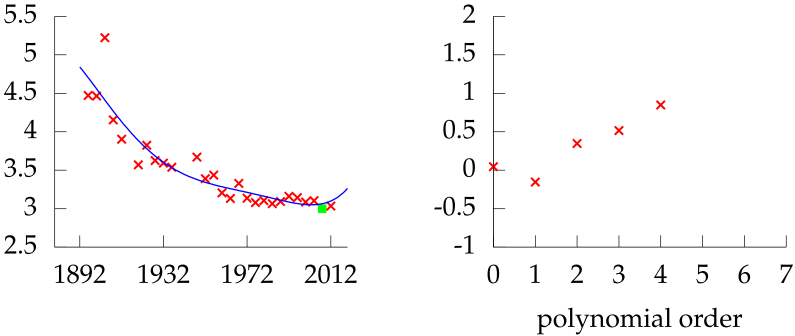
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



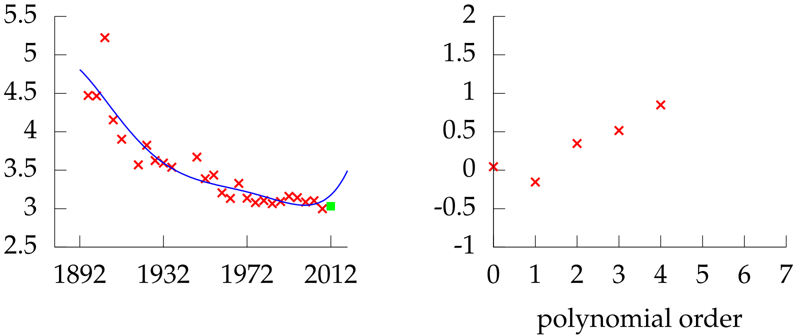
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



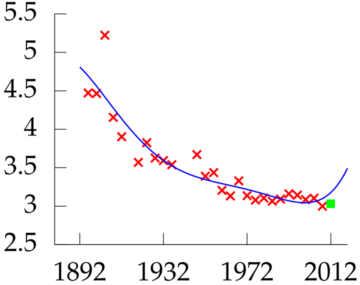
Polynomial order 5, training error -29.524, leave one out error 1.48.

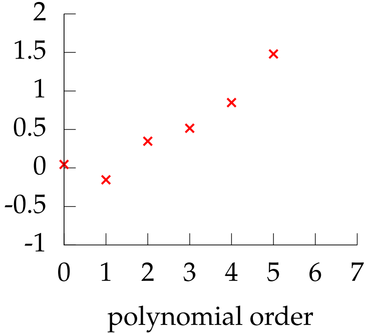
Leave One Out Error



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

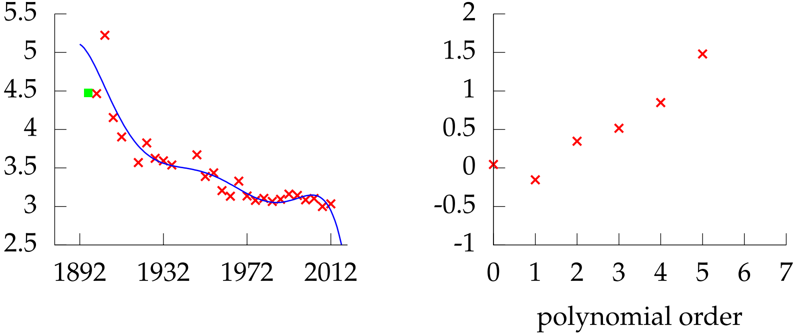
Leave One Out Error





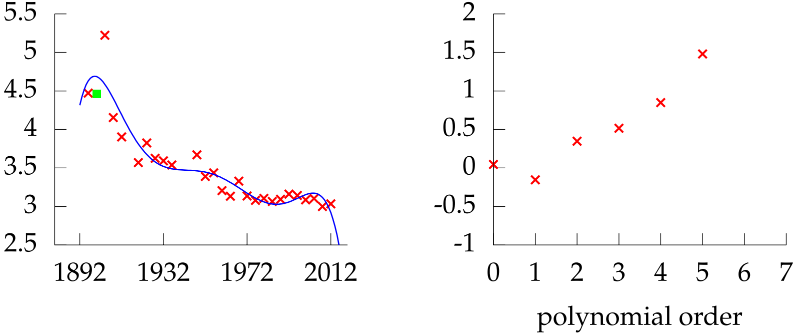
Polynomial order 5, training error -29.524, leave one out error 1.48.

Leave One Out Error



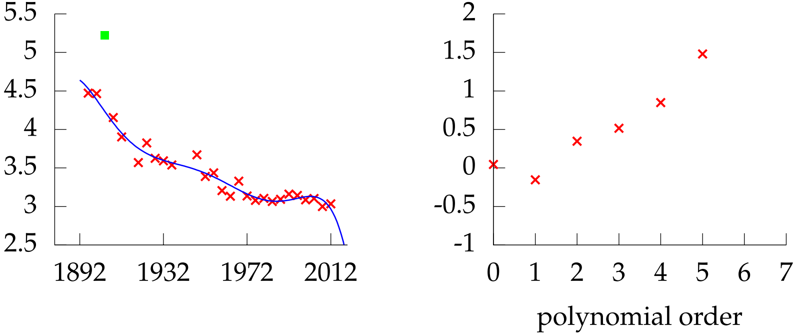
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



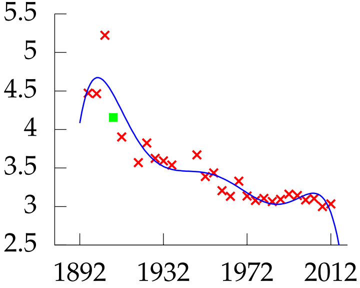
Polynomial order 6, training error -32.237, leave one out error 1.5047.

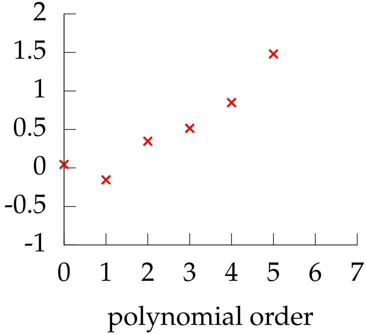
Leave One Out Error



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

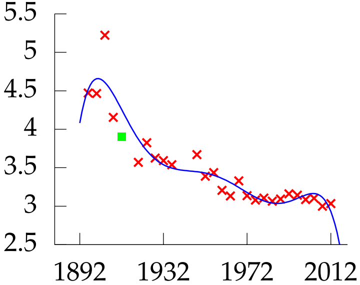
Leave One Out Error

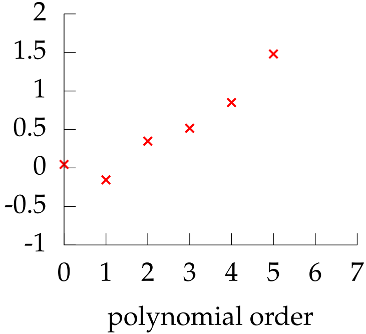




Polynomial order 6, training error -32.237, leave one out error 1.5047.

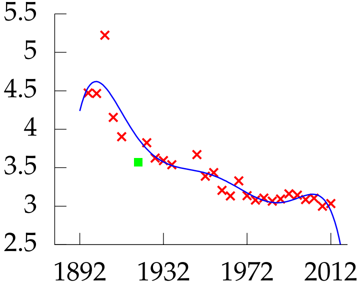
Leave One Out Error

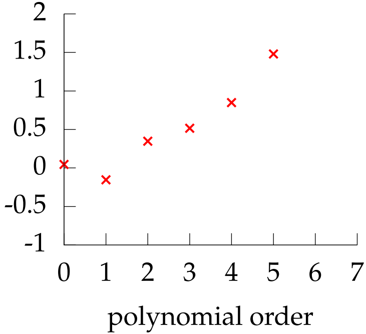




Polynomial order 6, training error -32.237, leave one out error 1.5047.

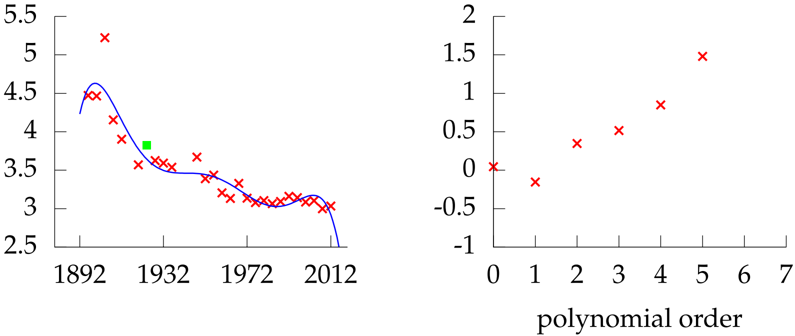
Leave One Out Error





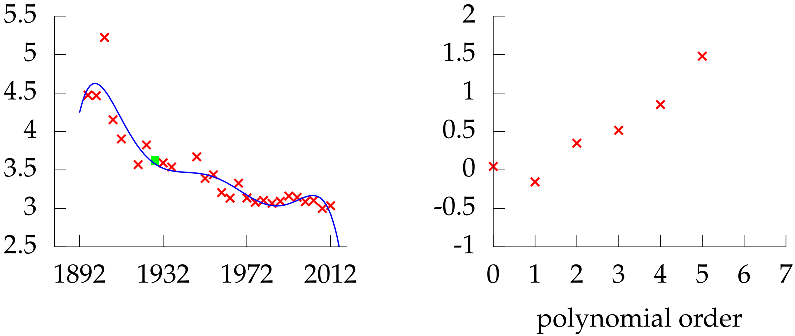
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



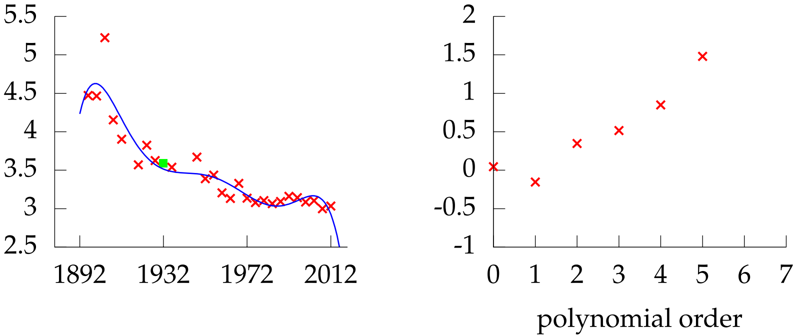
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



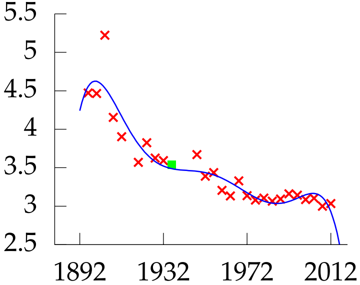
Polynomial order 6, training error -32.237, leave one out error 1.5047.

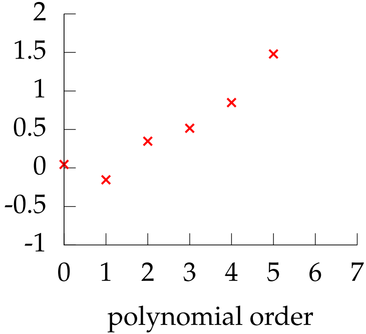
Leave One Out Error



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

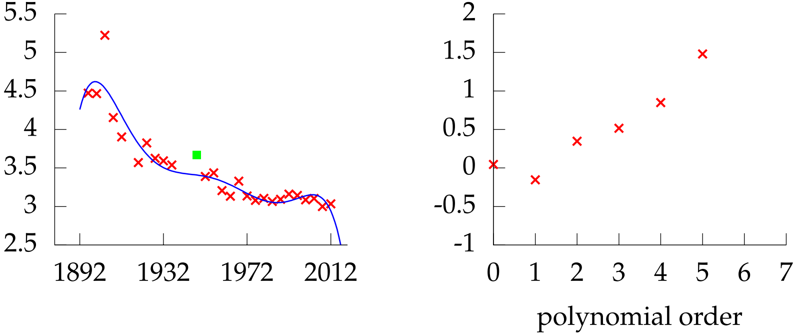
Leave One Out Error





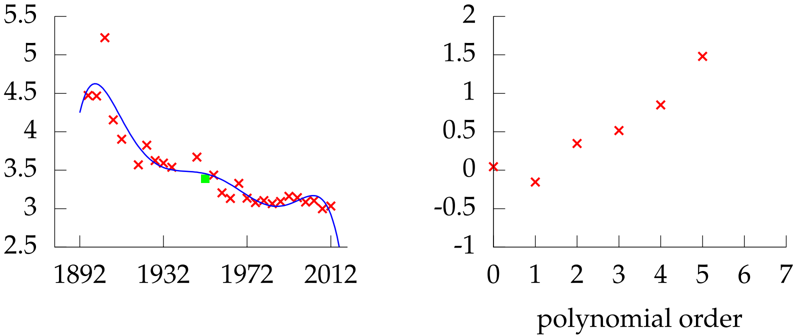
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



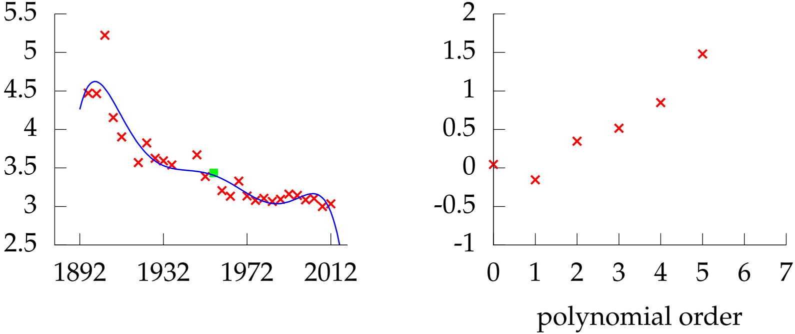
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



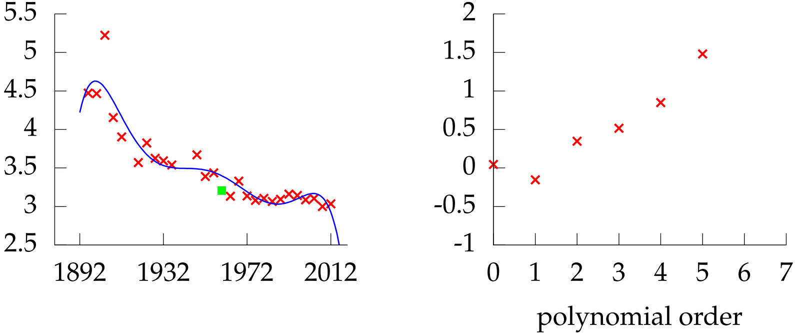
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



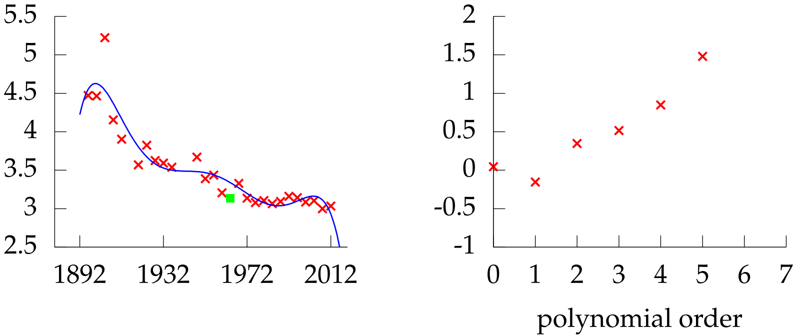
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



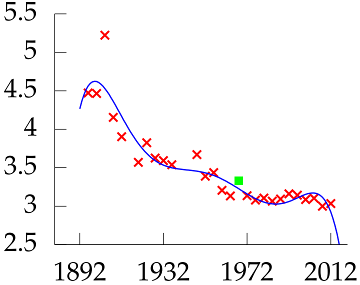
Polynomial order 6, training error -32.237, leave one out error 1.5047.

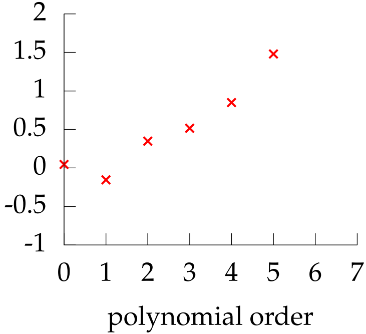
Leave One Out Error



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

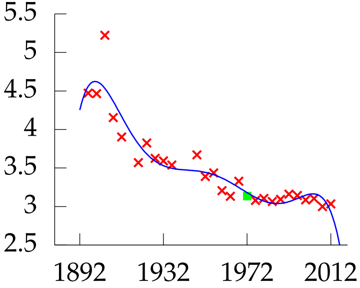
Leave One Out Error

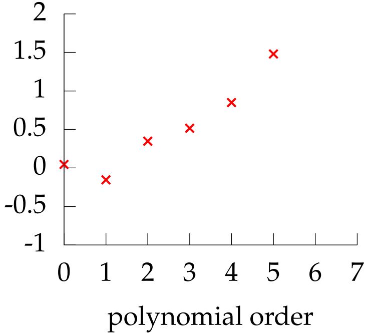




Polynomial order 6, training error -32.237, leave one out error 1.5047.

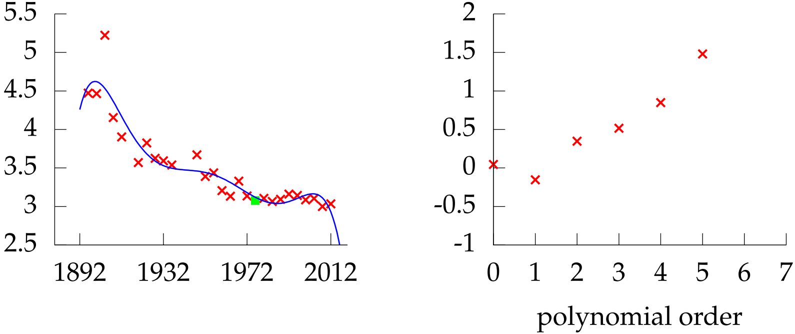
Leave One Out Error





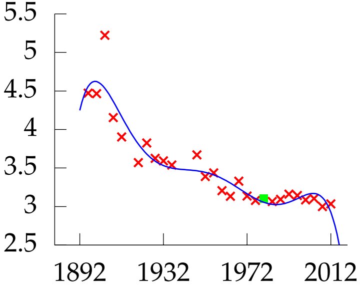
Polynomial order 6, training error -32.237, leave one out error 1.5047.

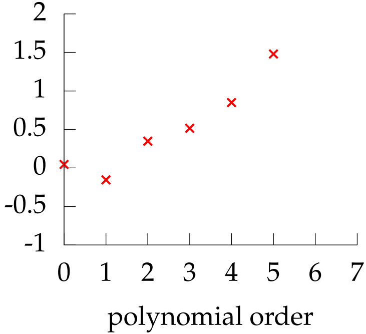
Leave One Out Error



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

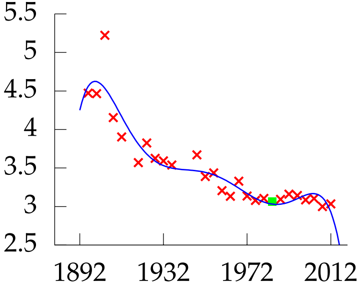
Leave One Out Error

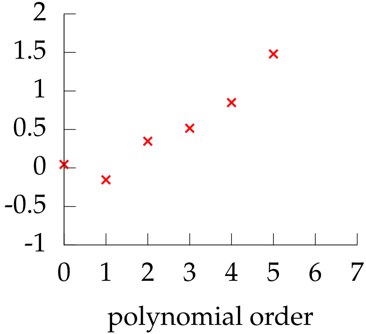




Polynomial order 6, training error -32.237, leave one out error 1.5047.

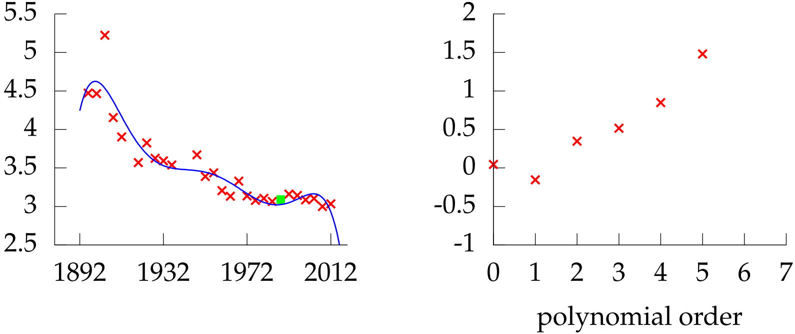
Leave One Out Error





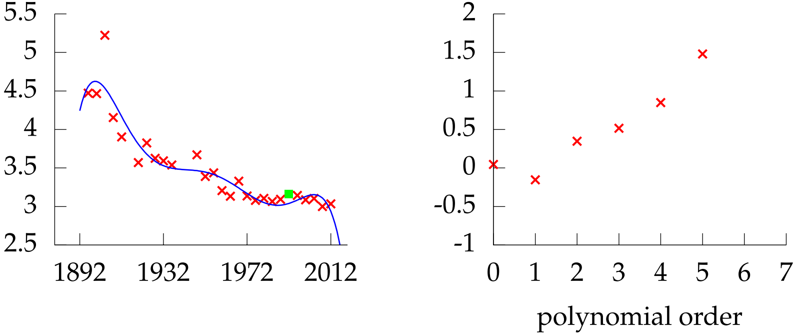
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



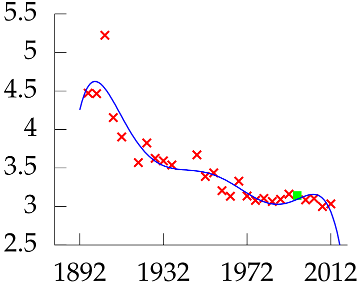
Polynomial order 6, training error -32.237, leave one out error 1.5047.

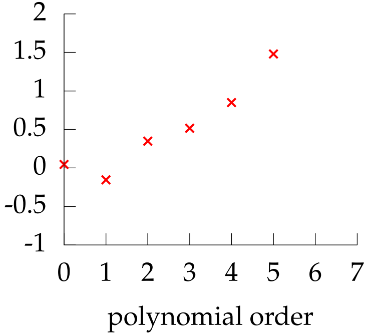
Leave One Out Error



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

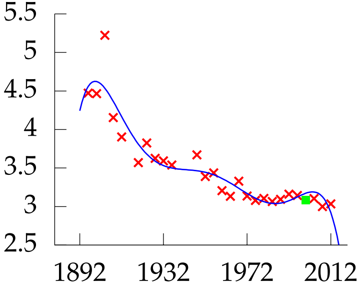
Leave One Out Error

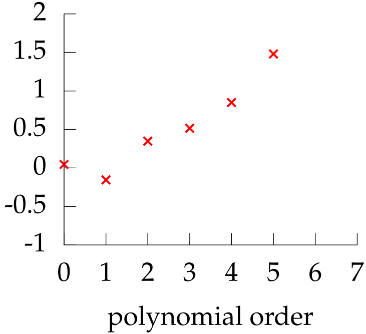




Polynomial order 6, training error -32.237, leave one out error 1.5047.

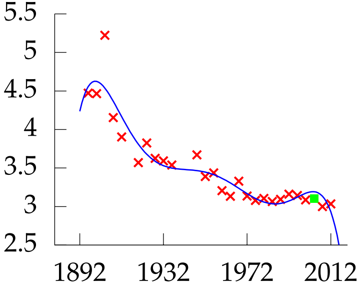
Leave One Out Error

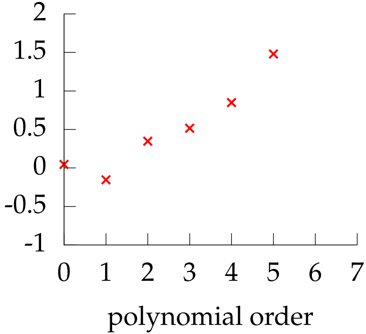




Polynomial order 6, training error -32.237, leave one out error 1.5047.

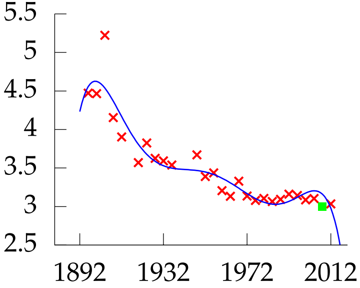
Leave One Out Error

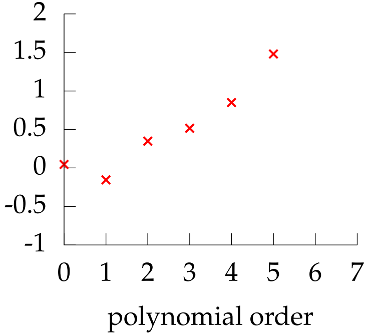




Polynomial order 6, training error -32.237, leave one out error 1.5047.

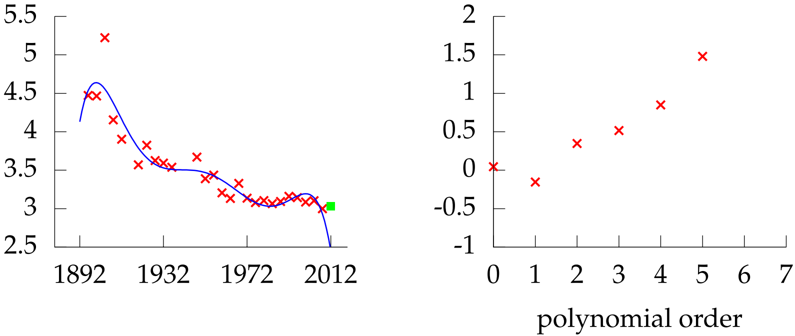
Leave One Out Error





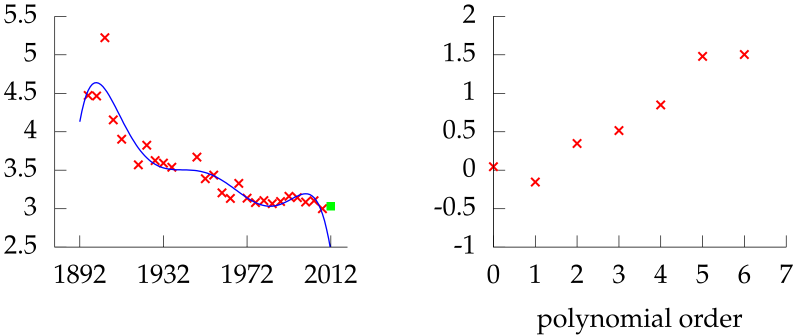
Polynomial order 6, training error -32.237, leave one out error 1.5047.

Leave One Out Error



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

Leave One Out Error

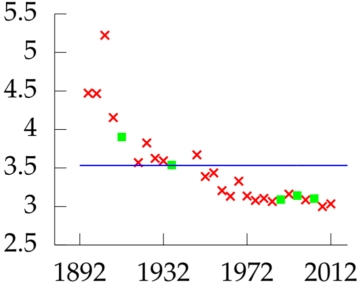


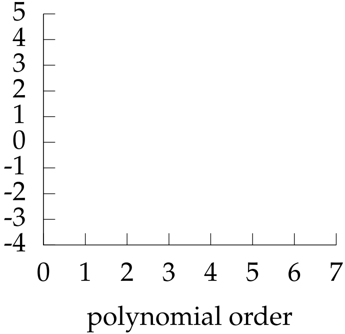
Polynomial order 6, training error -32.237, leave one out error 1.5047.

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.

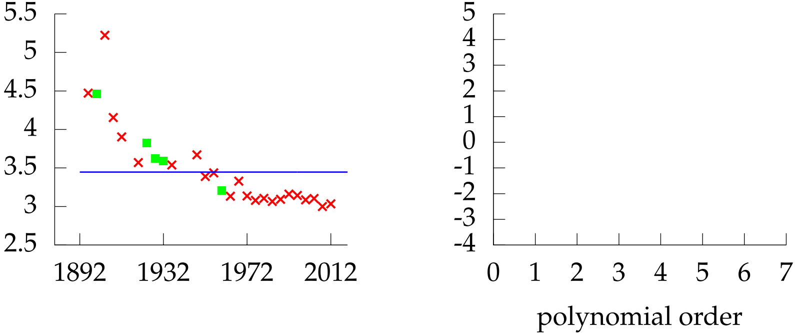
Cross Validation Error





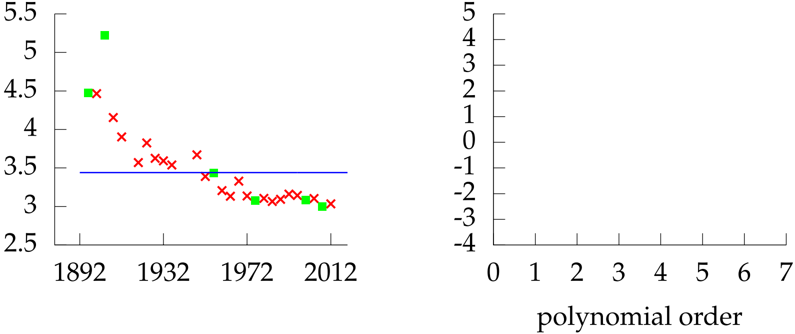
Polynomial order 0, training error -3.2644, leave one out error 0.045811.

Cross Validation Error



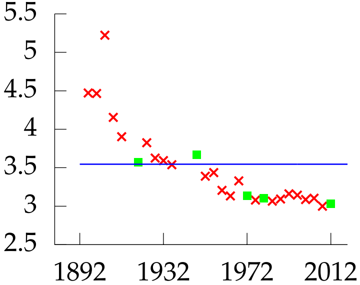
Polynomial order 0, training error -3.2644, leave one out error 0.045811.

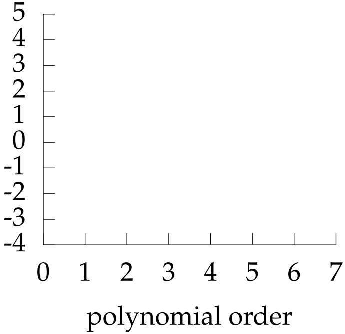
Cross Validation Error



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

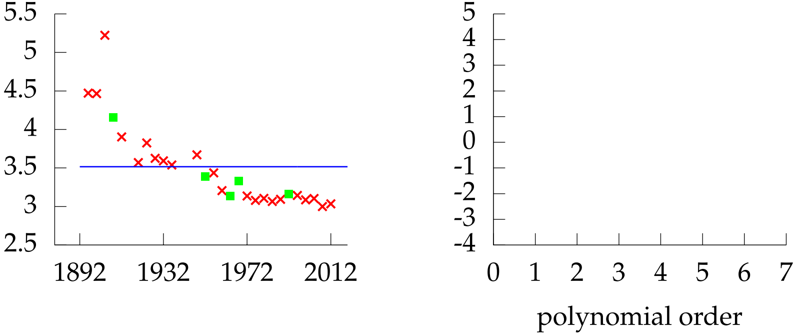
Cross Validation Error





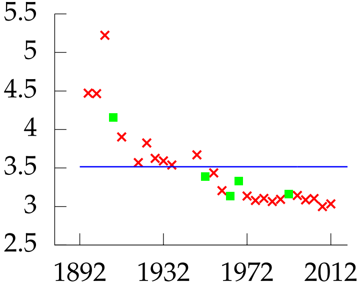
Polynomial order 0, training error -3.2644, leave one out error 0.045811.

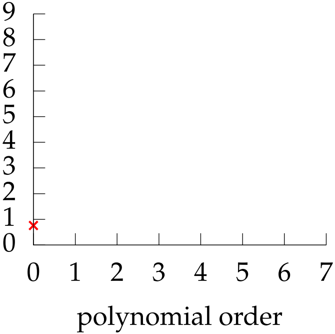
Cross Validation Error



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

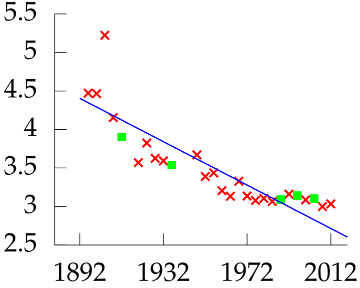
Cross Validation Error

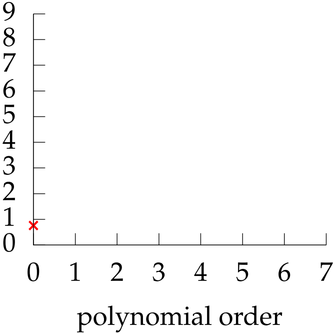




Polynomial order 0, training error -3.2644, leave one out error 0.045811.

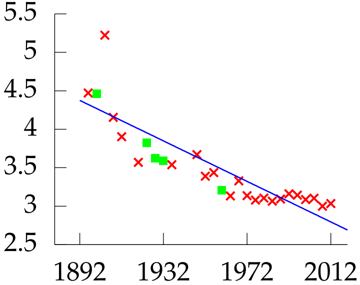
Cross Validation Error

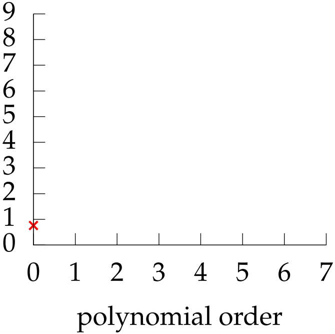




Polynomial order 1, training error -18.873, leave one out error -0.15413.

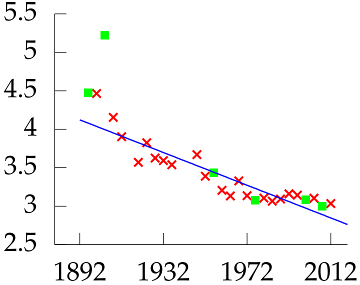
Cross Validation Error

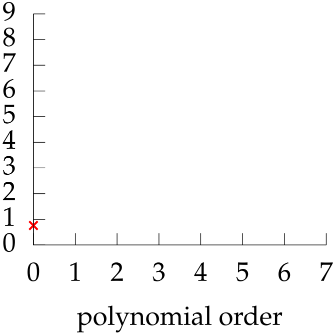




Polynomial order 1, training error -18.873, leave one out error -0.15413.

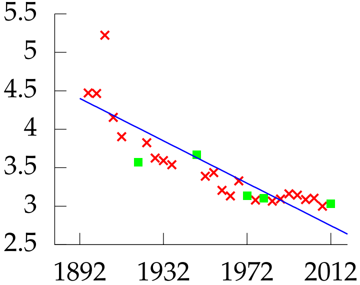
Cross Validation Error

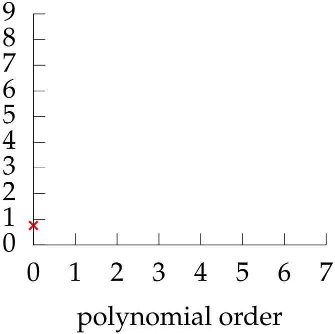




Polynomial order 1, training error -18.873, leave one out error -0.15413.

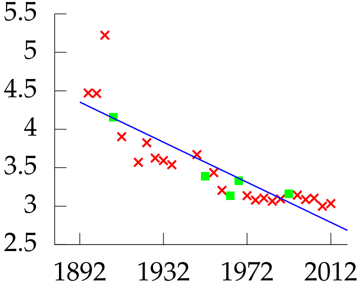
Cross Validation Error

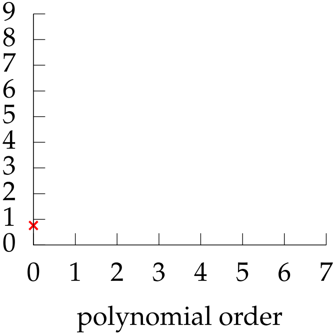




Polynomial order 1, training error -18.873, leave one out error -0.15413.

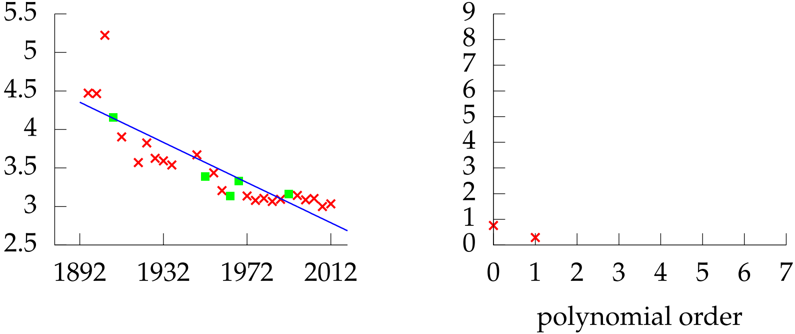
Cross Validation Error





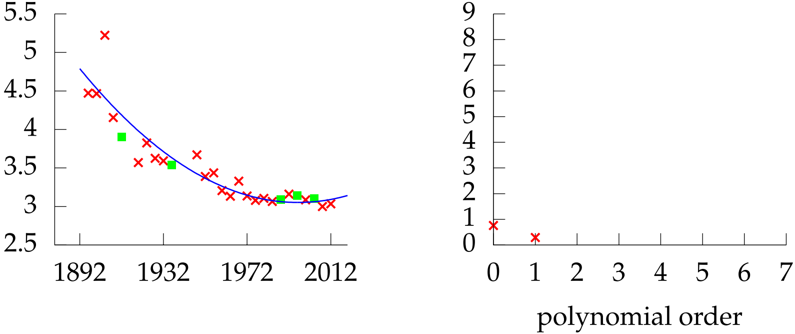
Polynomial order 1, training error -18.873, leave one out error -0.15413.

Cross Validation Error



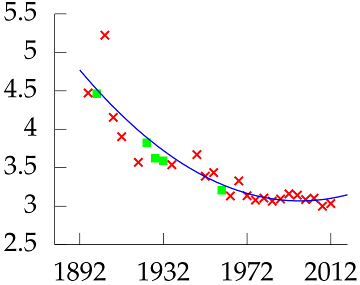
Polynomial order 1, training error -18.873, leave one out error -0.15413.

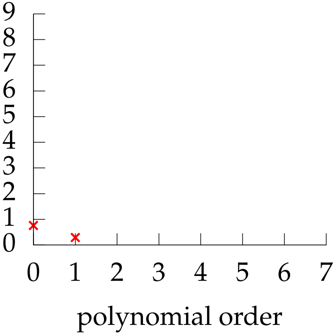
Cross Validation Error



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

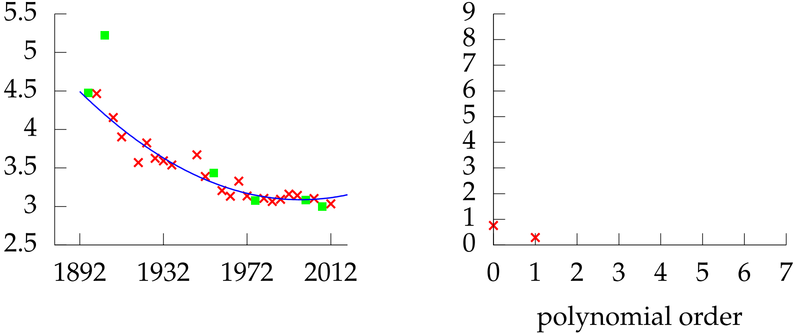
Cross Validation Error





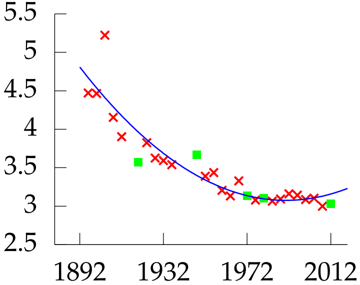
Polynomial order 2, training error -25.177, leave one out error 0.34669.

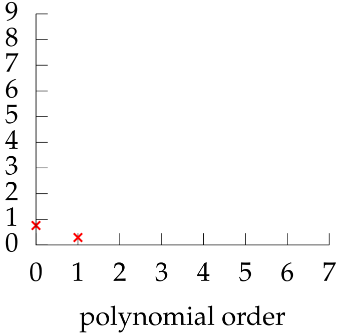
Cross Validation Error



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

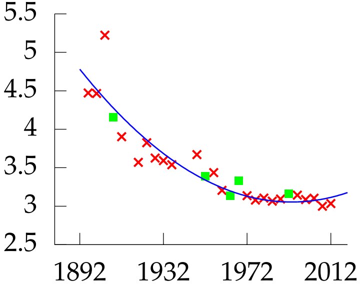
Cross Validation Error

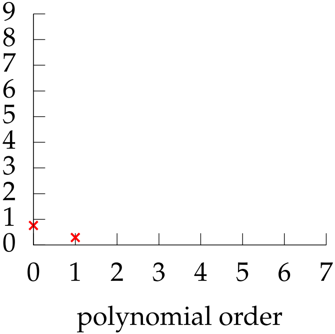




Polynomial order 2, training error -25.177, leave one out error 0.34669.

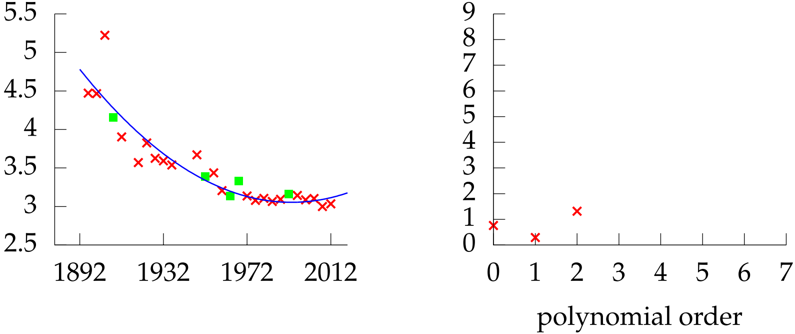
Cross Validation Error





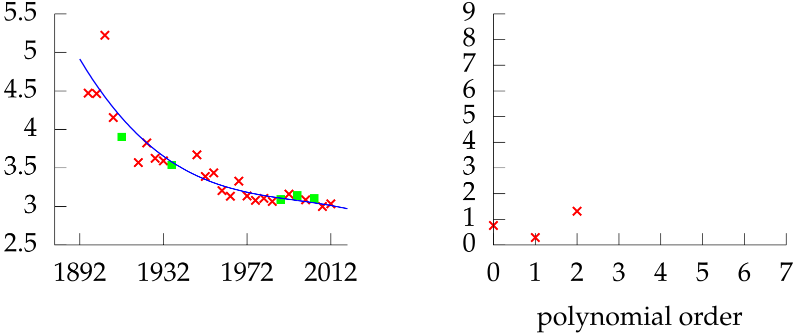
Polynomial order 2, training error -25.177, leave one out error 0.34669.

Cross Validation Error



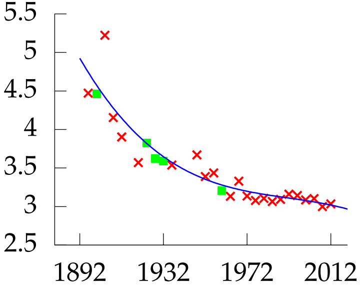
Polynomial order 2, training error -25.177, leave one out error 0.34669.

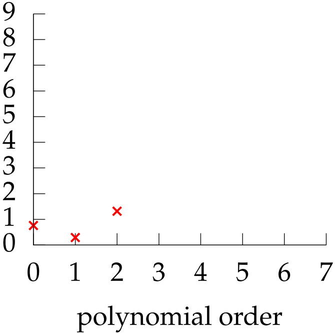
Cross Validation Error



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

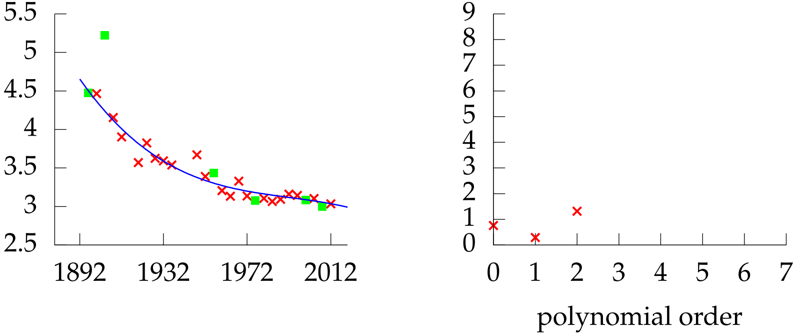
Cross Validation Error





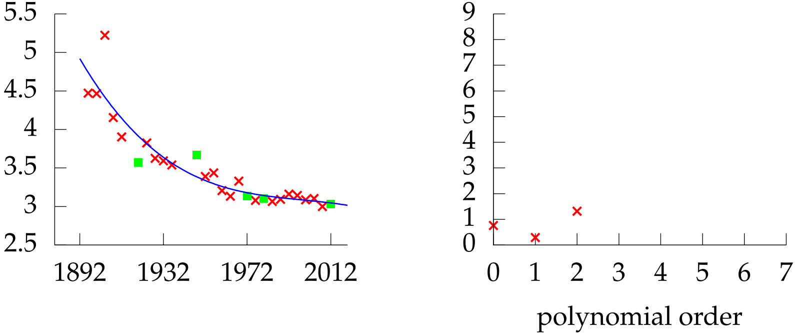
Polynomial order 3, training error -25.777, leave one out error 0.51621.

Cross Validation Error



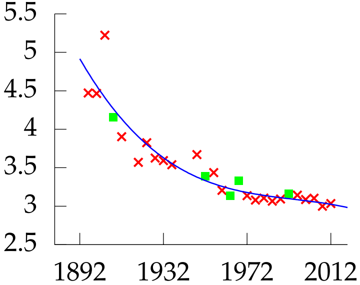
Polynomial order 3, training error -25.777, leave one out error 0.51621.

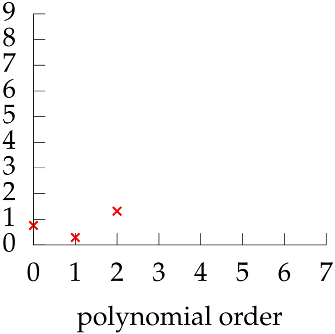
Cross Validation Error



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

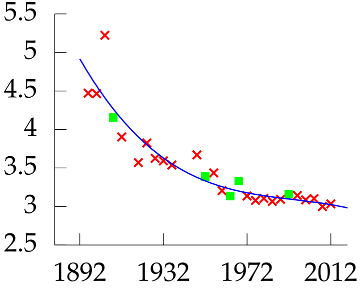
Cross Validation Error

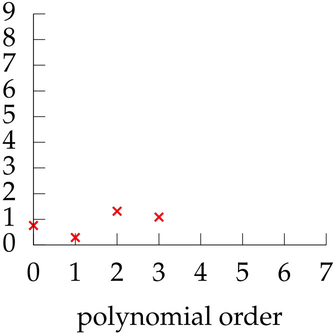




Polynomial order 3, training error -25.777, leave one out error 0.51621.

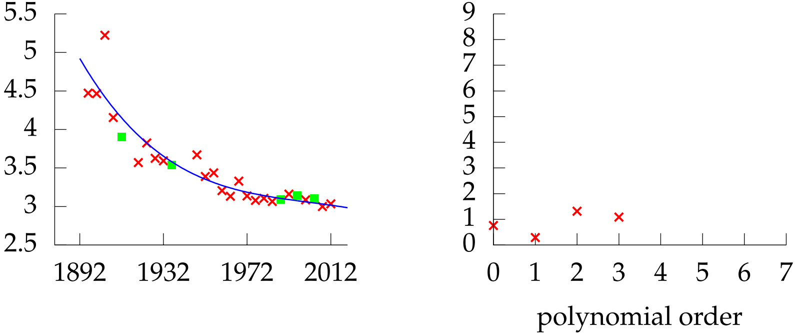
Cross Validation Error





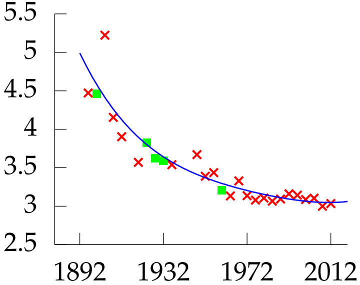
Polynomial order 3, training error -25.777, leave one out error 0.51621.

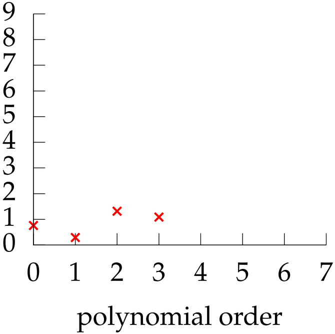
Cross Validation Error



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

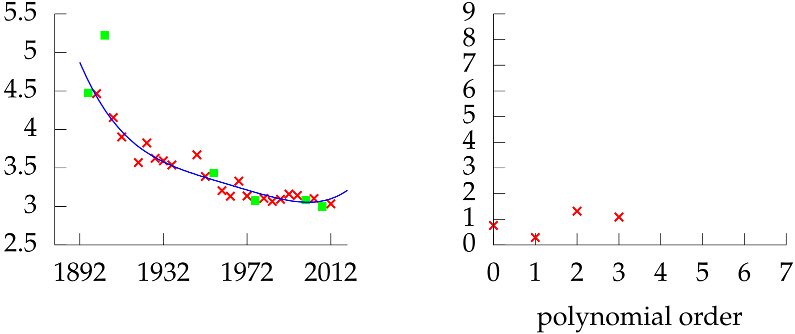
Cross Validation Error





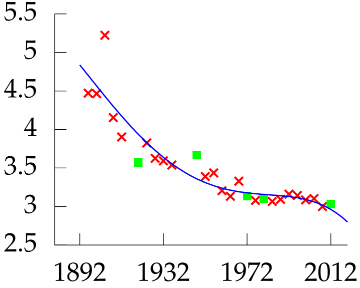
Polynomial order 4, training error -26.048, leave one out error 0.84844.

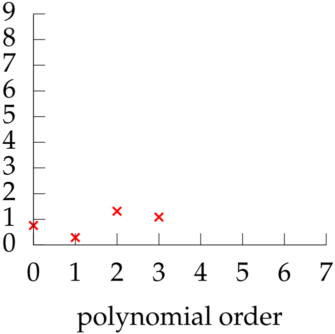
Cross Validation Error



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

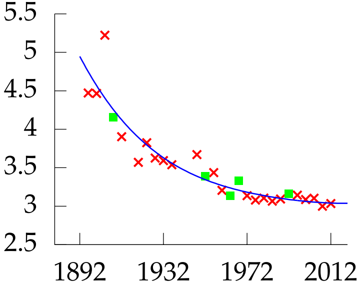
Cross Validation Error

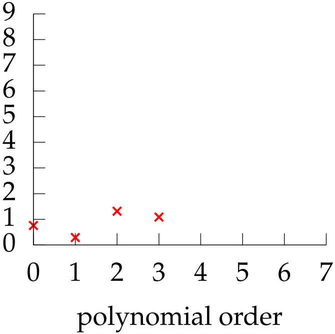




Polynomial order 4, training error -26.048, leave one out error 0.84844.

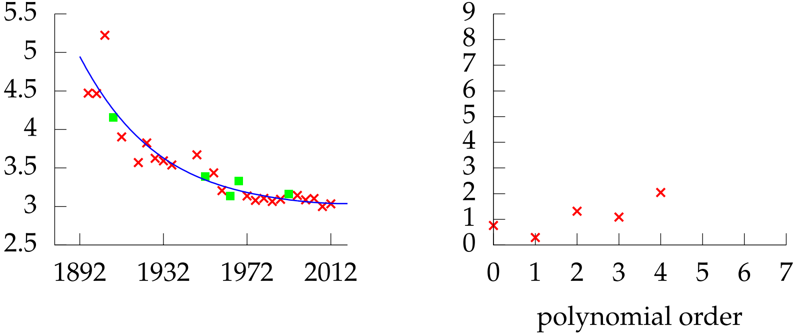
Cross Validation Error





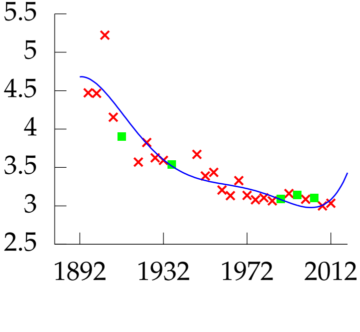
Polynomial order 4, training error -26.048, leave one out error 0.84844.

Cross Validation Error



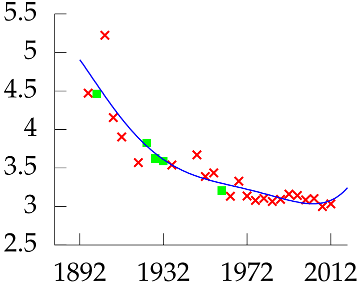
Polynomial order 4, training error -26.048, leave one out error 0.84844.

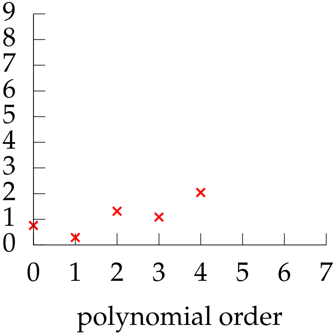
Cross Validation Error



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

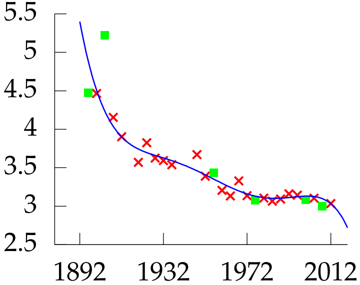
Cross Validation Error

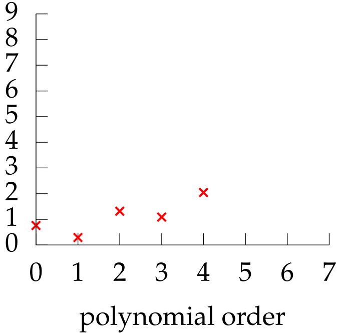




Polynomial order 5, training error -26.892, leave one out error 1.48.

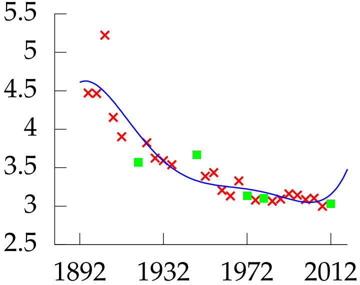
Cross Validation Error

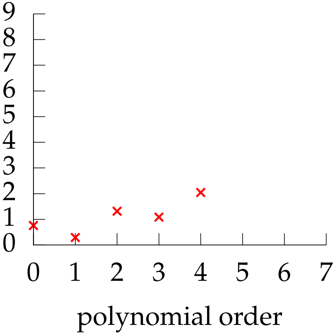




Polynomial order 5, training error -26.892, leave one out error 1.48.

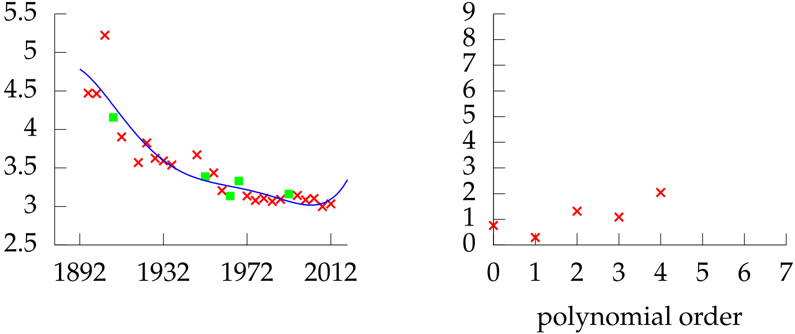
Cross Validation Error





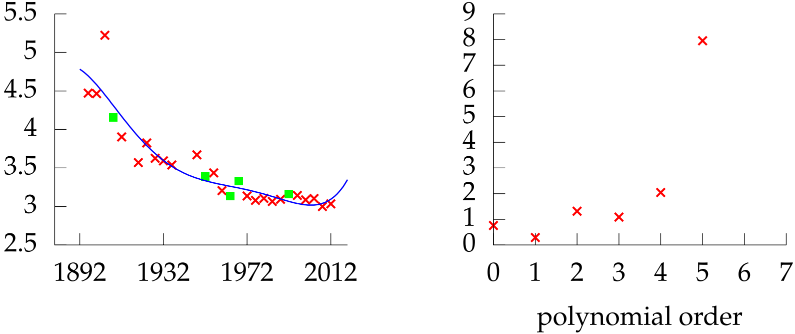
Polynomial order 5, training error -26.892, leave one out error 1.48.

Cross Validation Error



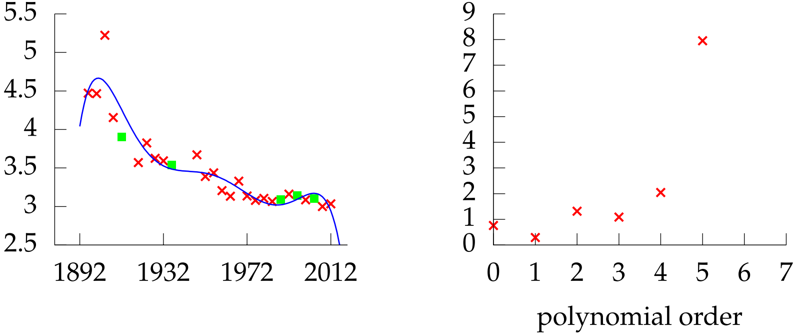
Polynomial order 5, training error -26.892, leave one out error 1.48.

Cross Validation Error



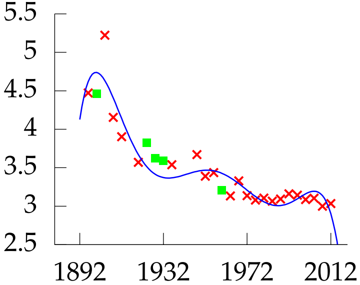
Polynomial order 5, training error -26.892, leave one out error 1.48.

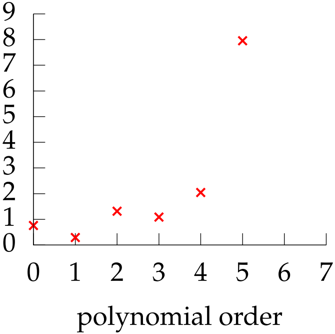
Cross Validation Error



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

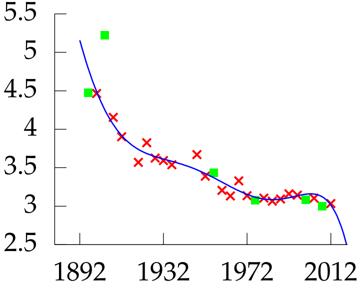
Cross Validation Error

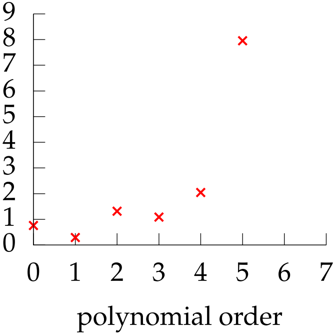




Polynomial order 6, training error -29.395, leave one out error 1.5047.

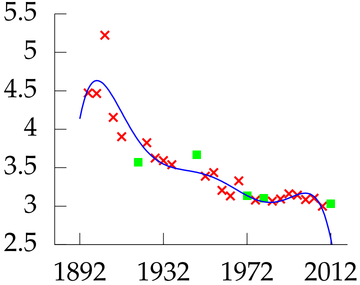
Cross Validation Error

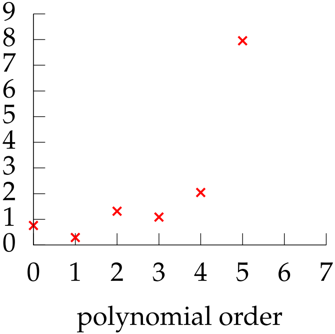




Polynomial order 6, training error -29.395, leave one out error 1.5047.

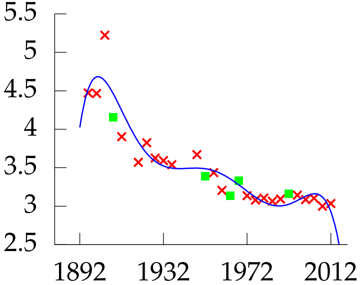
Cross Validation Error

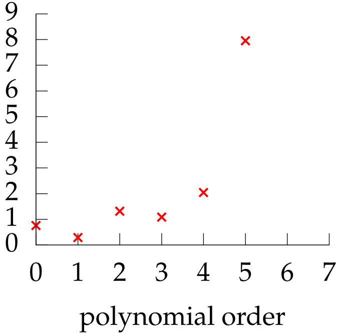




Polynomial order 6, training error -29.395, leave one out error 1.5047.

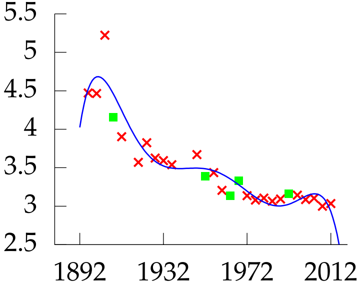
Cross Validation Error

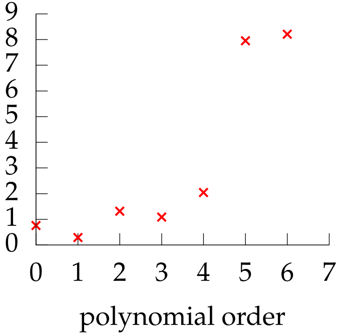




Polynomial order 6, training error -29.395, leave one out error 1.5047.

Cross Validation Error





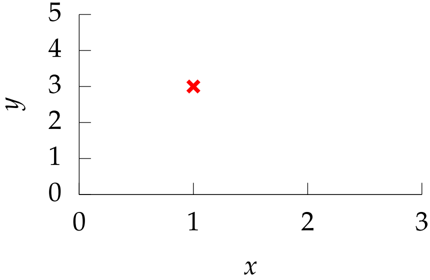
Polynomial order 6, training error -29.395, leave one out error 1.5047.

Outline

Underdetermined System

What about two unknowns and *one* observation?

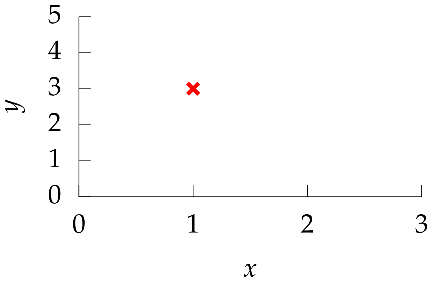
$$y_1 = mx_1 + c$$



Underdetermined System

Can compute m given c.

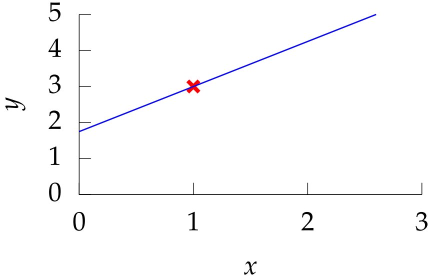
$$m = \frac{y_1 - c}{x}$$



Underdetermined System

Can compute m given c.

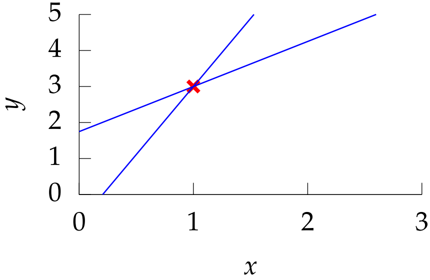
$$c = 1.75 \implies m = 1.25$$



Underdetermined System

Can compute m given c.

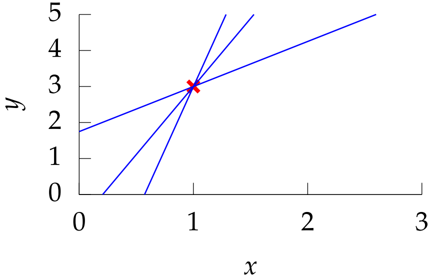
$$c = -0.777 \implies m = 3.78$$



Underdetermined System

Can compute m given c.

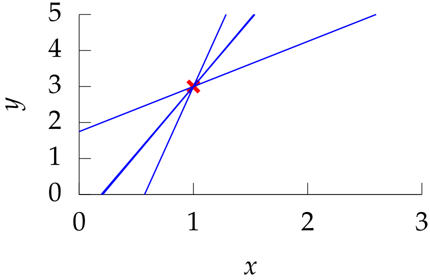
$$c = -4.01 \Longrightarrow m = 7.01$$



Underdetermined System

Can compute m given c.

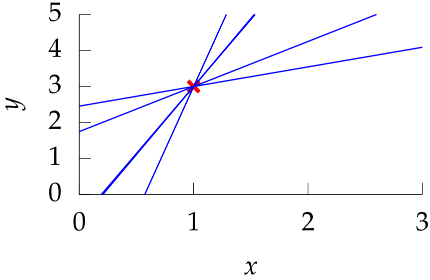
$$c = -0.718 \Longrightarrow m = 3.72$$



Underdetermined System

Can compute m given c.

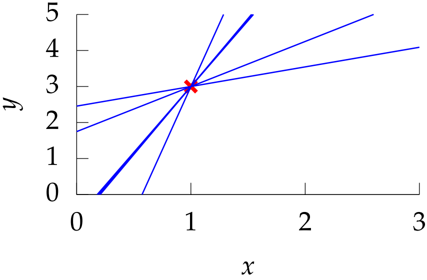
$$c = 2.45 \implies m = 0.545$$



Underdetermined System

Can compute m given c.

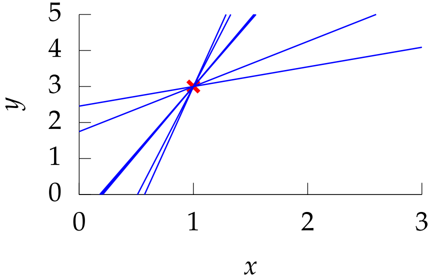
$$c = -0.657 \implies m = 3.66$$



Underdetermined System

Can compute m given c.

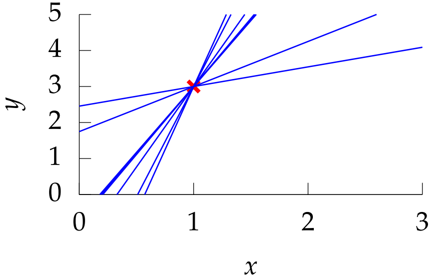
$$c = -3.13 \Longrightarrow m = 6.13$$



Underdetermined System

Can compute m given c.

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

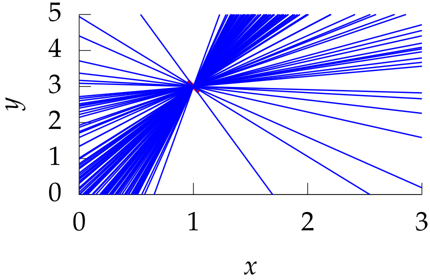


Underdetermined System

Can compute m given c. Assume

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

we find a distribution of solutions.



Probability for Under- and Overdetermined

- To deal with overdetermined introduced probability distribution for 'variable', ϵ_i.
- For underdetermined system introduced probability distribution for 'parameter', c.
- This is known as a Bayesian treatment.

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.

Outline

Basis Functions

Underdetermined Systems

Bayesian Regression

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:

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

Gaussian Noise

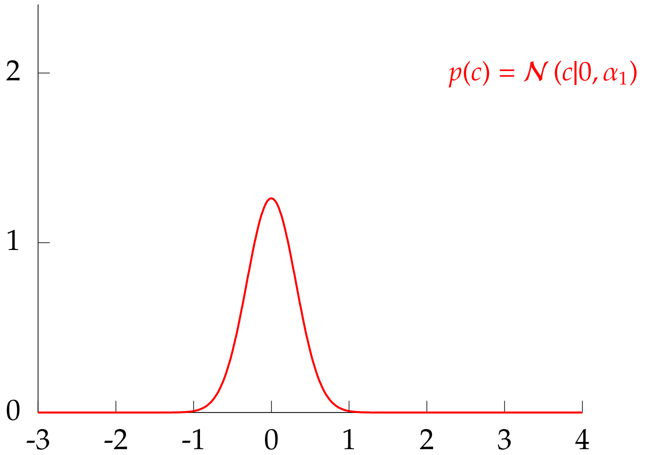


Figure: A Gaussian prior combines with a Gaussian likelihood for a Gaussian posterior.

Gaussian Noise

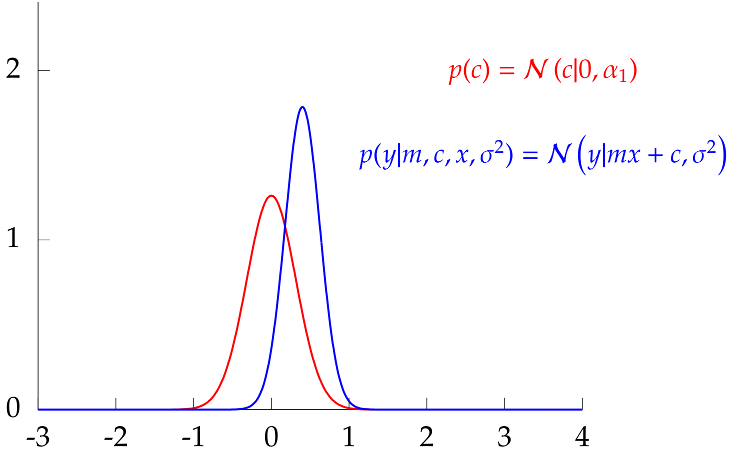


Figure: A Gaussian prior combines with a Gaussian likelihood for a Gaussian posterior.

Gaussian Noise

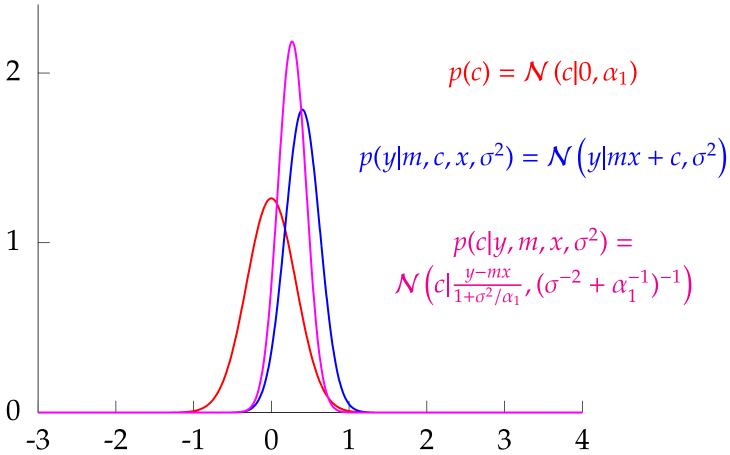


Figure: A Gaussian prior combines with a Gaussian likelihood for a Gaussian posterior.

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.

Multivariate Prior Distributions

- For general Bayesian inference need multivariate priors.
- E.g. for multivariate linear regression:

$$y_i = \sum_i w_j x_{i,j} + \epsilon_i$$

  (where we've dropped c for convenience), we need a prior over $\mathbf{w}$.
- This motivates a *multivariate* Gaussian density.
- We will use the multivariate Gaussian to put a prior *directly* on the function (a Gaussian process).

Multivariate Prior Distributions

- For general Bayesian inference need multivariate priors.
- E.g. for multivariate linear regression:

$$y_i = \mathbf{w}^\top \mathbf{x}_{i,:} + \epsilon_i$$

  (where we've dropped c for convenience), we need a prior over $\mathbf{w}$.
- This motivates a *multivariate* Gaussian density.
- We will use the multivariate Gaussian to put a prior *directly* on the function (a Gaussian process).

Two Dimensional Gaussian

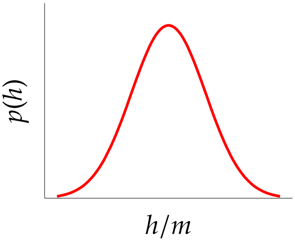
- Consider height, h/m and weight, w/kg.
- Could sample height from a distribution:

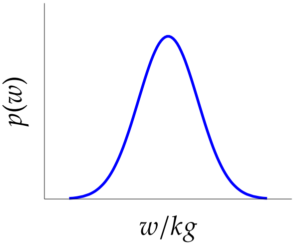
$$p(h) \sim \mathcal{N}(1.7, 0.0225)$$

- And similarly weight:

$$p(w) \sim \mathcal{N}(75, 36)$$

Height and Weight Models



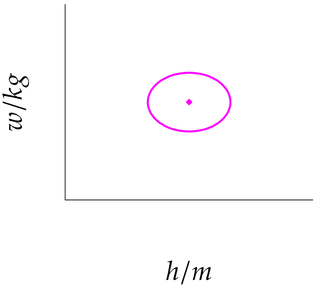


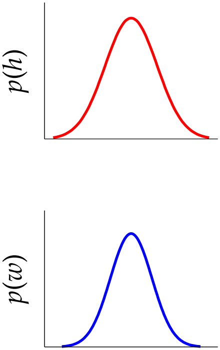
Gaussian distributions for height and weight.

Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution



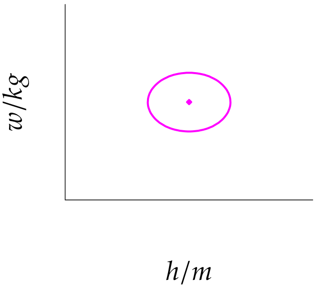


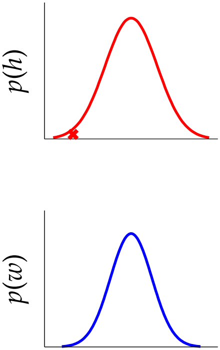
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution



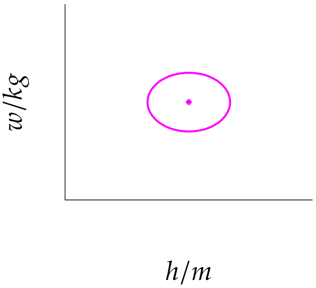


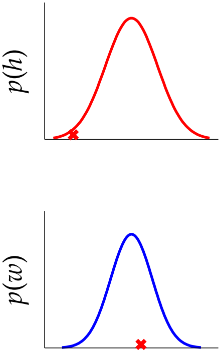
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

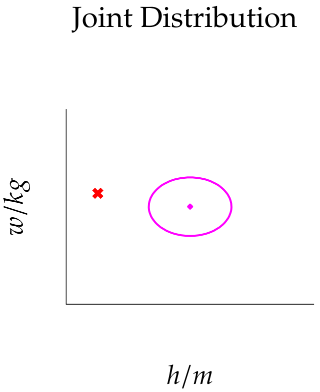
Joint Distribution

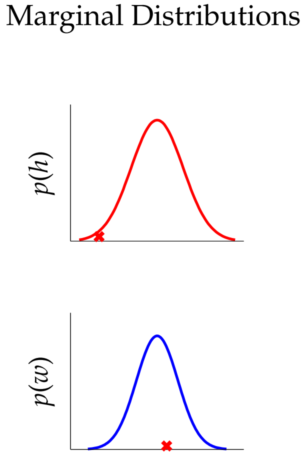




Samples of height and weight

Sampling Two Dimensional Variables

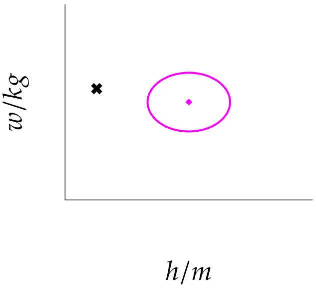




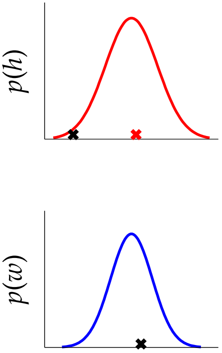
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution



Marginal Distributions

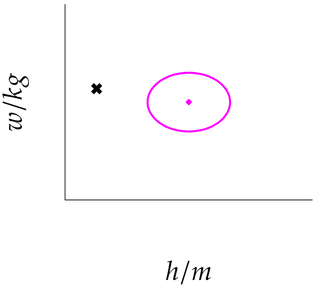


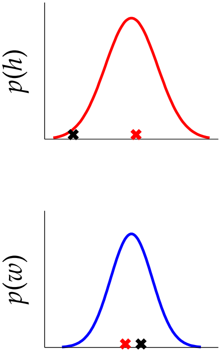
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution

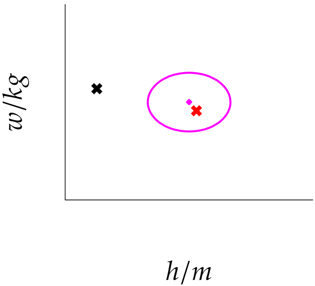




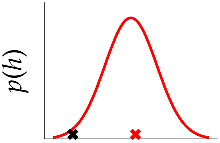
Samples of height and weight

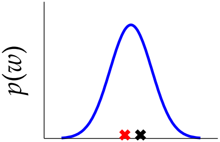
Sampling Two Dimensional Variables

Joint Distribution



Marginal Distributions

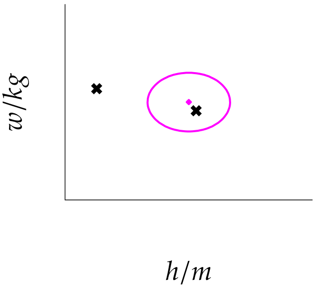




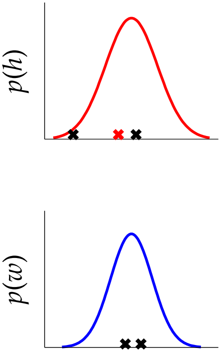
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution

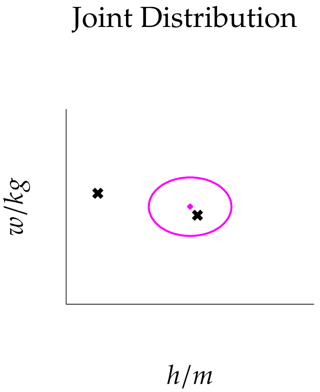


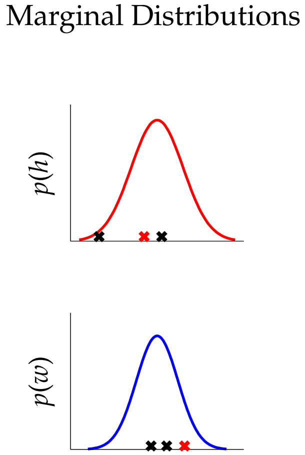
Marginal Distributions



Samples of height and weight

Sampling Two Dimensional Variables



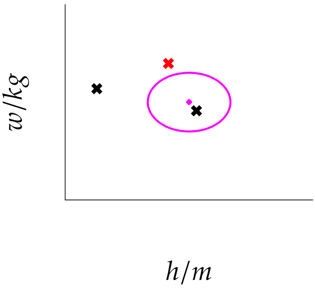


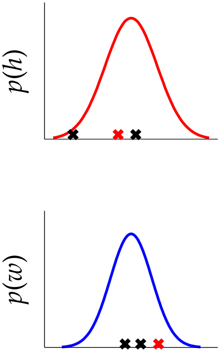
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

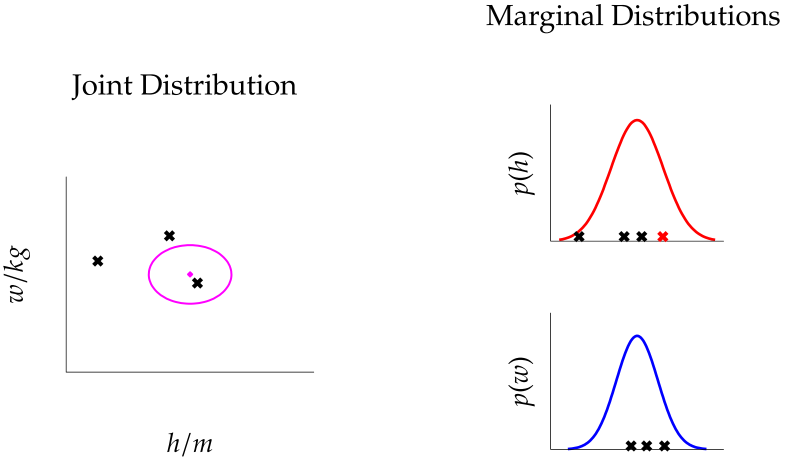
Joint Distribution





Samples of height and weight

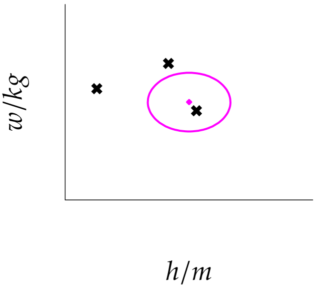
Sampling Two Dimensional Variables



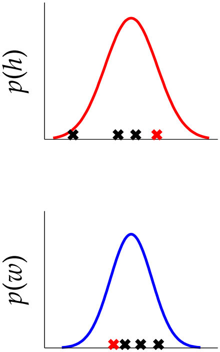
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution



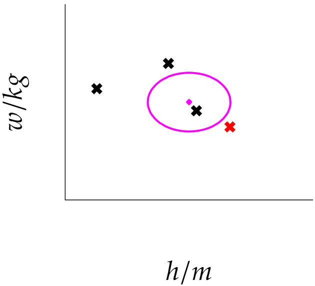
Marginal Distributions



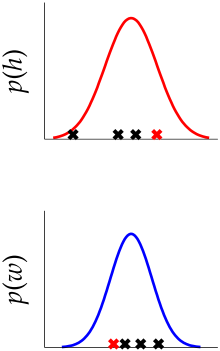
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution



Marginal Distributions

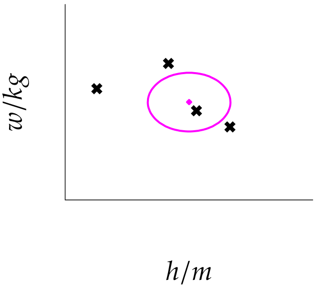


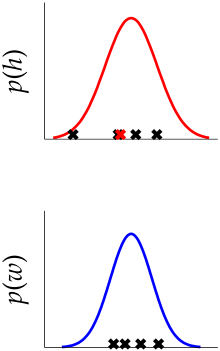
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution



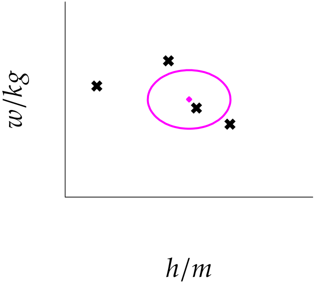


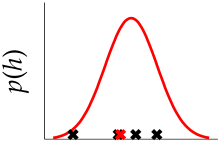
Samples of height and weight

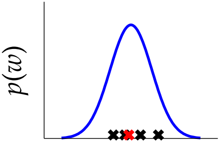
Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution

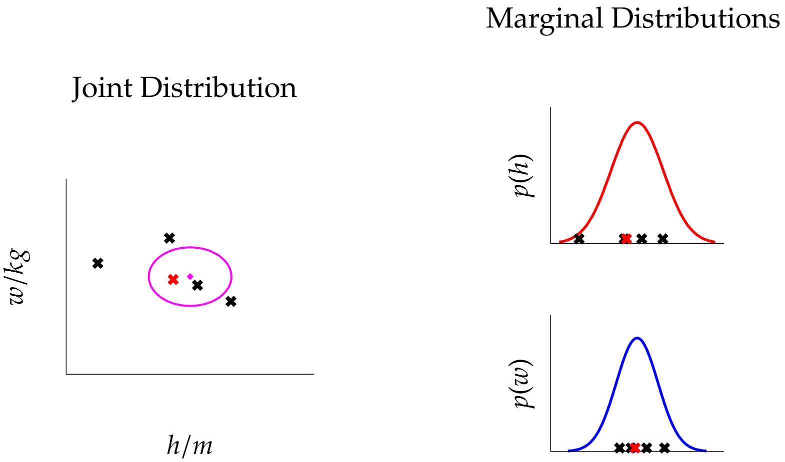






Samples of height and weight

Sampling Two Dimensional Variables

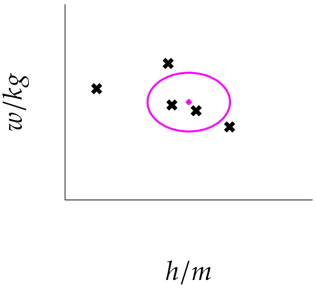


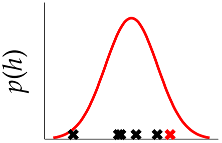
Samples of height and weight

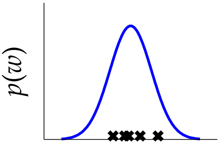
Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution





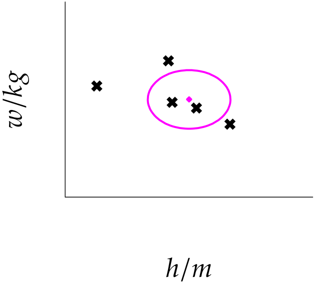


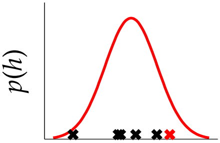
Samples of height and weight

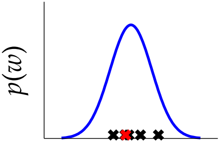
Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution





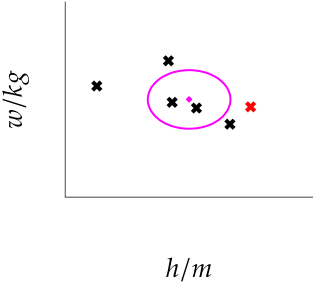


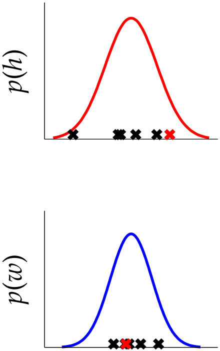
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

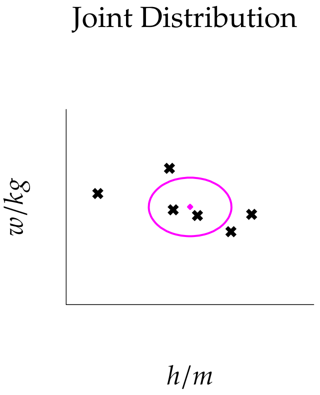
Joint Distribution

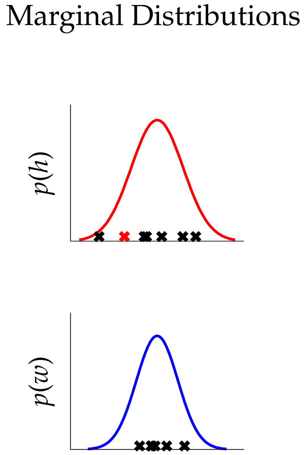




Samples of height and weight

Sampling Two Dimensional Variables



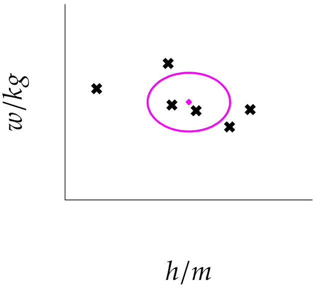


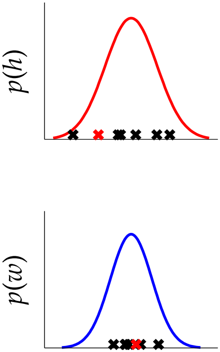
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

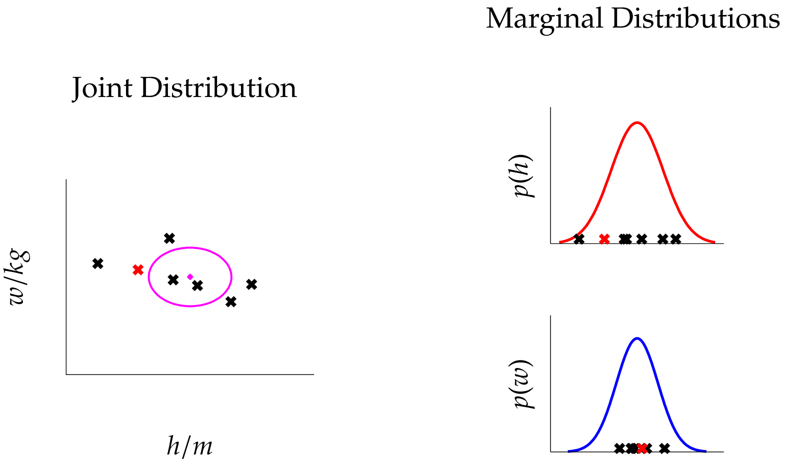
Joint Distribution





Samples of height and weight

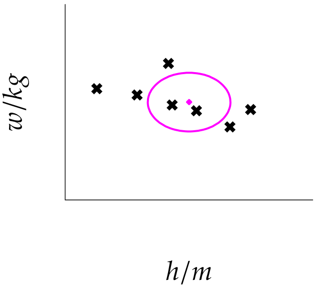
Sampling Two Dimensional Variables



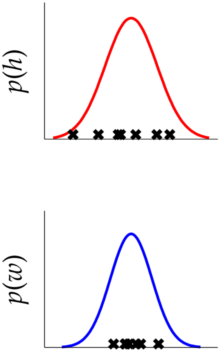
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution



Marginal Distributions



Samples of height and weight

Independence Assumption

- This assumes height and weight are independent.

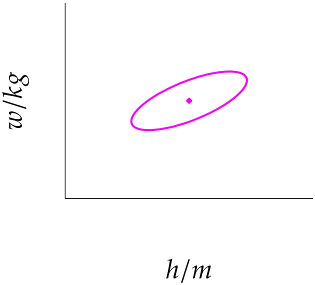
$$p(h, w) = p(h)p(w)$$

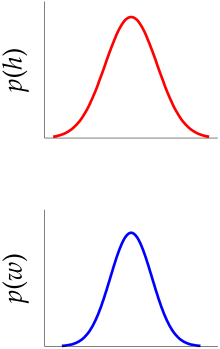
- In reality they are dependent (body mass index) $= \frac{w}{h^2}$.

Sampling Two Dimensional Variables

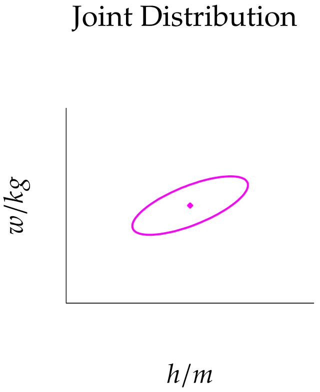
Marginal Distributions

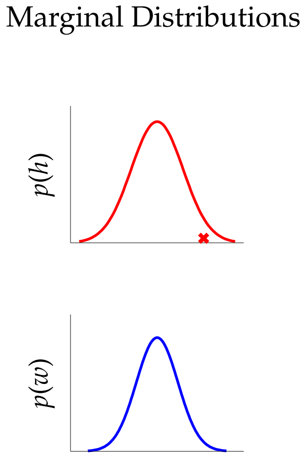
Joint Distribution





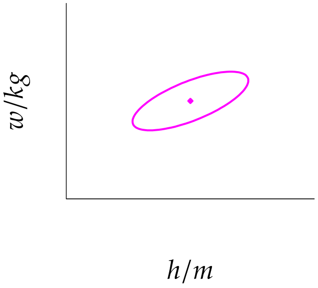
Sampling Two Dimensional Variables



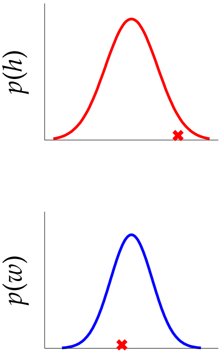


Sampling Two Dimensional Variables

Joint Distribution



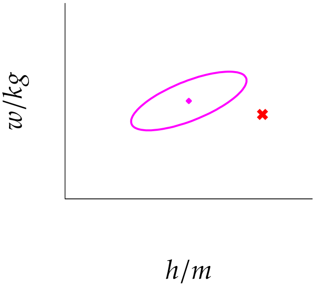
Marginal Distributions

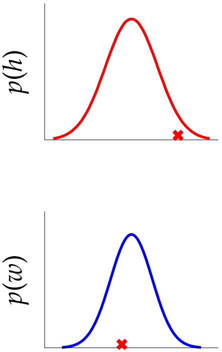


Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution

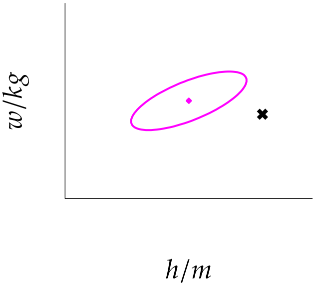


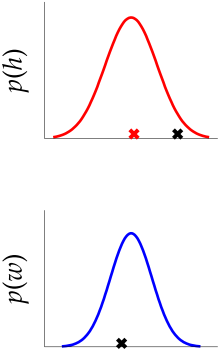


Sampling Two Dimensional Variables

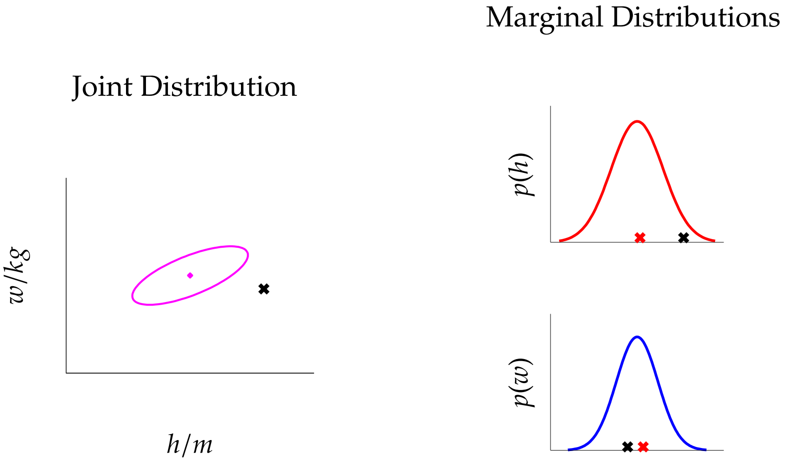
Marginal Distributions

Joint Distribution

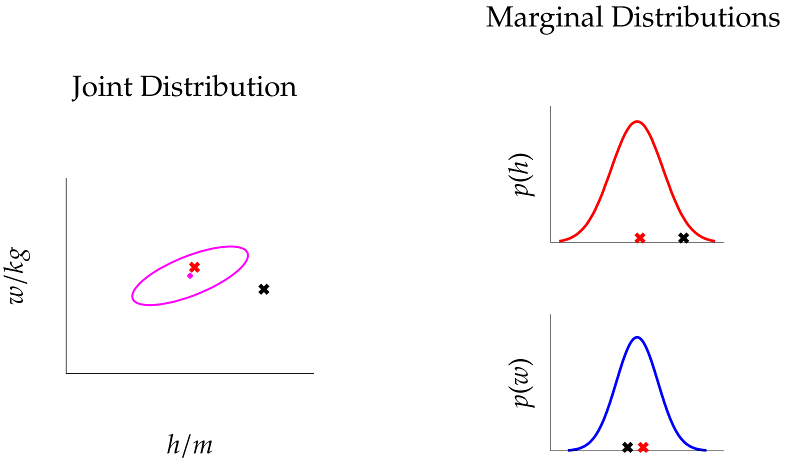




Sampling Two Dimensional Variables



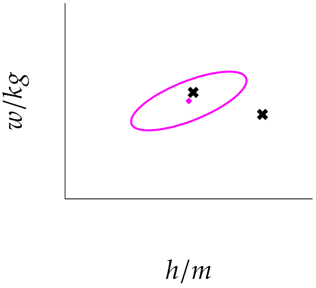
Sampling Two Dimensional Variables

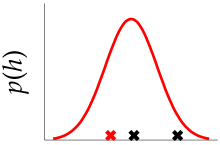


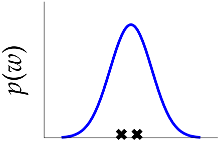
Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution



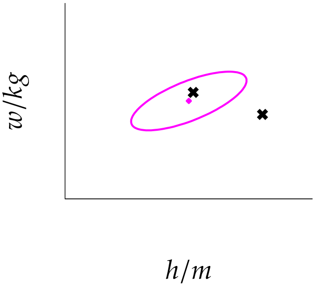


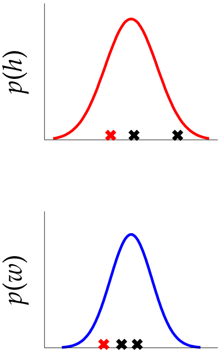


Sampling Two Dimensional Variables

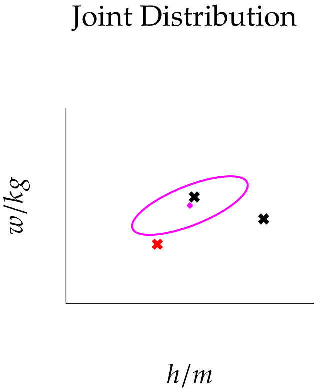
Marginal Distributions

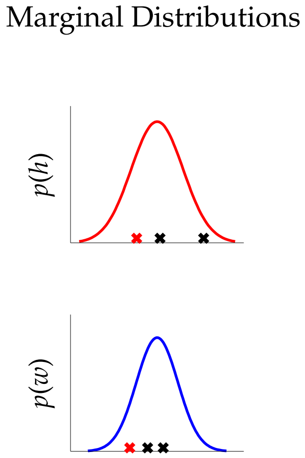
Joint Distribution





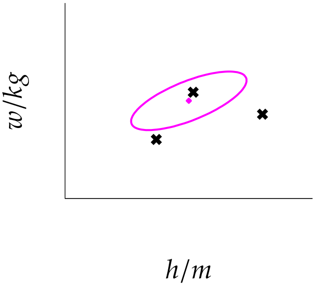
Sampling Two Dimensional Variables



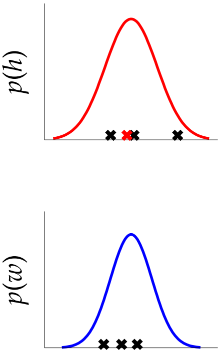


Sampling Two Dimensional Variables

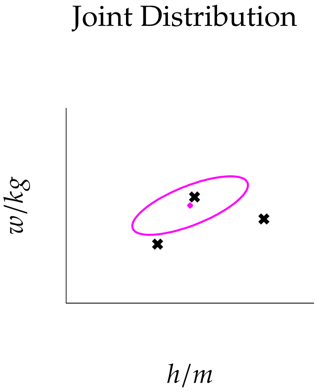
Joint Distribution

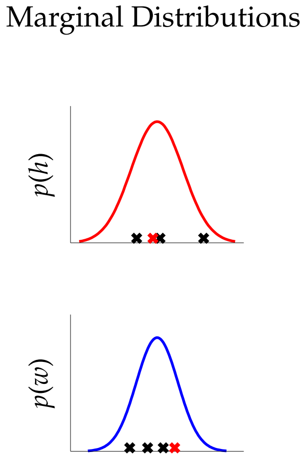


Marginal Distributions



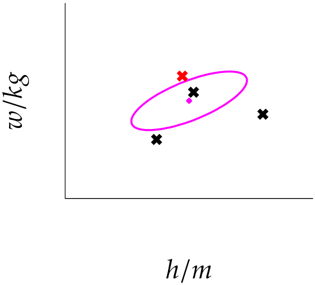
Sampling Two Dimensional Variables



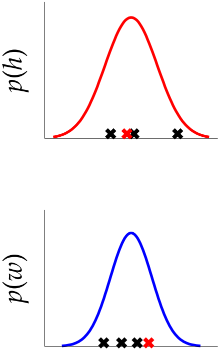


Sampling Two Dimensional Variables

Joint Distribution



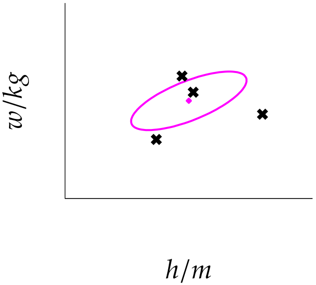
Marginal Distributions

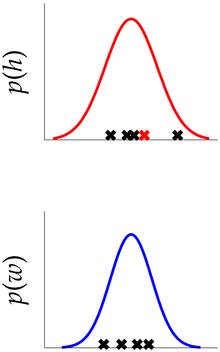


Sampling Two Dimensional Variables

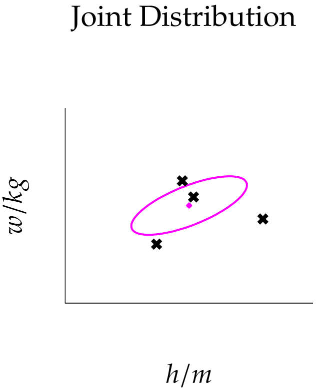
Marginal Distributions

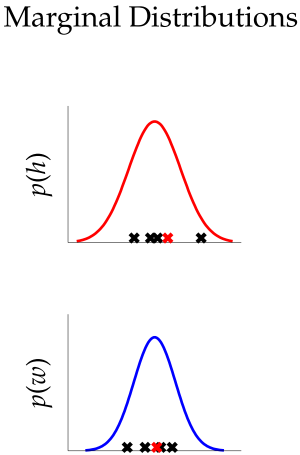
Joint Distribution



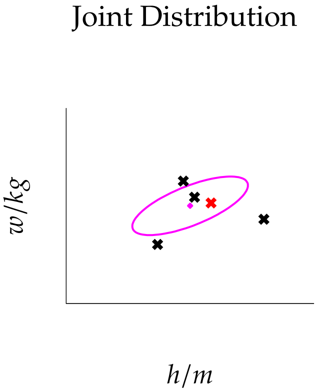


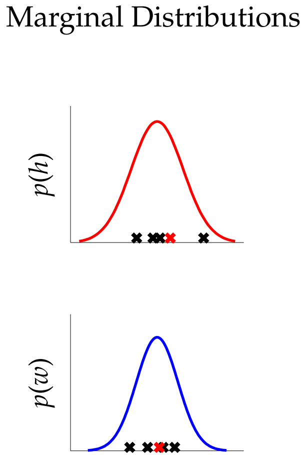
Sampling Two Dimensional Variables





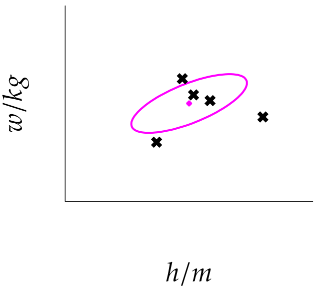
Sampling Two Dimensional Variables



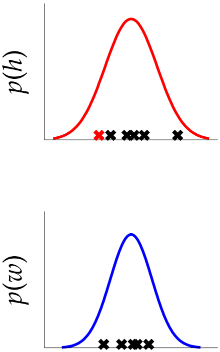


Sampling Two Dimensional Variables

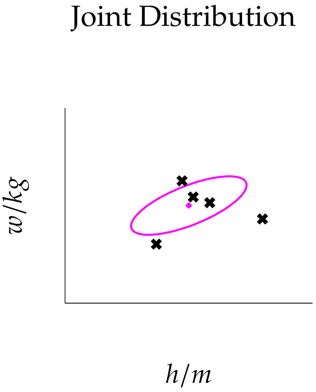
Joint Distribution

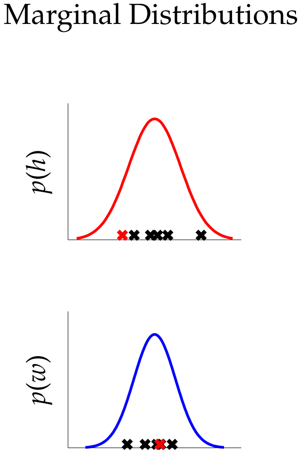


Marginal Distributions



Sampling Two Dimensional Variables

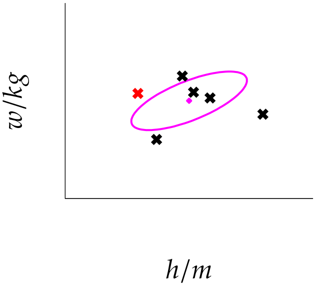


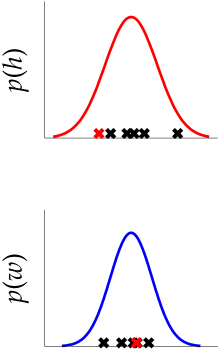


Sampling Two Dimensional Variables

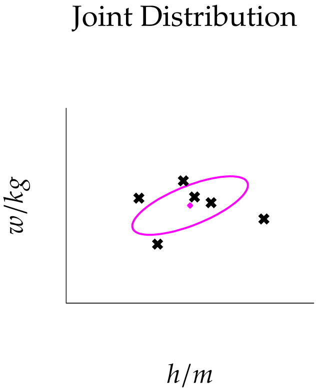
Marginal Distributions

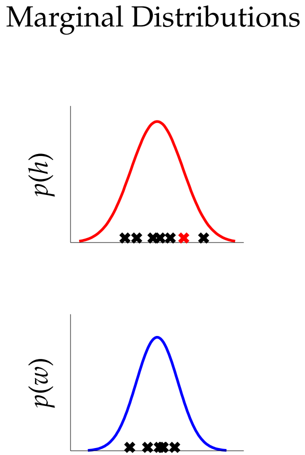
Joint Distribution





Sampling Two Dimensional Variables

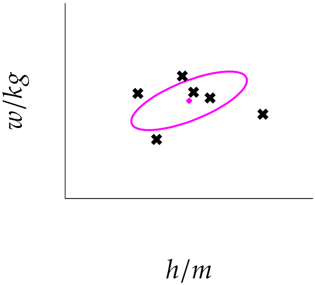


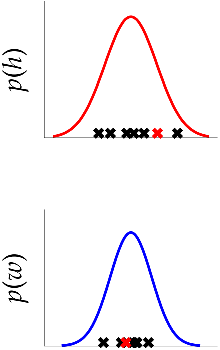


Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution

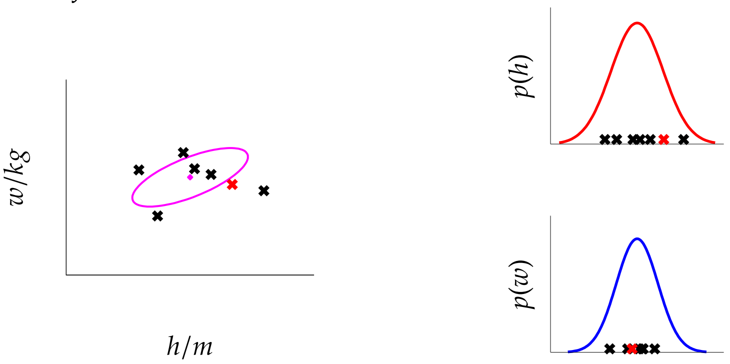




Sampling Two Dimensional Variables

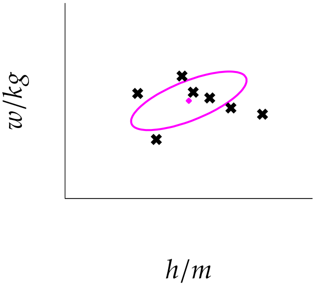
Marginal Distributions

Joint Distribution

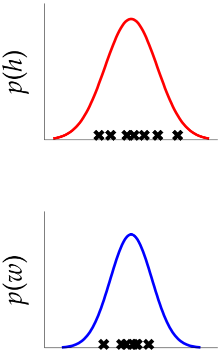


Sampling Two Dimensional Variables

Joint Distribution



Marginal Distributions



Independent Gaussians

$$p(w, h) = p(w)p(h)$$

Independent Gaussians

$$p(w,h) = \frac{1}{\sqrt{2\pi\sigma_1^2}\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\frac{(w-\mu_1)^2}{\sigma_1^2} + \frac{(h-\mu_2)^2}{\sigma_2^2}\right)\right)$$

Independent Gaussians

$$p(w, h) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} w \\ h \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)^\top \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} \left(\begin{bmatrix} w \\ h \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)\right)$$

Independent Gaussians

$$p(\mathbf{y}) = \frac{1}{2\pi\,|\mathbf{D}|}\exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \mathbf{D}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix $\mathbf{R}$.

$$p(\mathbf{y}) = \frac{1}{2\pi \left|\mathbf{D}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \mathbf{D}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix $\mathbf{R}$.

$$p(\mathbf{y}) = \frac{1}{2\pi \left|\mathbf{D}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{R}^{\top}\mathbf{y} - \mathbf{R}^{\top}\boldsymbol{\mu})^{\top}\mathbf{D}^{-1}(\mathbf{R}^{\top}\mathbf{y} - \mathbf{R}^{\top}\boldsymbol{\mu})\right)$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix $\mathbf{R}$.

$$p(\mathbf{y}) = \frac{1}{2\pi \left|\mathbf{D}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \mathbf{R}\mathbf{D}^{-1}\mathbf{R}^\top(\mathbf{y}-\boldsymbol{\mu})\right)$$

this gives a covariance matrix:

$$\mathbf{C}^{-1} = \mathbf{R}\mathbf{D}^{-1}\mathbf{R}^\top$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix **R**.

$$p(\mathbf{y}) = \frac{1}{2\pi \left|\mathbf{C}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^{\top}\mathbf{C}^{-1}(\mathbf{y} - \boldsymbol{\mu})\right)$$

this gives a covariance matrix:

$$\mathbf{C} = \mathbf{R}\mathbf{D}\mathbf{R}^{\top}$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Multivariate Consequence

- If

$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

Multivariate Consequence

- If

$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

- And

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

Multivariate Consequence

- If
$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$
- And
$$\mathbf{y} = \mathbf{W}\mathbf{x}$$
- Then
$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{W}\boldsymbol{\mu}, \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^\top\right)$$

References I

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.