Statistics in Cosmology

Day 3 – Model Selection

11th TRR33 Winter School in Cosmology
10-16 December 2017, Passo del Tonale - Italy

Emille E. O. Ishida

Laboratoire de Physique de Clermont - Université Clermont-Auvergne
Clermont Ferrand, France

Model Selection in Cosmology

$$\chi^2 \equiv -2 \ln \mathcal{L}.$$

$$\chi^2_{\rm tot} = \chi^2_{\rm CMB} + \chi^2_{\rm SN} + \chi^2_{\rm lens} + \chi^2_{\rm LSS} + \dots$$

Is the Universe flat?

Does Dark Energy evolve?

Is there evidence for modified gravity?

Model Selection

Information criteria

Goal: penalize models with more parameters

$$AIC = -2\ln\mathcal{L}|_{\hat{\theta}} + 2k$$

$$AICc = AIC + \frac{2k(k+1)}{N-k-1}$$

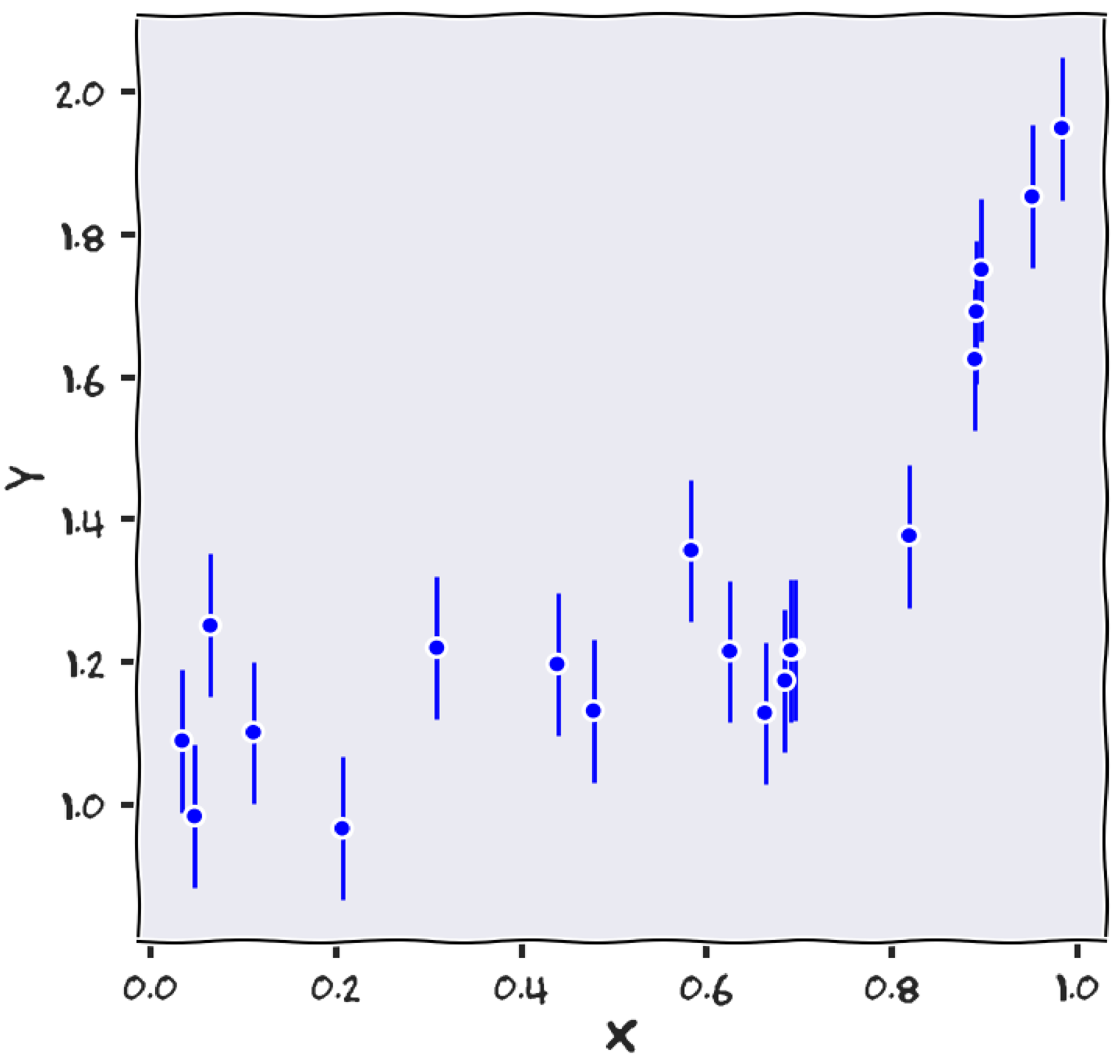
$$BIC = -2\ln\mathcal{L}|_{\hat{\theta}} + k\ln N$$

K → number of parameters
N → number of data points

These are only valid in very specific conditions

Model Selection

Information criteria

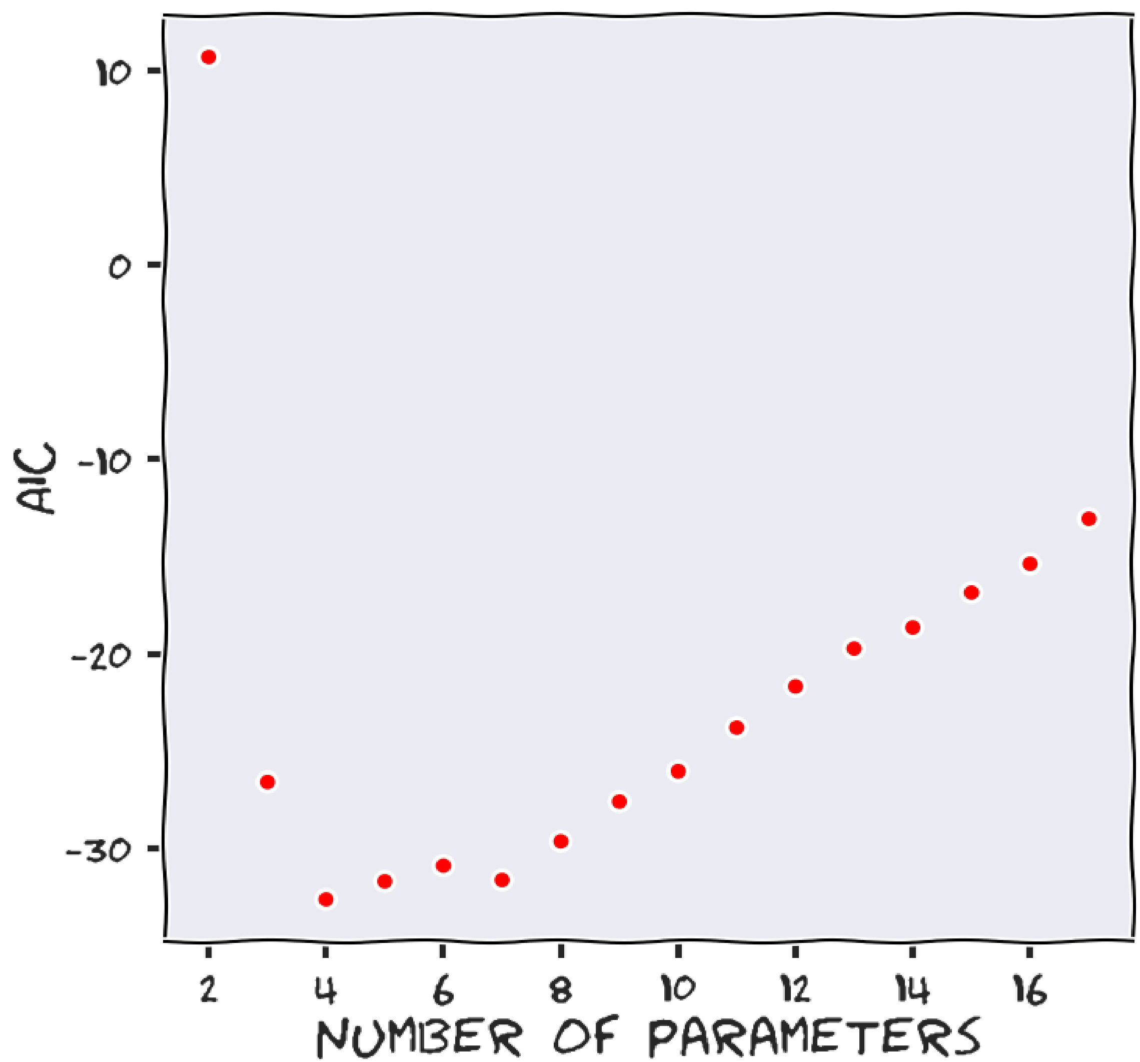


Which model better describe this data?

$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\dots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

Model Selection

Information criteria

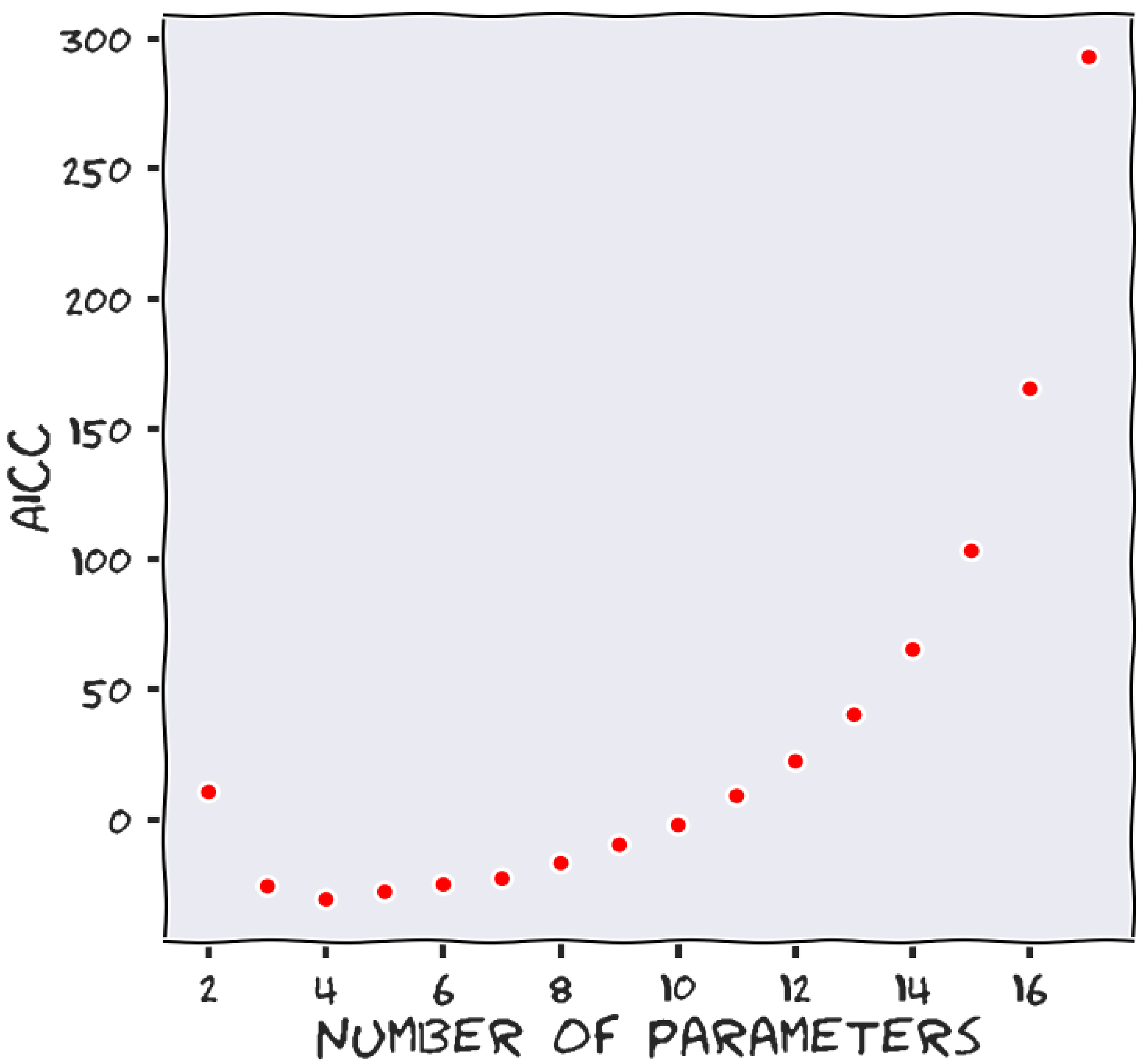


Which model better describe this data?

$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\dots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

Model Selection

Information criteria



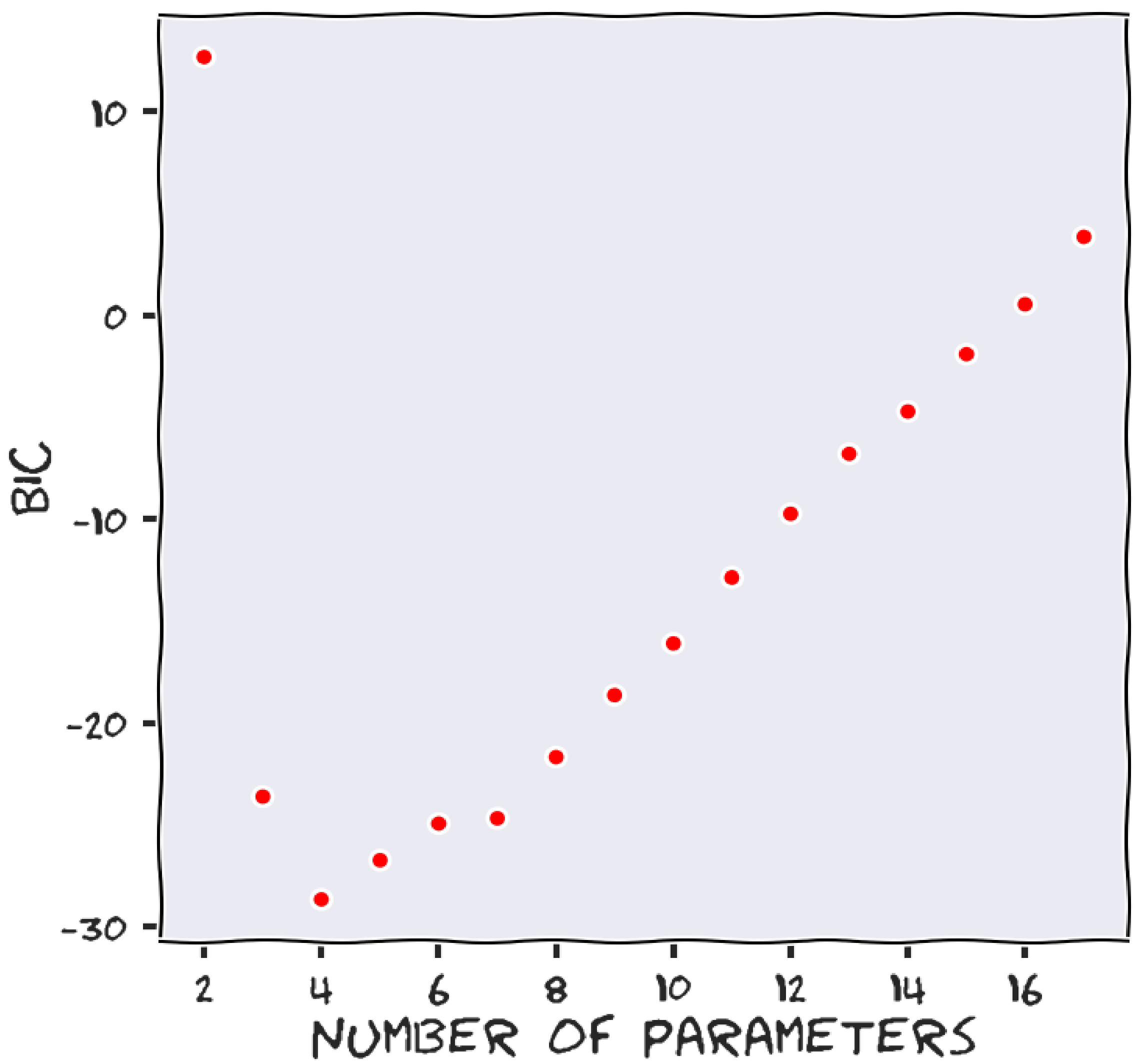
Which model better describe this data?

$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\dots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

Model Selection

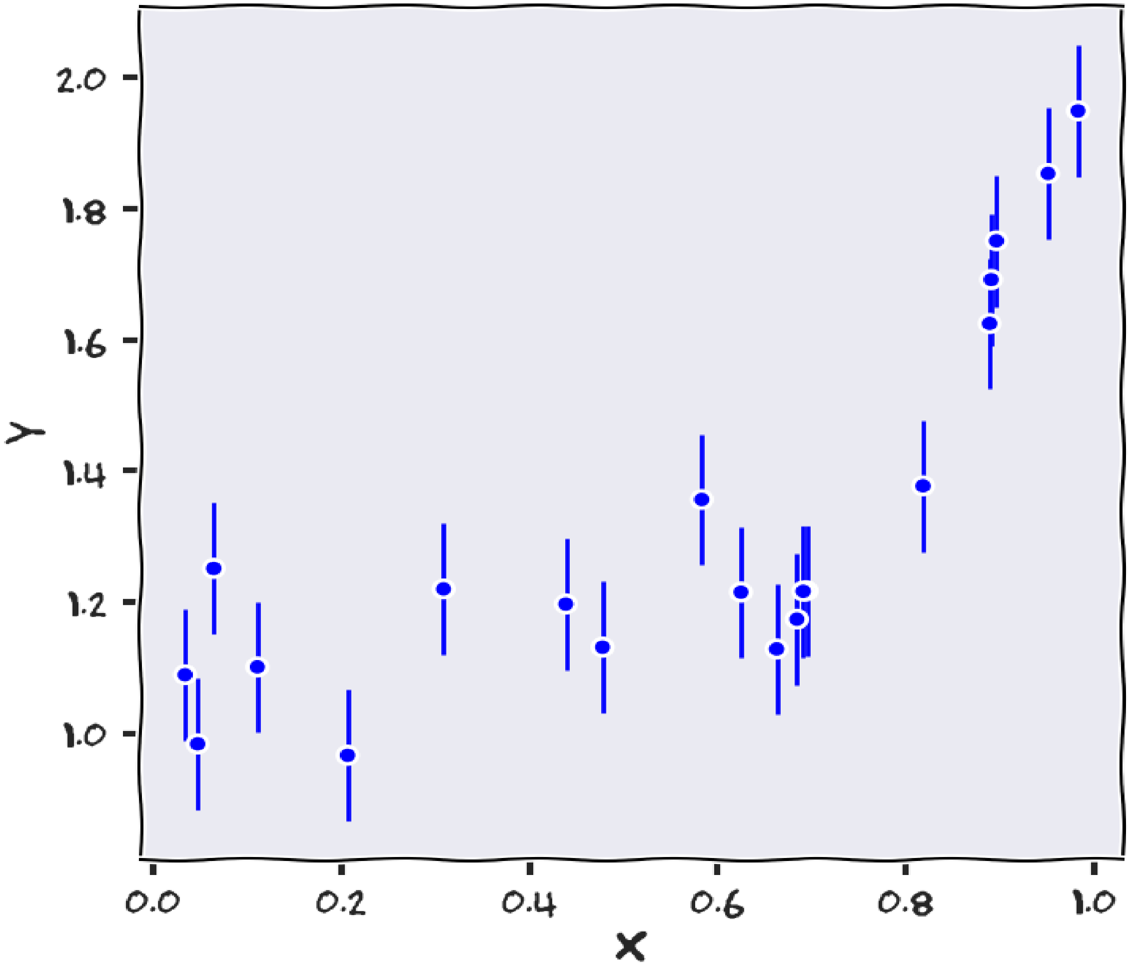
Information criteria

Which model better describe this data?



$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\dots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

Cross-Validation

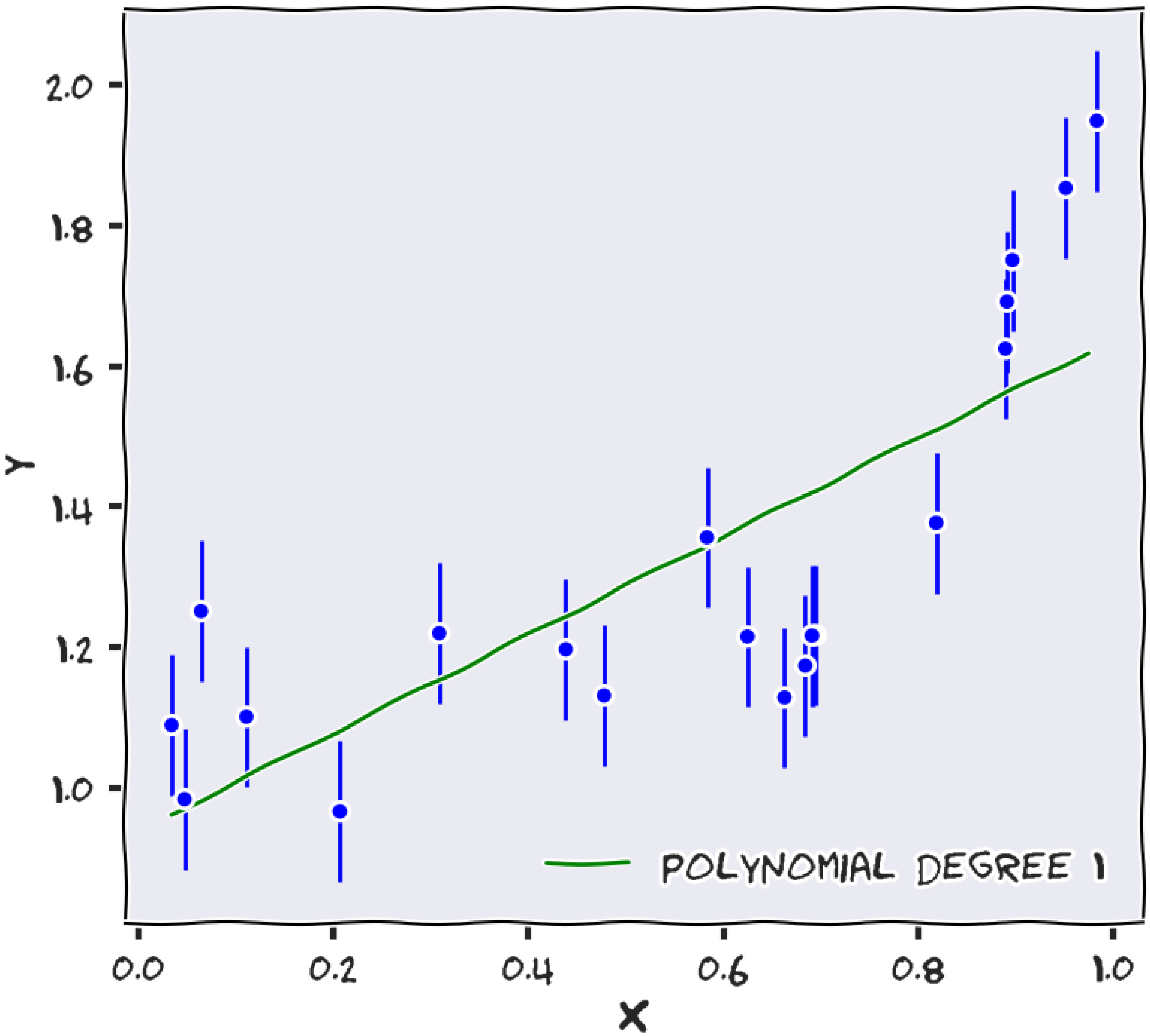


Which model better describe this data?

$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\cdots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

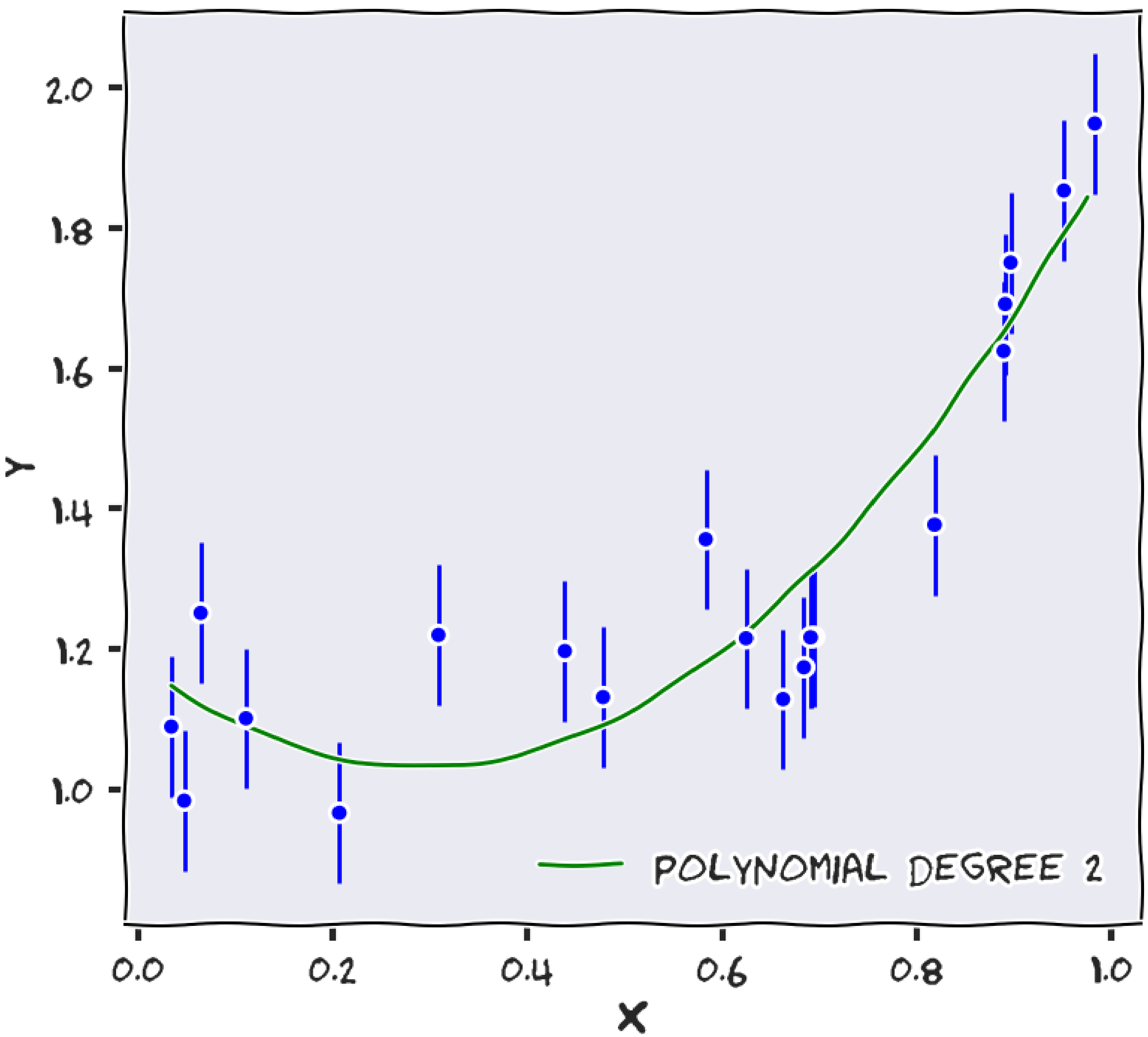
Cross-Validation

Which model better describe this data?



$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\cdots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

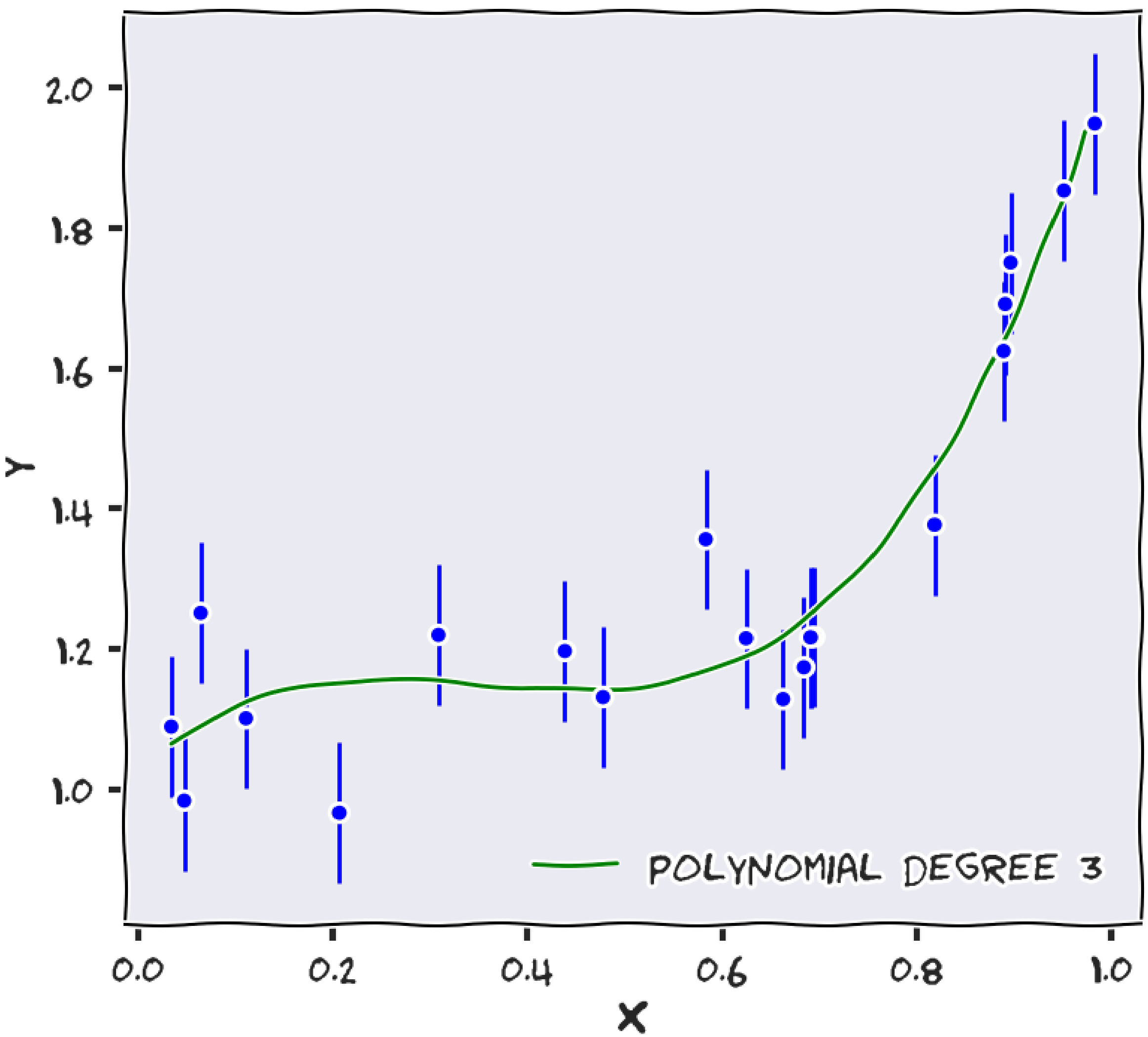
Cross-Validation



Which model better describe this data?

$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\cdots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

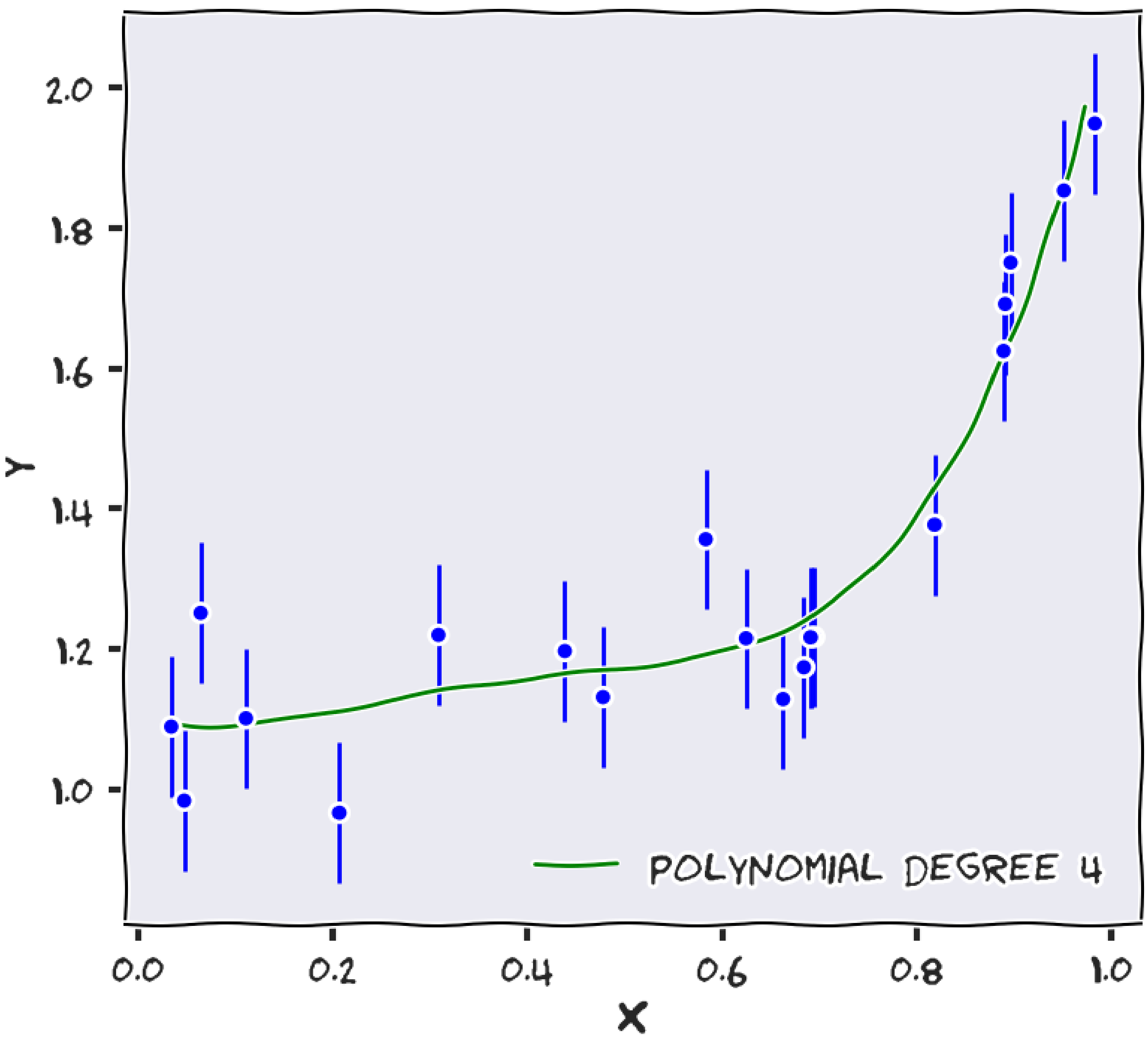
Cross-Validation



Which model better describe this data?

$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\cdots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

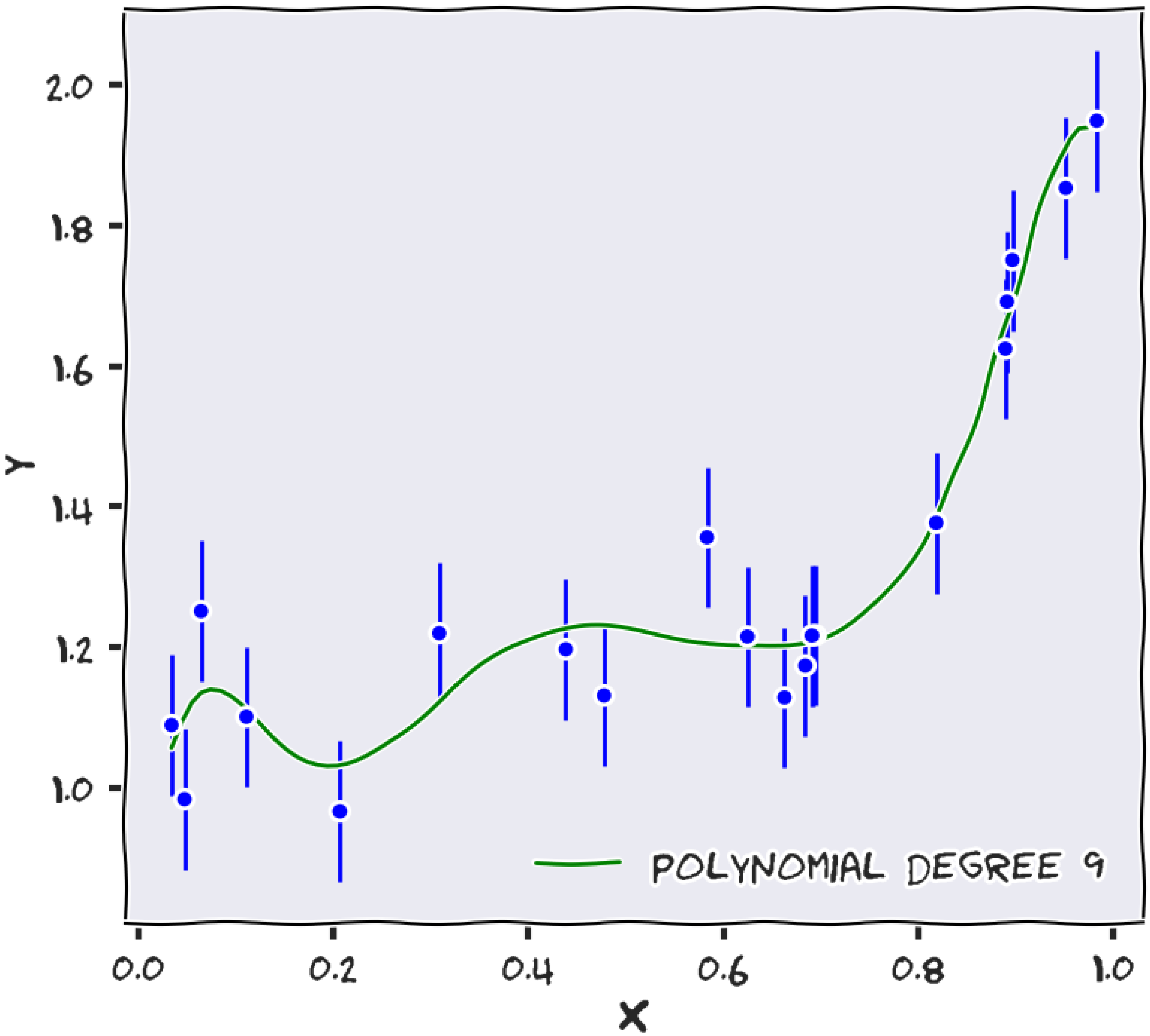
Cross-Validation



Which model better describe this data?

$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\ldots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

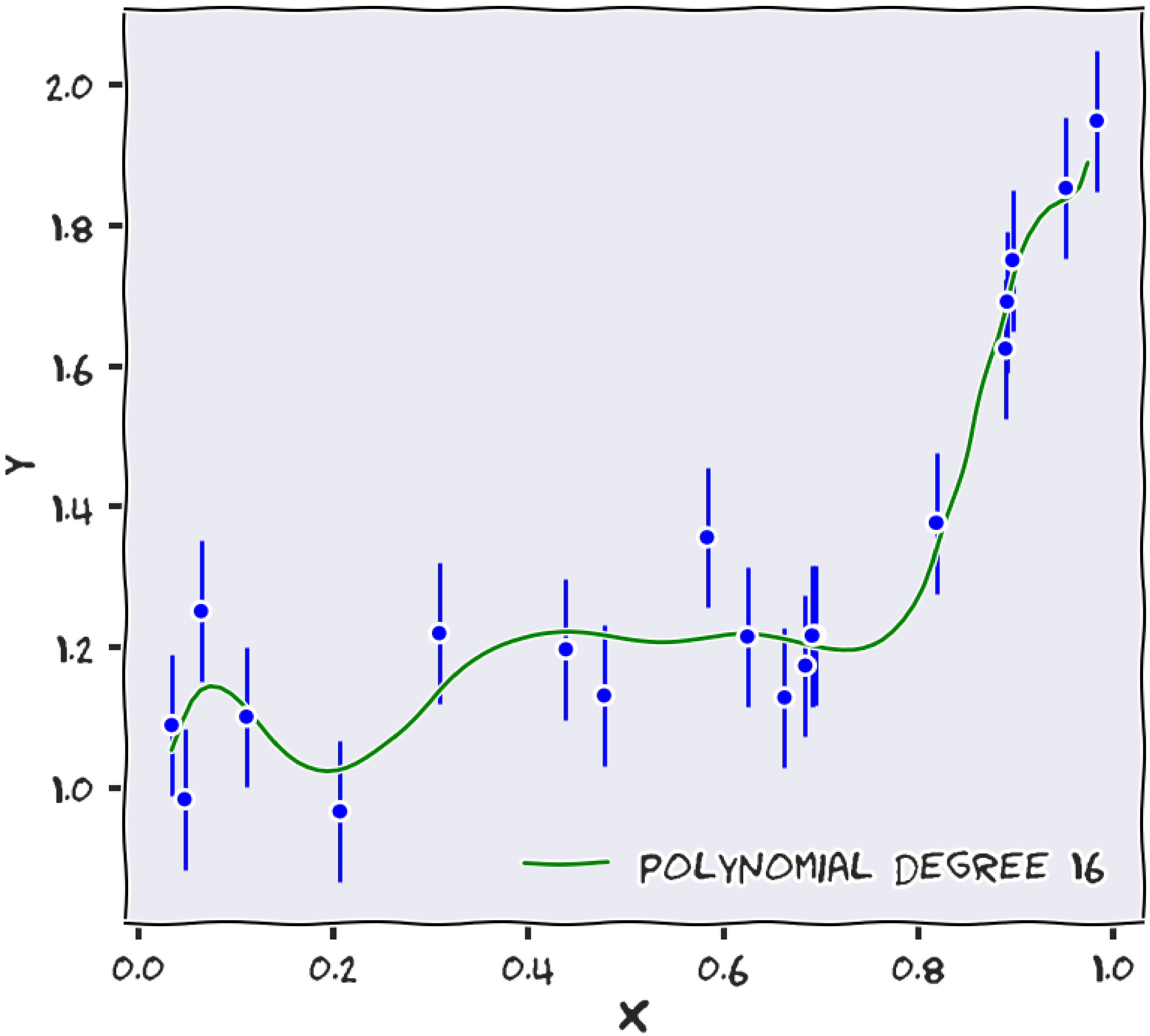
Cross-Validation



Which model better describe this data?

$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\cdots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

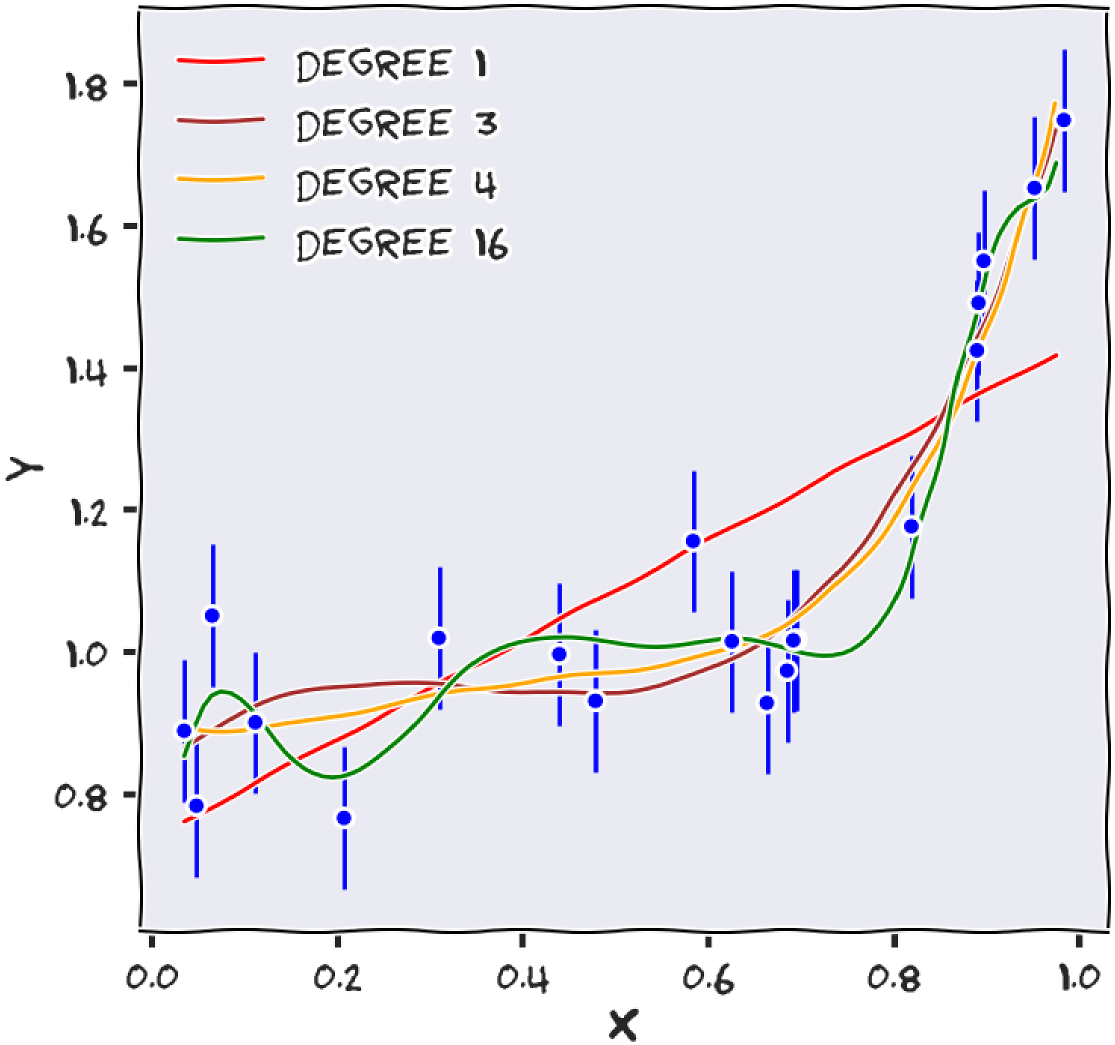
Cross-Validation



Which model better describe this data?

$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\cdots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

Cross-Validation

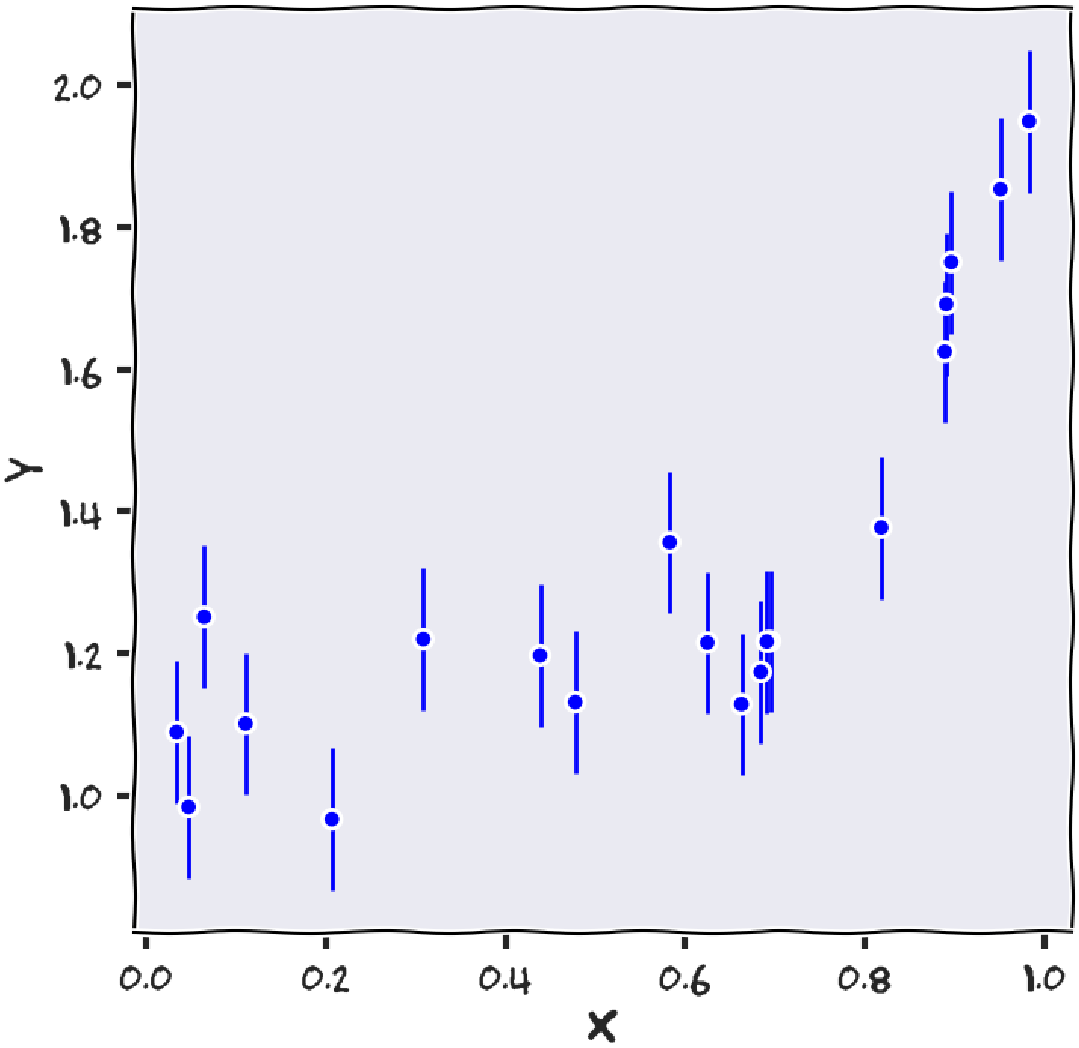


Which model better describe this data?

$$
\begin{aligned}
M_1 : y &= \theta_0 + \theta_1 x \\
M_2 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 \\
M_3 : y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\
&\dots \\
M_k : y &= \sum_{i=0}^{k} \theta_i x^i
\end{aligned}
$$

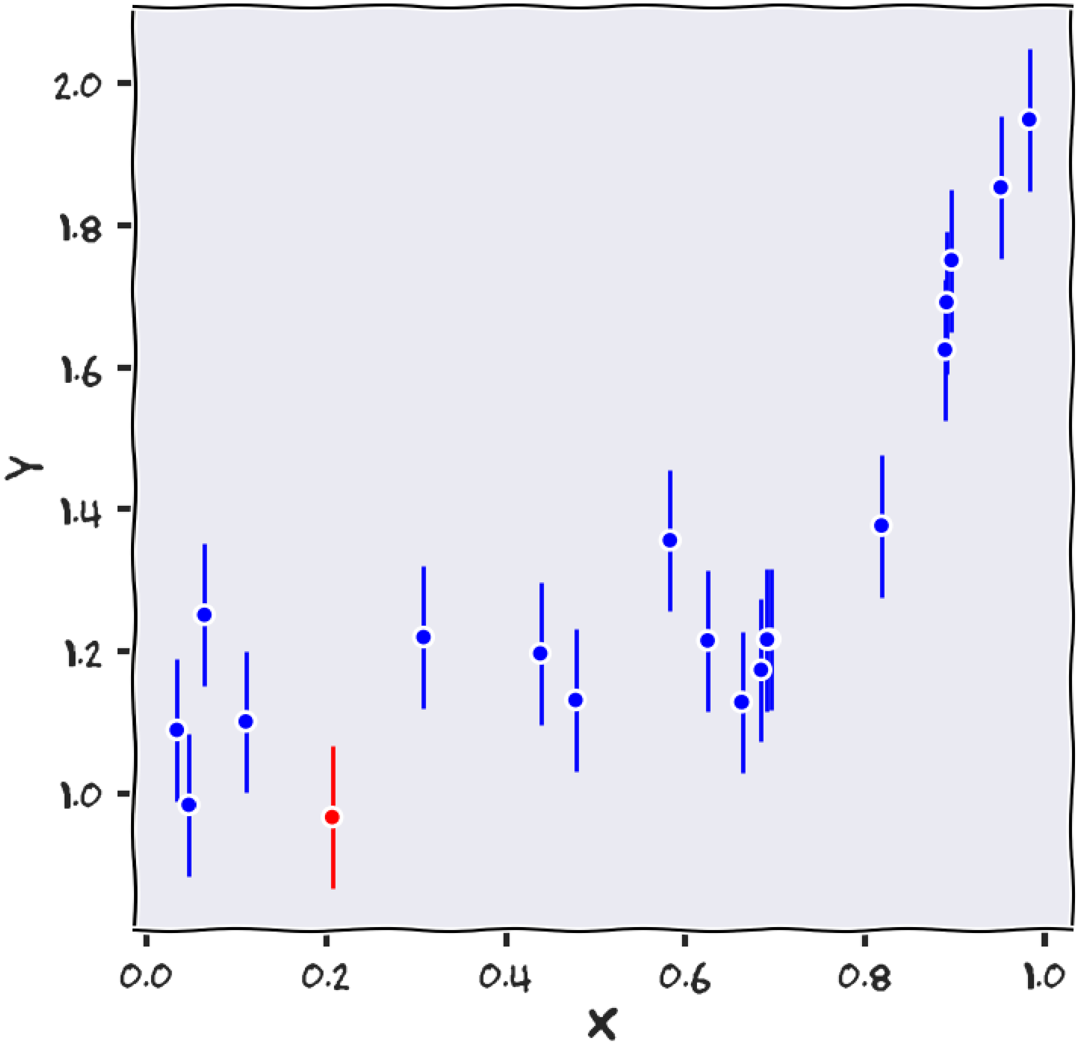
Cross-Validation

Balancing fit and predictiveness



Cross-Validation

Balancing fit and predictiveness

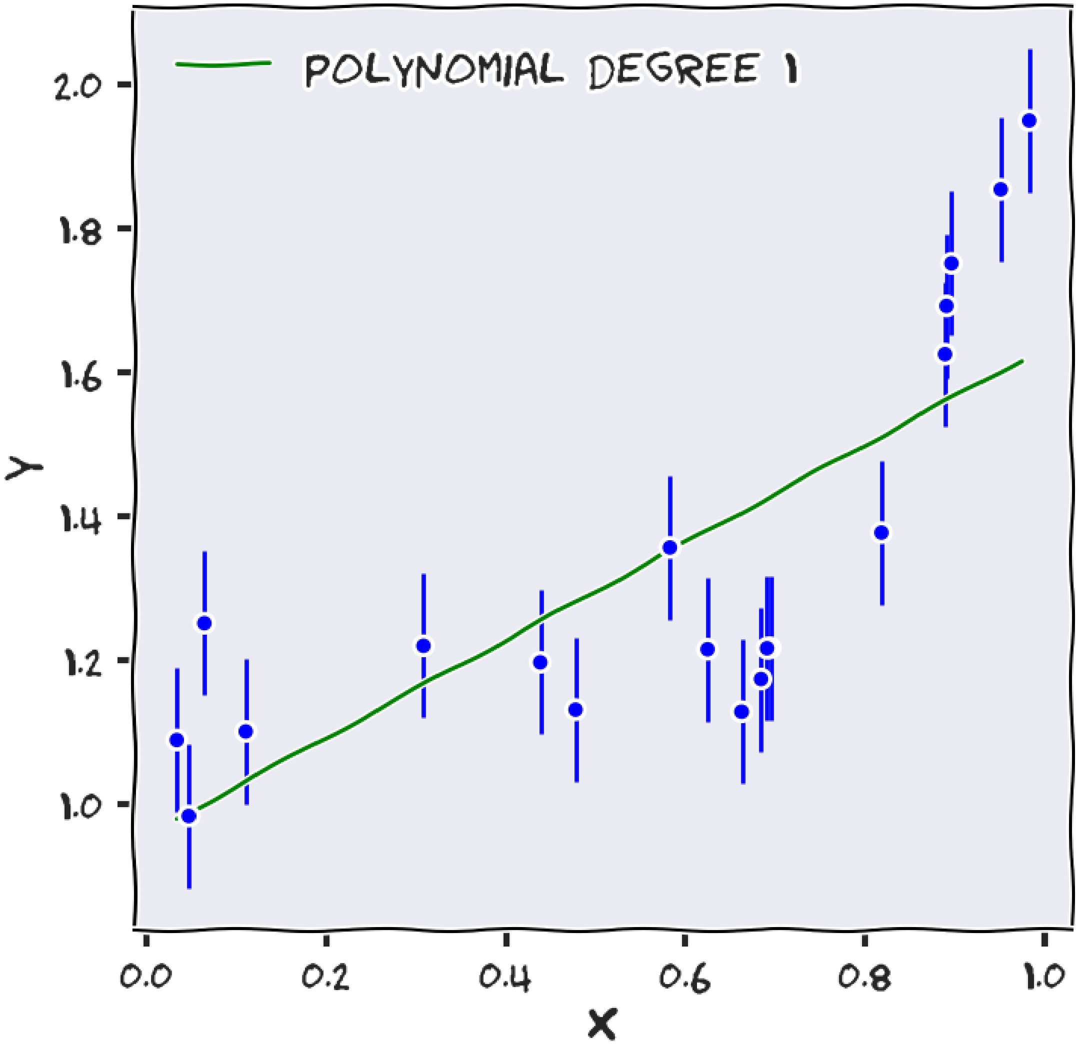


Algorithm:

1 – remove 1 data point from the sample

Cross-Validation

Balancing fit and predictiveness



Algorithm:

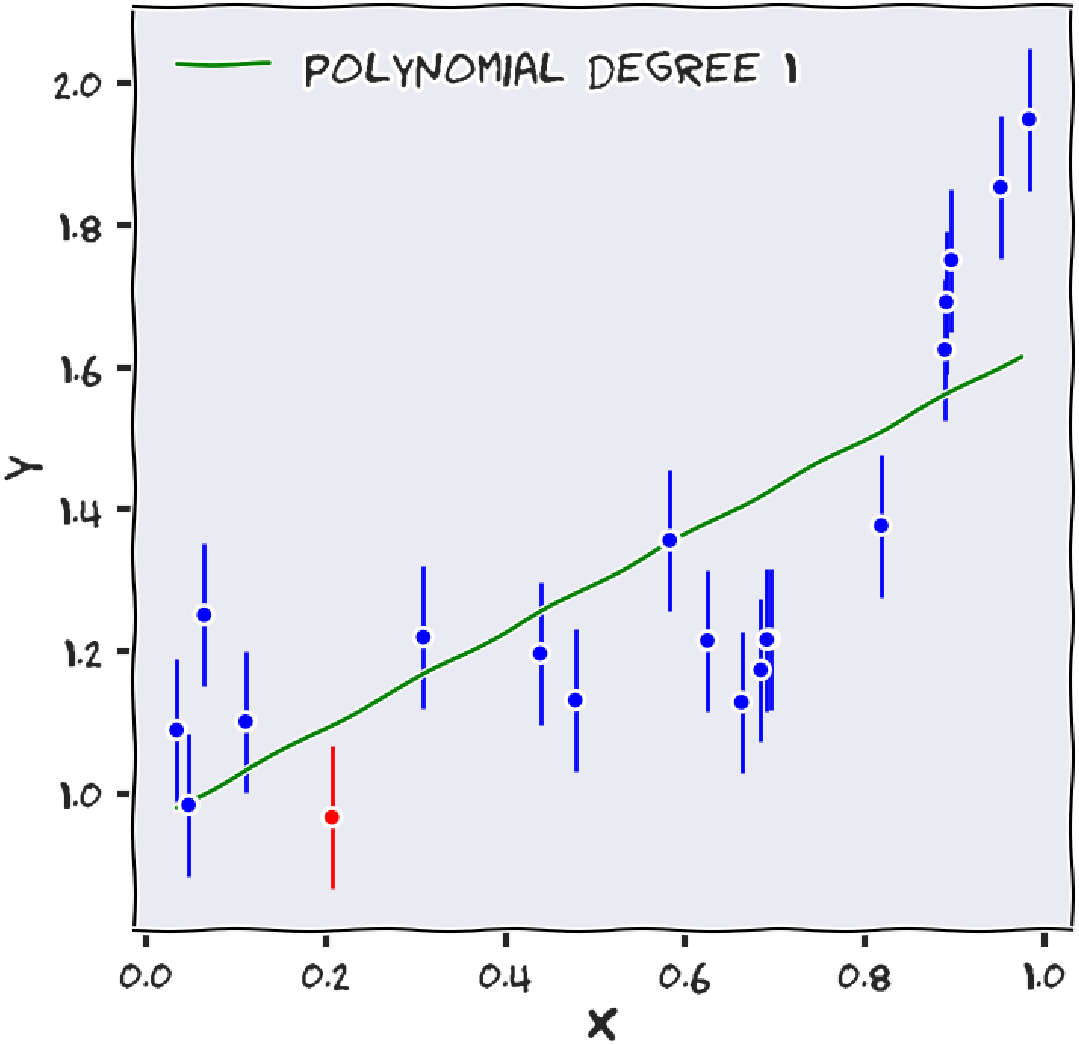
For all models:

1 – remove 1 data point from the sample

2 – fit the model with the remaining data

Cross-Validation

Balancing fit and predictiveness



<u>Algorithm:</u>

For each model:

1 – remove 1 data point from the sample

2 – fit the model with the remaining data

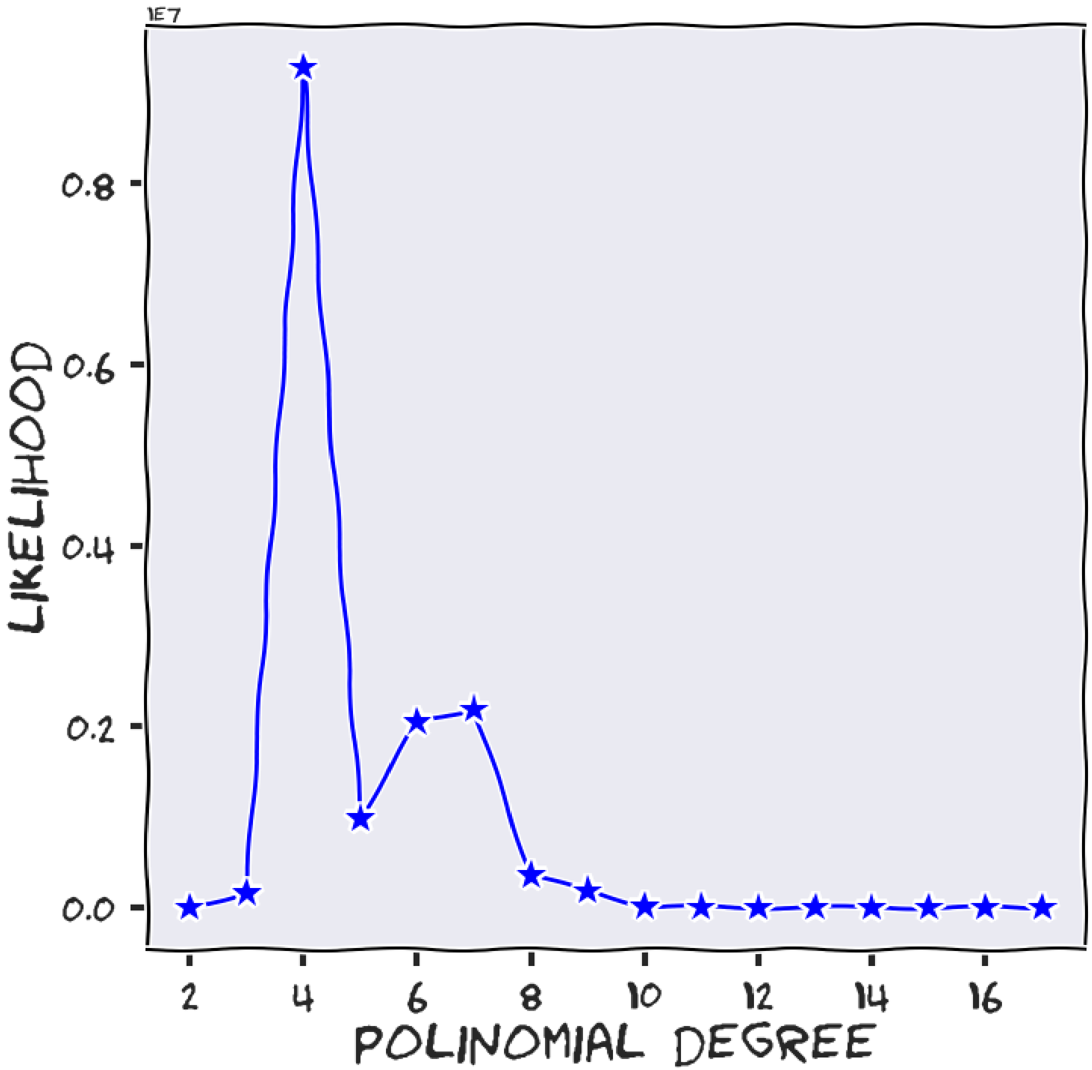
3 – calculate the likelihood of the excluded point under this model

4 – repeat 1-3 for all data points

5 – Calculate the likelihood of all points

Cross-Validation

Balancing fit and predictiveness



Algorithm:

For each model:

1 – remove 1 data point from the sample

2 – fit the model with the remaining data

3 – calculate the likelihood of the excluded point under this model

4 – repeat 1-3 for all data points

5 – Model likelihood = product of the likelihood of all points

Choose the model with highest model likelihood.

Bayesian Evidence

Bayes theorem

$$P(\vec{\theta}|\mathcal{D}) = \frac{P(\mathcal{D}|\vec{\theta})P(\vec{\theta})}{P(\mathcal{D})}$$

Bayesian Evidence

Bayes theorem

$$P(\vec{\theta}|\mathcal{D},\mathcal{M}) = \frac{P(\mathcal{D}|\vec{\theta},\mathcal{M})P(\vec{\theta})}{P(\mathcal{D}|\mathcal{M})}$$

Evidence

Bayesian Evidence

Bayes theorem

$$P(\vec{\theta}|\mathcal{D},\mathcal{M}) = \frac{P(\mathcal{D}|\vec{\theta},\mathcal{M})P(\vec{\theta})}{P(\mathcal{D}|\mathcal{M})}$$

Evidence

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})}$$

Bayesian Evidence

Bayes theorem

$$P(\vec{\theta}|\mathcal{D},\mathcal{M}) = \frac{P(\mathcal{D}|\vec{\theta},\mathcal{M})P(\vec{\theta})}{P(\mathcal{D}|\mathcal{M})}$$

Evidence

Model posterior

$$\boxed{P(\mathcal{M}|\mathcal{D})} = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})}$$

Bayesian Evidence

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})}$$

Goal: compare posteriors of 2 different models!

$B_{12} \leftarrow$ The Bayes factor

$$\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} = \boxed{\frac{P(\mathcal{D}|\mathcal{M}_1)}{P(\mathcal{D}|\mathcal{M}_2)}}\frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}$$

Bayesian Evidence

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})}$$

Goal: compare posteriors of 2 different models!

B_{12} ← The Bayes factor

$$\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} = \boxed{\frac{P(\mathcal{D}|\mathcal{M}_1)}{P(\mathcal{D}|\mathcal{M}_2)}}\frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}$$

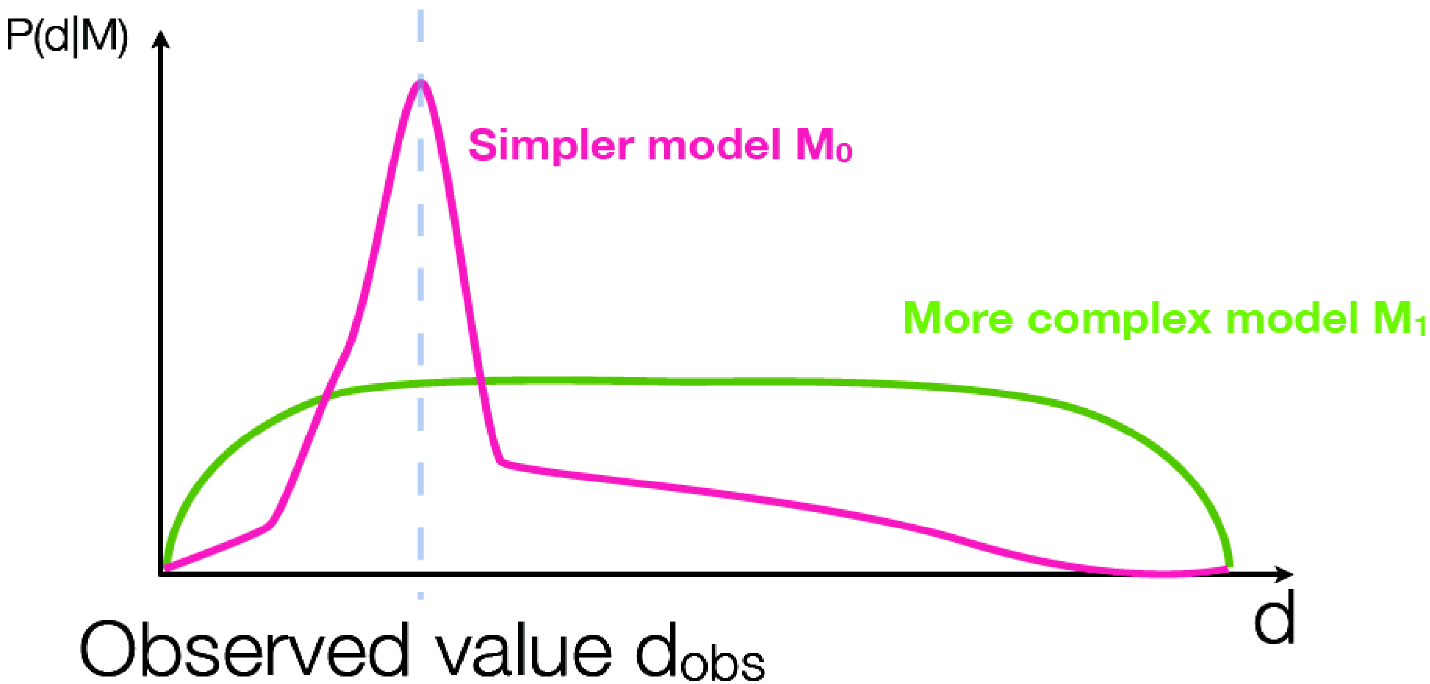
Posterior odds = Bayes factor x prior odds

Bayesian Evidence

Occam's razor

$$E(\mathcal{D}|\mathcal{M}) = \int_{\Omega_\theta} \mathcal{L}(\mathcal{D}|\bar{\theta}) P(\bar{\theta}) d\bar{\theta}$$

Volume weighted likelihood



Plot by Roberto Trotta, 2011

Bayesian Evidence

Jeffrey's scale

|lnB|	relative odds	favoured model's probability	Interpretation
< 1.0	< 3:1	< 0.750	not worth mentioning
< 2.5	< 12:1	0.923	weak
< 5.0	< 150:1	0.993	moderate
> 5.0	> 150:1	> 0.993	strong

Table by Roberto Trotta, 2011

Nesseris and Garcia-Belido, Is the Jeffreys' scale a reliable tool for Bayesian model comparison in cosmology?, JCAP, 08, article id. 036, 2013

Bayesian Evidence

Computation

$$E(\mathcal{D}|\mathcal{M}) = \int_a \int_b \mathcal{L}(\vec{x}, \vec{y}|a, b, \sigma)p(a)p(b)dadb$$

Nested Sampling

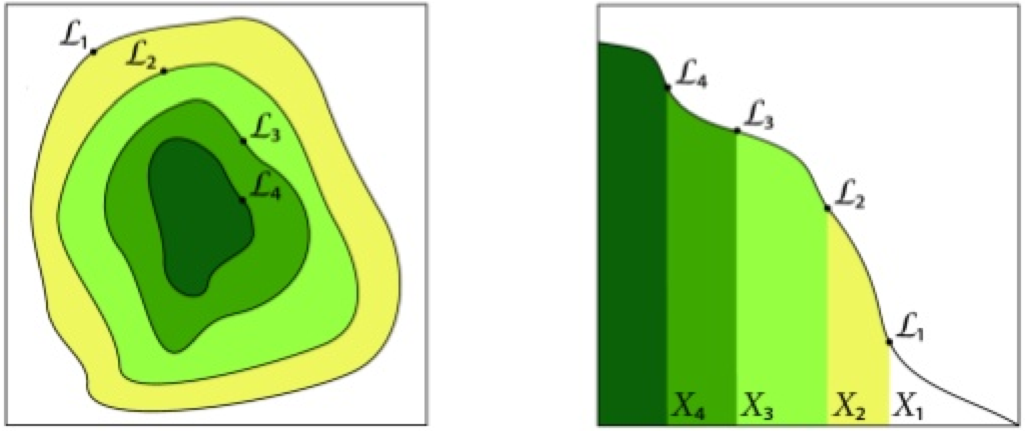


Image credit: Feroz et al., 2013

A **population of points** are randomly sampled. For iteration, i, **the point with lowest likelihood value, L_i,** is removed from the live point set and **replaced** by another point drawn from the prior under the **constraint that its likelihood is higher than** L_i

Multinest

DIAMONDS

CosmoSIS

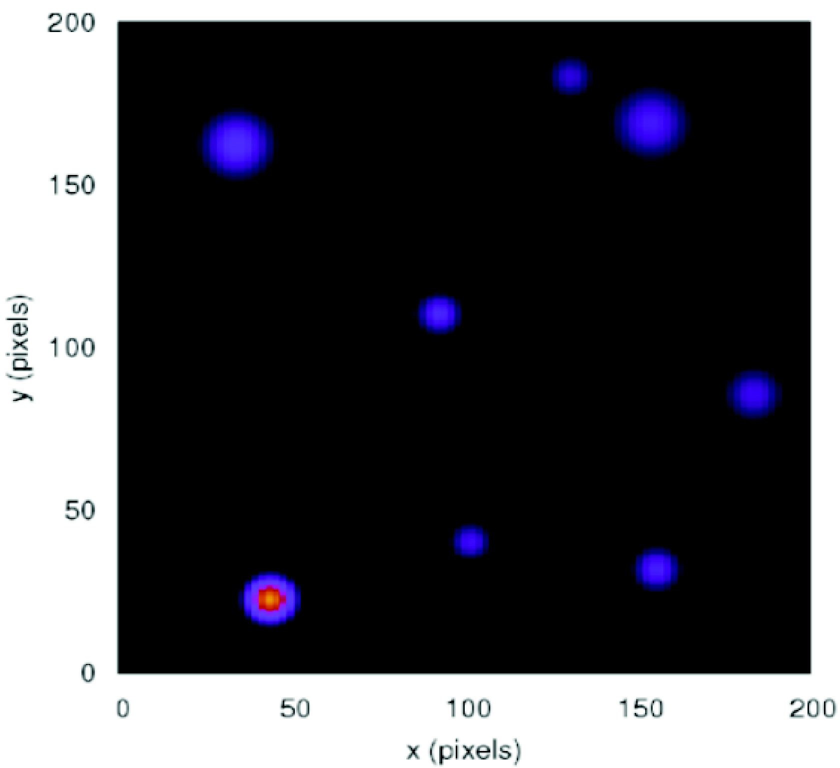
Bayesian Evidence

Application

A “simple” example: how many sources?

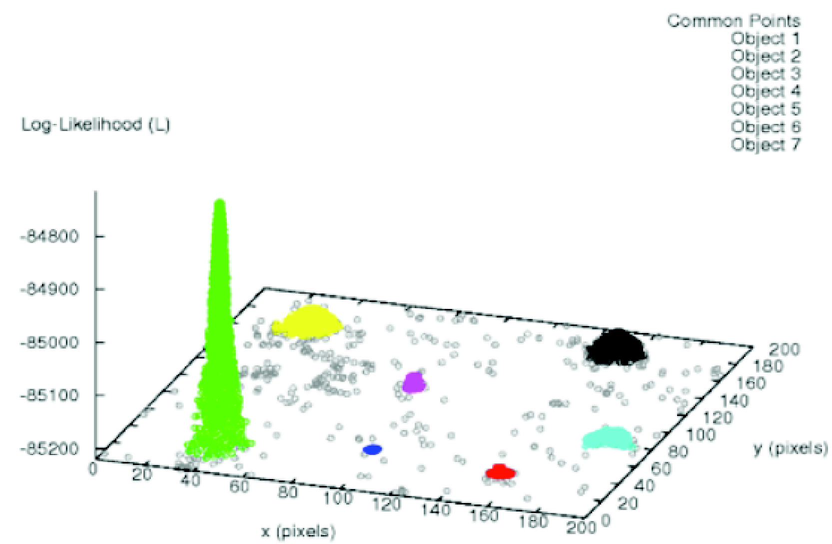
Imperial College London

Feroz and Hobson (2007)



Bayesian reconstruction

7 out of 8 objects correctly identified.
Mistake happens because 2 objects very close.



Summary

- Information criteria should be used with parsimony

- Cross-validation is a nice alternative which uses a non-prohibit amount of computational resources

- Bayesian evidences are proven to be very useful in cosmology, given the complexity of the models .. but the computational cost is also very high

What if your likelihood is

not available?