Model selection

CWI Autumn School - Scientific Machine Learning and Dynamical Systems

Urban Fasel

Imperial College London

Literature

- **Model selection**
  - SL Brunton, JL Proctor, JN Kutz (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems.
  - NM Mangan, JN Kutz, SL Brunton, JL Proctor (2017) Model selection for dynamical systems via sparse regression and information criteria.
  - S Maddu, BL Cheeseman, IF Sbalzarini, CL Muller (2019) Stability selection enables robust learning of partial differential equations from limited noisy data.

Model selection

Model selection is **not simply about reducing error**, but about producing a model that has high degree of **interpretability**, **generalization**, and **predictive capabilities.**

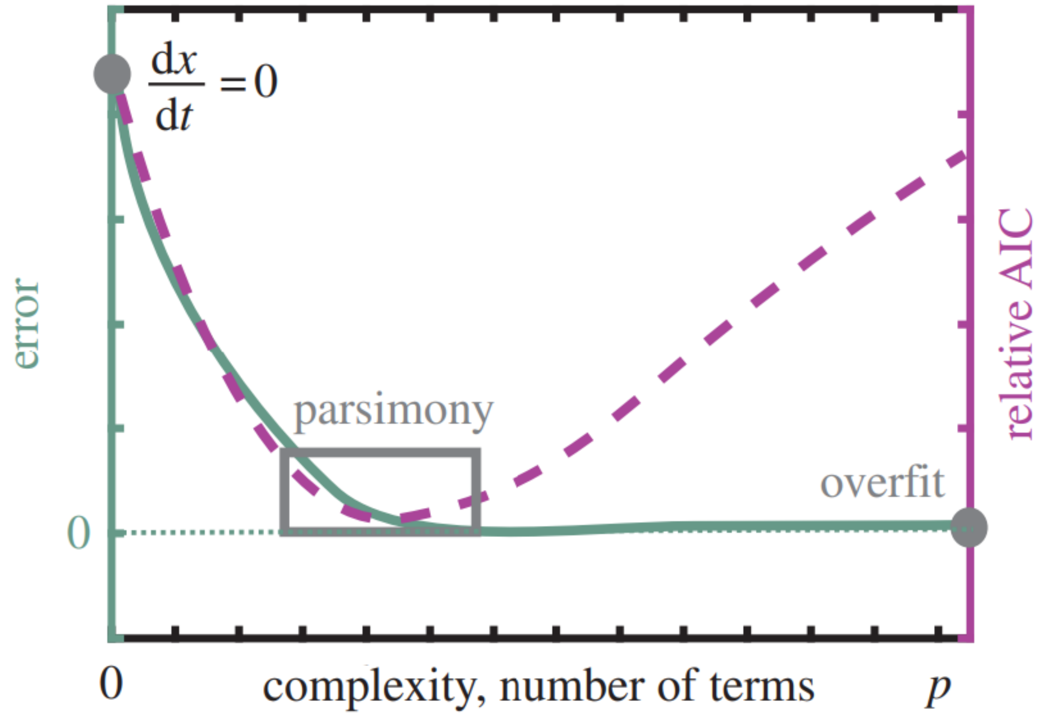
→ selecting **accurate** and **sparse** model

Model selection methods

- Cross validation
- Akaike information criteria
- Stability selection

Tutorials

- Compare methods on Lorenz system data

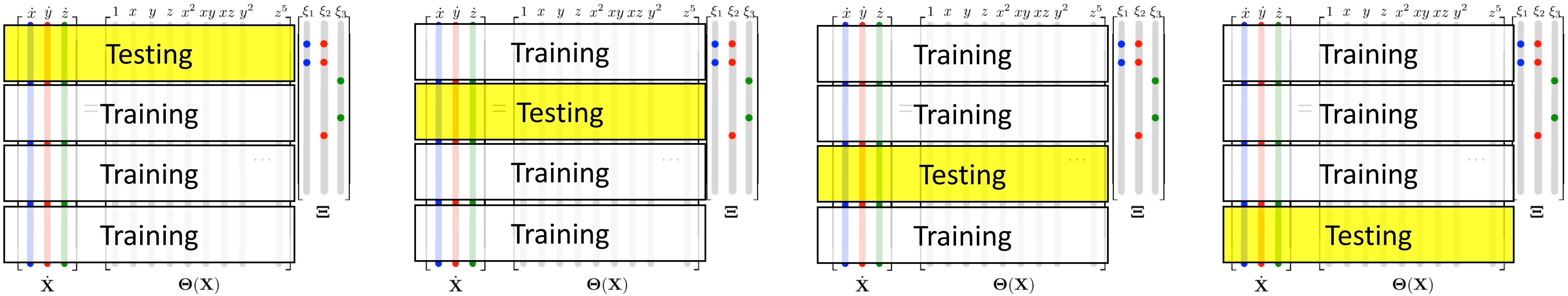


Model selection – k-fold cross-validation

Model selection: sweep through λ-path → repeat following 3 steps for increasing λ (model complexity)

1. Partition data in k (random) subsamples. *(e.g. 4-fold CV shown here)*
2. Build k SINDy models:
   - use $k-1$ subsamples for training
   - test the model using the withheld testing sample: $\epsilon_k = \left\|\dot{\mathbf{X}}_k - \boldsymbol{\Theta}(\mathbf{X}_k)\boldsymbol{\xi}_k\right\|_2^2$
3. Average the test scores ϵ_k → $\epsilon_\lambda = \frac{1}{k}\sum \epsilon_k$

Select model (choose λ) with lowest test score ϵ_λ

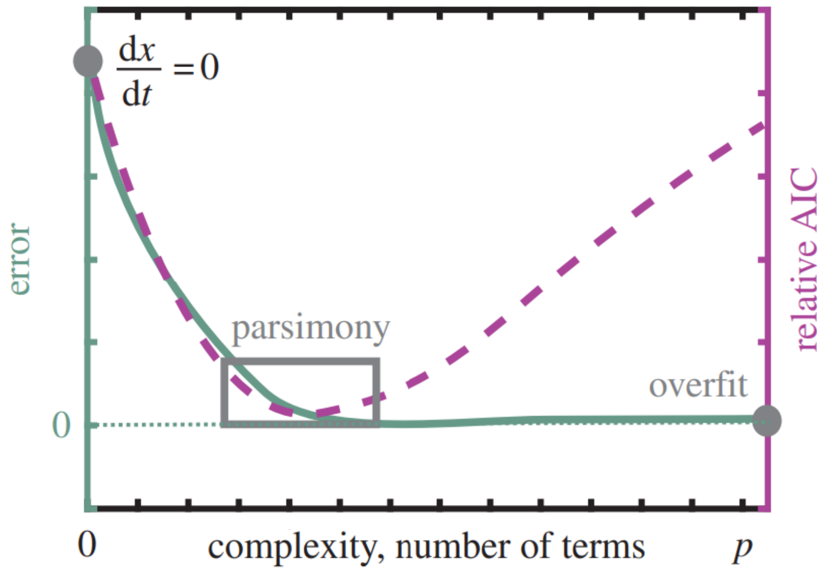


Model selection – Akaike information criteria

Sum of model prediction error plus number of active terms

- Model quality relative to other models
- Increased information criteria score for larger, overfit models
  - Creates min in AIC curve: allows for intuitive model selection

- $\text{AIC} \;=\; -2\ln\big(L(\boldsymbol{x},\hat{\mu})\big) + 2k$
  - Max log likelihood L of model + number of parameters k of model
    - $L(\boldsymbol{x},\hat{\mu}) = P(\boldsymbol{x}|\mu)$, likelihood function: conditional probability of observations $\boldsymbol{x}$ given the parameters μ of a candidate model
  - **SINDy** likelihood function: model prediction error $\rightarrow$ $E_{avg} = \sum_{\tau} |y_i - g(x_i;\mu)|$
    - y_i observed outcomes (time series data)
    - $g(x_i;\mu)$ SINDy model prediction (integrated SINDy ODE)
  - $\text{AIC} \;= m\ln\left(\left(\sum_{i=1}^{m} E_{avg}\right)^2/m\right) + 2k$
    - m number of test time series (e.g. starting from different initial conditions)



NM Mangan, JN Kutz, SL Brunton, JL Proctor (2017) [Model selection for dynamical systems via sparse regression and information criteria]

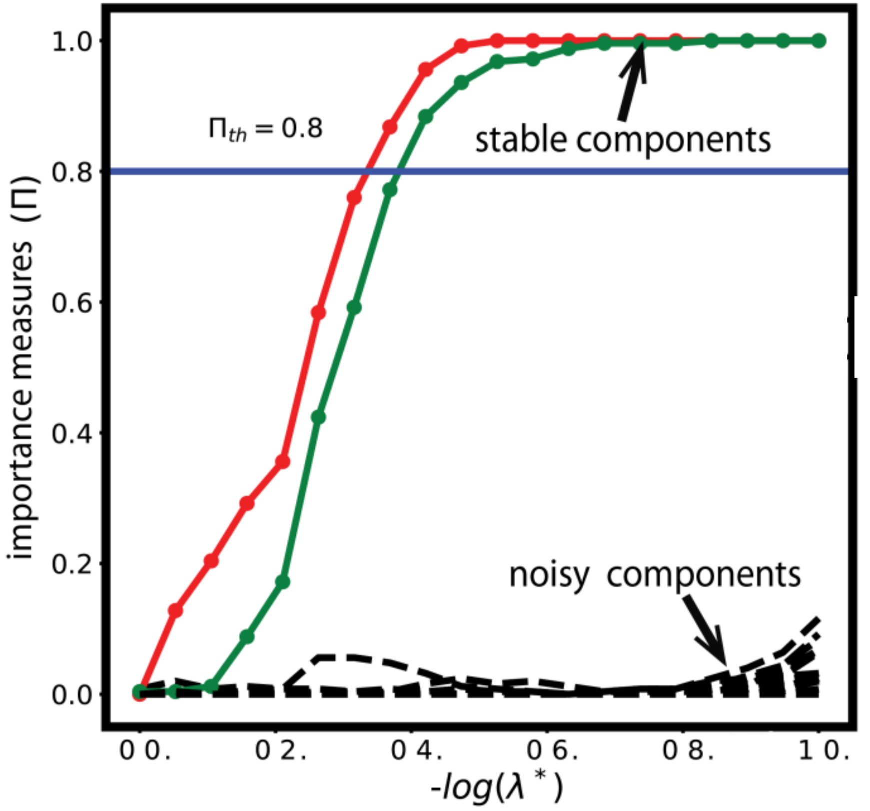
Model selection – stability selection

Calculate the stability (importance) of each coefficient over the regularization path λ

1. Generate B random subsamples (size $N/2$) of the library and derivative data without replacement

2. Compute B SINDy models over regularization path λ

3. Calculate the λ-dependent stability (or importance) Π_k^λ
   - Probability of a coefficient to be non-zero
   - Find stable support: $\Pi_k^\lambda > \pi_{th} = 0.8$
     - e.g. coefficient is non-zero in 80% of the models



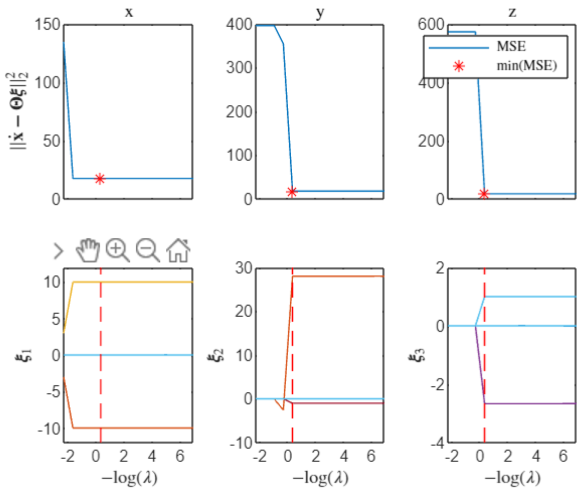
S Maddu, BL Cheeseman, IF Sbalzarini, CL Muller (2019) [Stability selection enables robust learning of partial differential equations from limited noisy data].

Tutorial summary

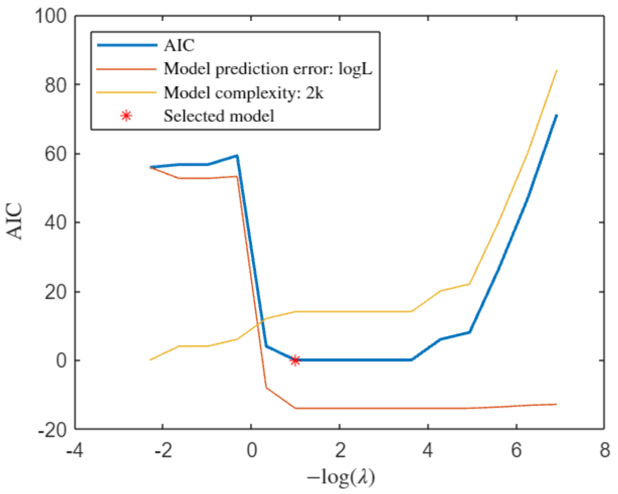
Model selection is **not simply about reducing error**, but about producing a model that has high degree of **interpretability**, **generalization**, and **predictive capabilities.**

→ selecting **accurate** and **sparse** model

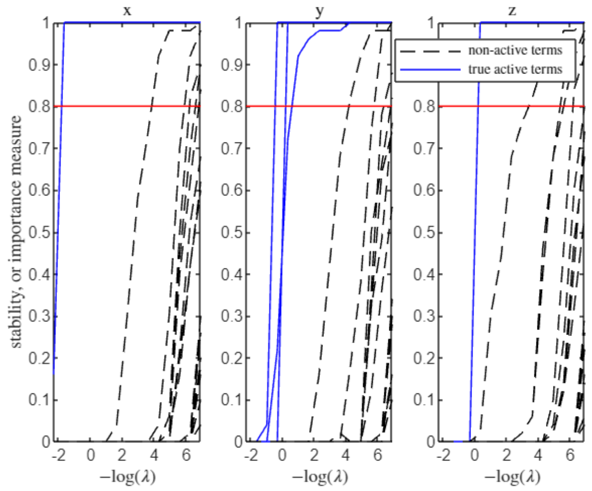
Cross validation



Akaike information criteria



Stability selection



Coding examples

- **MATLAB live scripts**
  - Test different model selection methods on dysts database ODEs: Database, paper, PySINDy
- **PySINDy**
  - AIC is implemented in example 16