Model UQ

- No model of a physical system is strictly true
- The probability of a model being strictly true is zero
- Given limited information, some models may be relied upon for describing the system

Let $\mathcal{M} = \{M_1, M_2, \ldots\}$ be the set of all models

- $p(M_k|I)$ is the probability that M_k is the model behind the available information
  - Model Plausibility
- Parameter estimation from data is conditioned on the model

$$p(\theta|D, M_k) = \frac{p(D|\theta, M_k)\pi(\theta|M_k)}{p(D|M_k)}$$

Bayesian Model Comparison

Evidence (marginal likelihood) for M_k:

$$p(D|M_k) = \int p(D|\theta, M_k)\pi(\theta|M_k)\mathrm{d}\theta$$

Bayes Factor B_{ij}:

$$B_{ij} = \frac{p(D|M_i)}{p(D|M_j)}$$

Plausibility of M_k:

$$p(M_k|D, \mathcal{M}) = \frac{p(D|M_k)\ \pi(M_k|\mathcal{M})}{\sum_s p(D|M_s)\pi(M_s|\mathcal{M})} \qquad k = 1, \ldots$$

Posterior odds:

$$\frac{p(M_i|D, \mathcal{M})}{p(M_j|D, \mathcal{M})} = B_{ij}\ \frac{\pi(M_i|\mathcal{M})}{\pi(M_j|\mathcal{M})}$$

Marginal Likelihood example

- Consider Fitting with data from a truth model

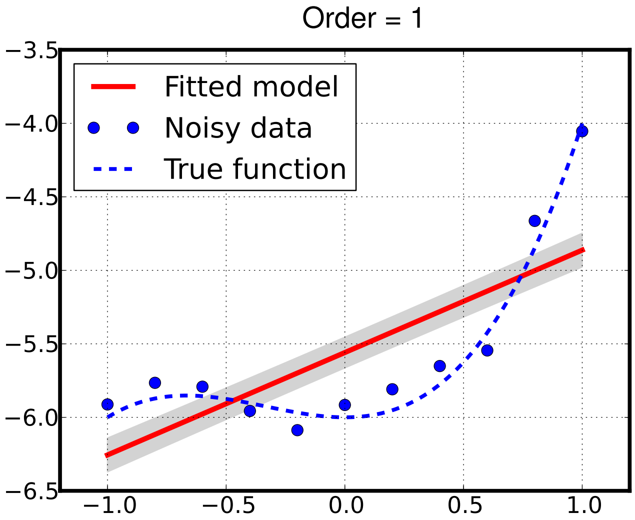
$$y_t = x^3 + x^2 - 6$$

- Gaussian *iid* additive noise model with fixed variance s
- Bayesian regression with a Gaussian Likelihood, *iid* and given s
- Consider a set of Legendre Polynomial expansion models, order 1-10

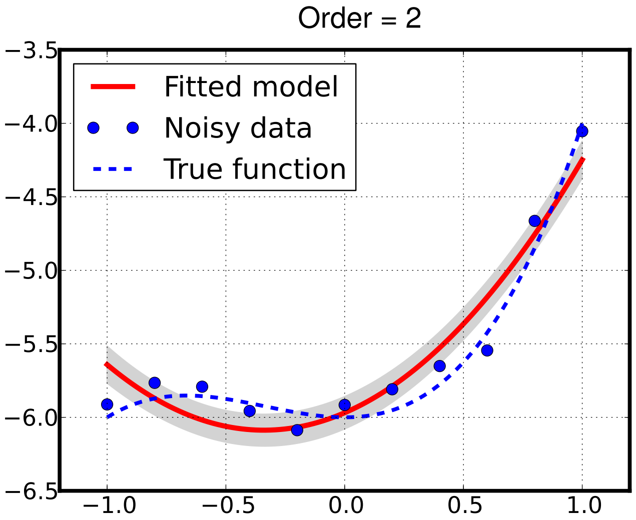
$$y_m = \sum_{k=0}^{P} c_k \psi_k(x)$$

- Uniform priors $[-D, D]$ on all coefficients

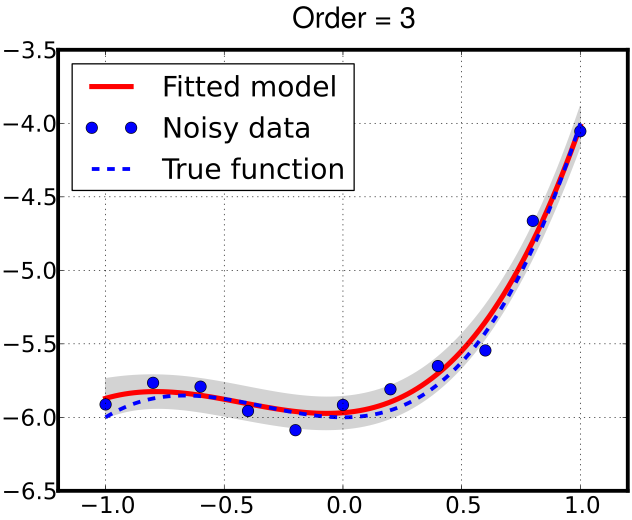
Too much model complexity leads to overfitting



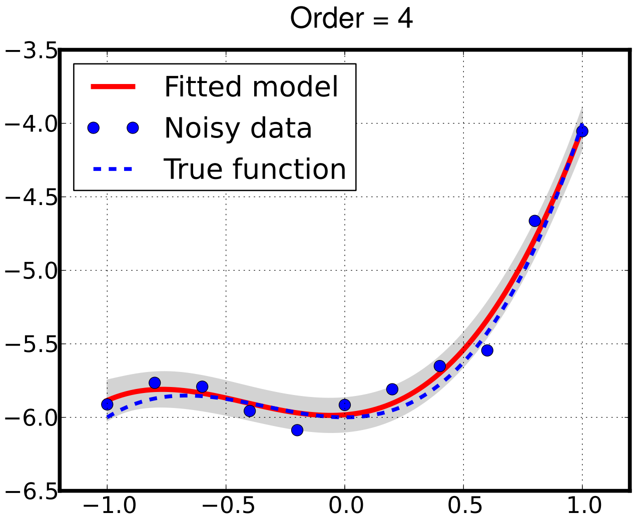
Too much model complexity leads to overfitting



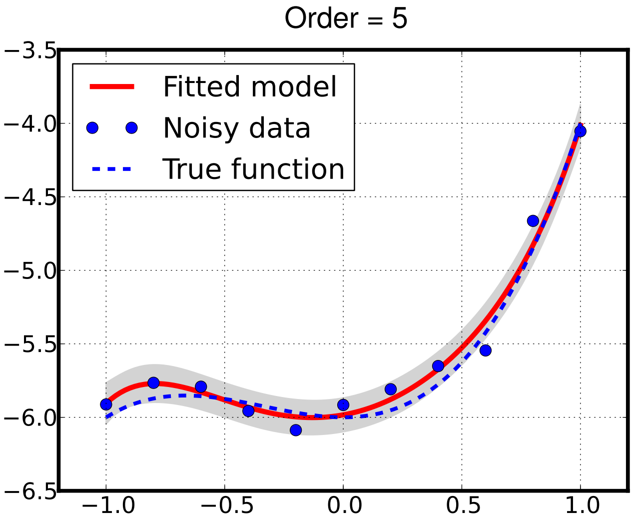
Too much model complexity leads to overfitting



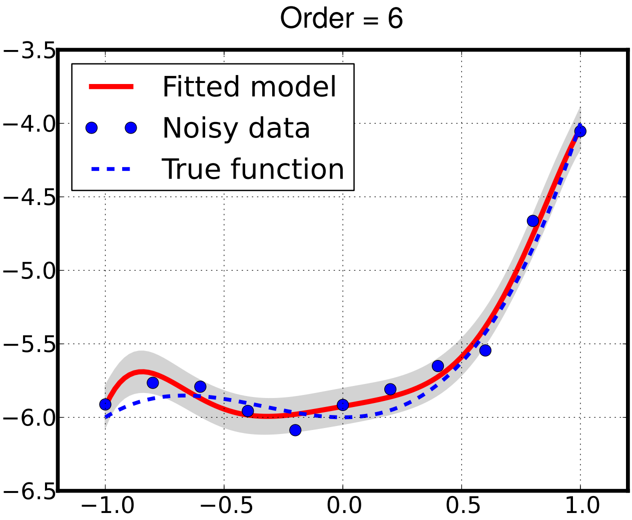
Too much model complexity leads to overfitting



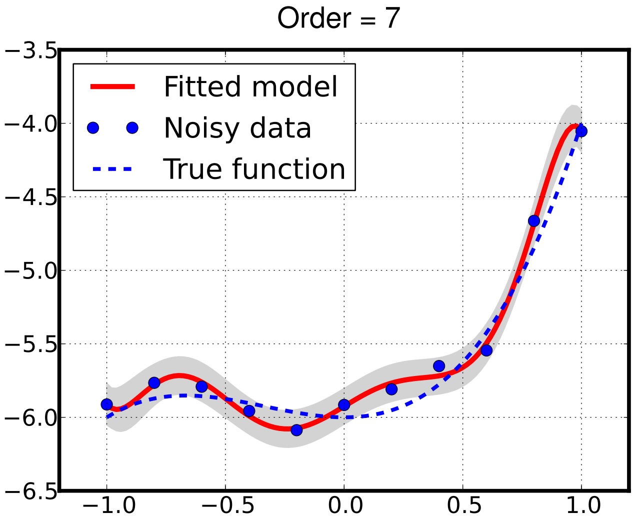
Too much model complexity leads to overfitting



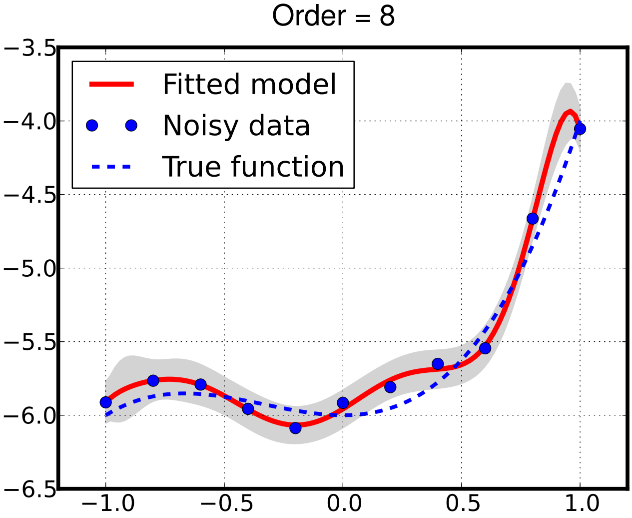
Too much model complexity leads to overfitting



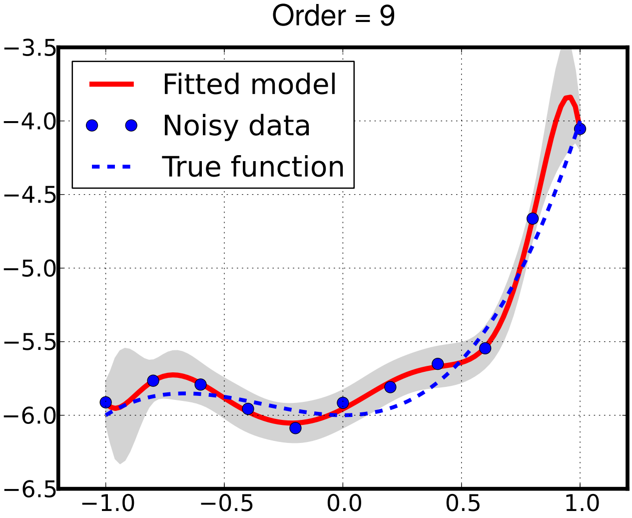
Too much model complexity leads to overfitting



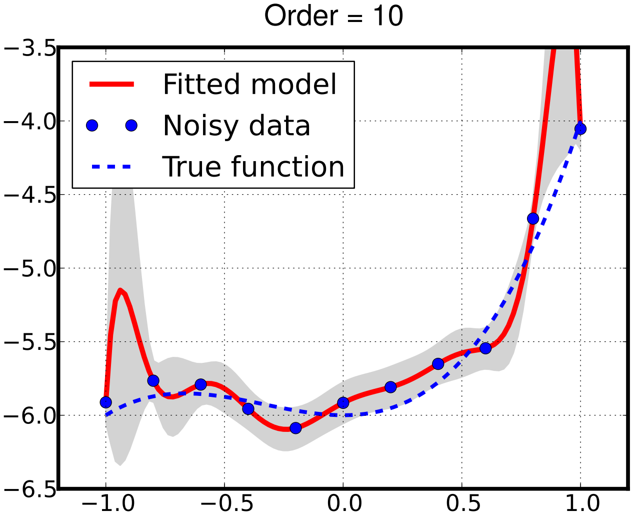
Too much model complexity leads to overfitting



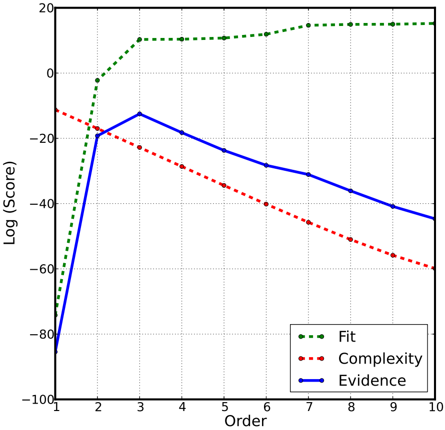
Too much model complexity leads to overfitting



Too much model complexity leads to overfitting

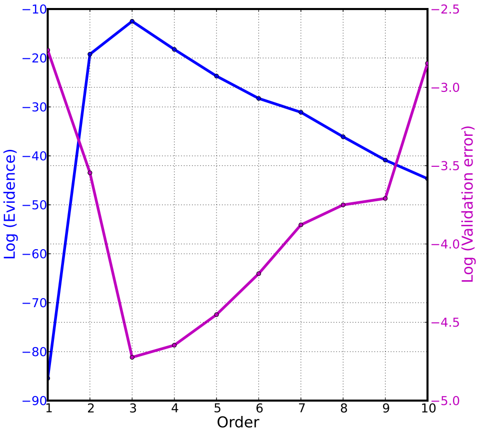


Evidence – Marginal Likelihood



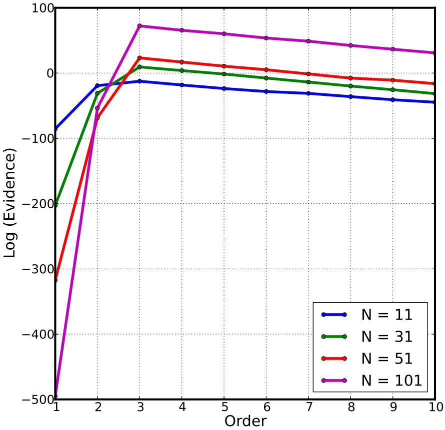
- Log evidence: sum of two scores, balances complexity & fit
- Peaks at 3rd order

Evidence and Validation Error



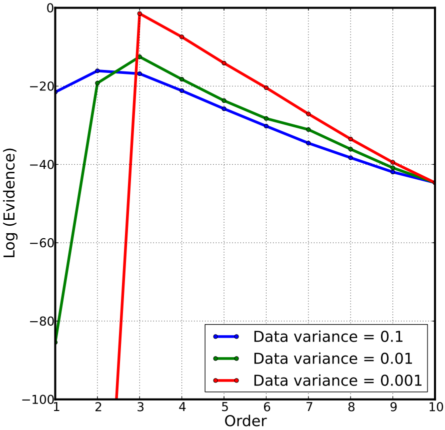
- Validation error – ℓ_2 error for a random set of 1000 points
- Validation error is minimal at the 3rd-order evidence peak

Evidence



- Discrimination among models is more clear-cut with higher amount of data

Evidence



- Discrimination among models is more clear-cut with less data noise

Prediction

Consider that a model

$$y_m = f(x, \lambda)$$

was fitted according to

$$y = f(x, \lambda) + \epsilon, \quad \epsilon \sim N(0, \sigma^2),$$

providing:

- The posterior $p(\lambda, \sigma|D)$
- The marginal posterior $p(\lambda|D)$

Define:

- Pushed forward posterior (PFP) distribution : $p(y_m|x, D)$
- Posterior predictive (PP) distribution : $p(y|x, D)$

Pushed forward posterior (PFP)

- PFP distribution $p(y_m|x, D)$
- Push-forward of the marginal posterior measure on λ through $f(x, \lambda)$
- PFP random process

$$
\begin{aligned}
Y_m(x, \omega) &= f(x, \lambda(\omega)) \\
&\sim p(y_m|x, D)
\end{aligned}
$$

- The PFP provides the uncertain prediction by the calibrated model

Posterior Predictive (PP)

Posterior Predictive distribution $p(y|x, D)$

- With $\alpha \equiv (\lambda, \sigma)$,

$$p(y|x, D) = \int p(y|x, \alpha, D)p(\alpha|D)\mathrm{d}\alpha$$

PP random process

$$\begin{aligned} Y^{PP}(x, \omega) &= \mathrm{E}_{\alpha}[Y(x, \omega)] \\ &\sim p(y|x, D) \end{aligned}$$

provides the marginal prediction of the data. Where

$$Y(x, \omega) = f(x, \lambda) + \epsilon(\omega, \sigma)$$

is the PFP data predictor

Validation

- Validity is a statement of model utility for predicting a given observable under given conditions
- Inspection of model utility requires accounting for uncertainty
- Statistical tool-chest for model validation
  - Cross-validation
  - Bayes Factor
  - Model Plausibility
  - Posterior Odds
  - Posterior predictive:

$$p(\tilde{D}|D, M_k) = \int p(\tilde{D}|\theta, M_k)p(\theta|D, M_k)d\theta$$