Along the **SINDy** Frontier

Allison Duprey, Fanuel Sisay, Natasha Stewart, and Yangxinyu Xie

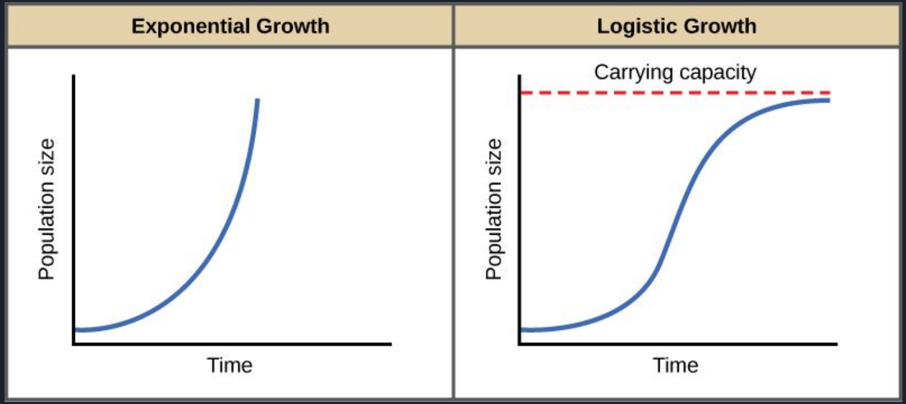
Postdoc Mentor: Dr. John Nardini
SAMSI 2019

Overview

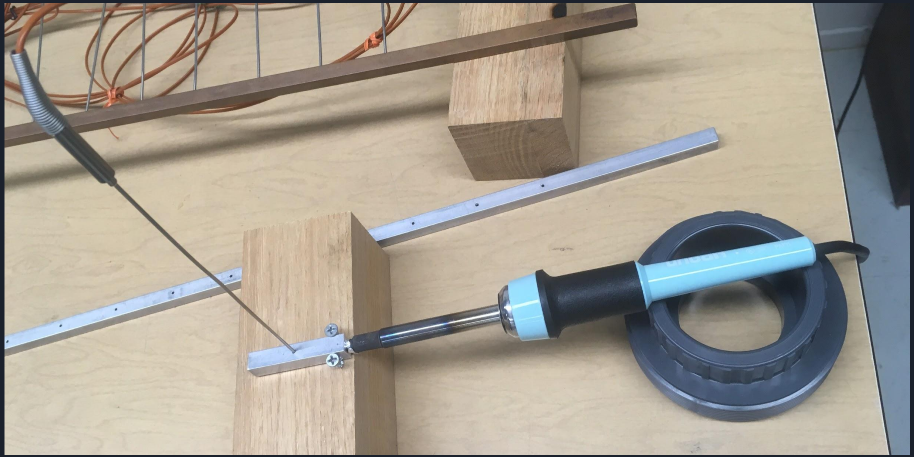
- Motivation
- Background
- Methods
- Results
- Limitations & Future Directions

Motivation

- Understanding data
- Common ODE Models
    - Logistic
    - Exponential



Motivation



- Heated rod experiment
- Sparse regression (SINDy)

Background - **SINDy**

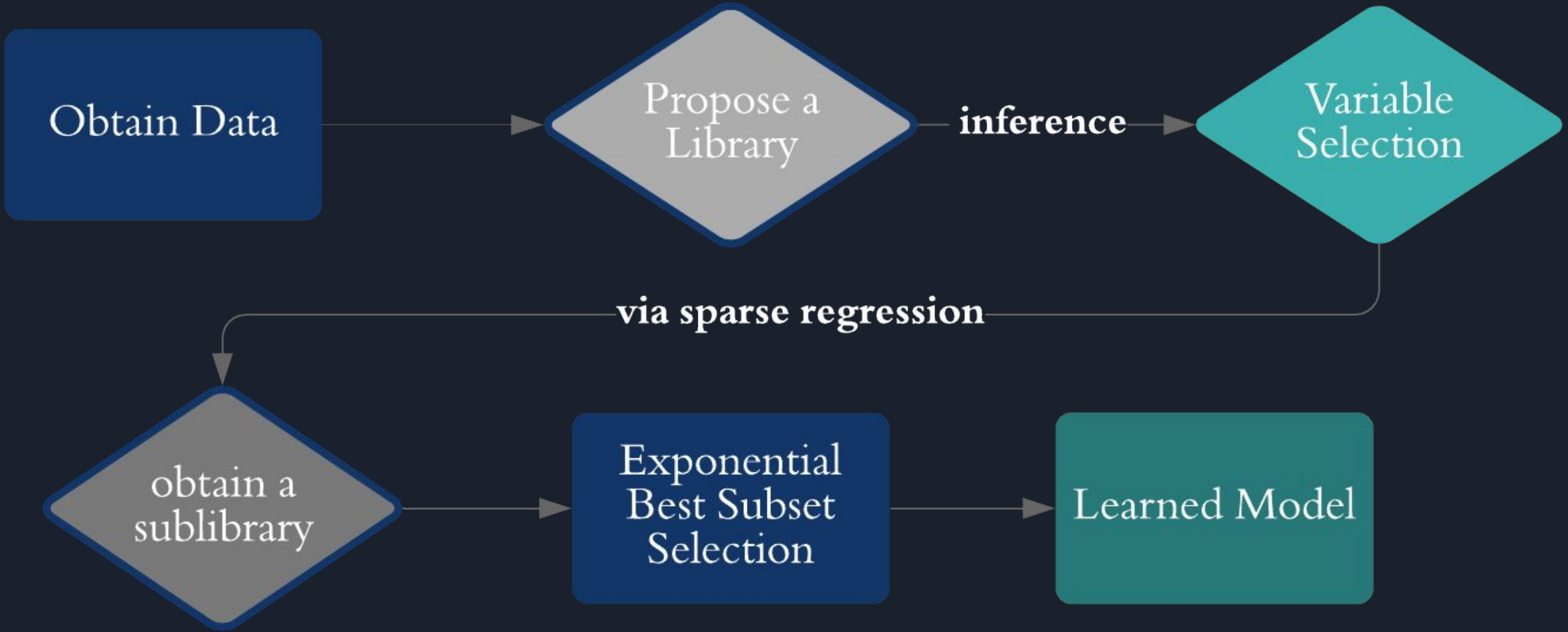
<u>S</u>parse **<u>i</u>**dentification of **<u>n</u>**onlinear **<u>dy</u>**namical systems

A data-driven process to learn an ODE's form -

$$\Theta = [1, y, \ldots y^p, sin(y), cos(y) \ldots]$$

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *113*(15), 3932-3937.

Background - **SINDy**



Background

- Least-Squares Regression
    - Loss Function (Residual Sum of Squares)

$$\hat{\xi} = \underset{\xi}{\operatorname{argmin}} \left(|| \frac{dy}{dt} - \Theta\xi ||_2^2 \right)$$

- Sparse Regression
    - Loss Function + Penalty for Large Coefficients
    - Prevent Overfitting

$$\hat{\xi} = \{\text{Loss}\} + \lambda\{\text{penalty}\}$$

Background

- Ridge Regression
  - L2 Penalty

$$\hat{\xi} = \underset{\xi}{\operatorname{argmin}} \left(|| \frac{dy}{dt} - \Theta\xi||_2^2 + \lambda||\xi||_2^2 \right)$$

- LASSO
  - L1 Penalty

$$\hat{\xi} = \underset{\xi}{\operatorname{argmin}} \left(|| \frac{dy}{dt} - \Theta\xi||_2^2 + \lambda||\xi|| \right)$$

Background

- Best subset selection

$$\text{AIC} = -\frac{2}{N}\log \text{lik} + 2\frac{d}{N}$$

$$\text{BIC} = -2\log \text{lik} + d\log N$$

We are not aware of any study that has performed a full analysis of how LASSO performs when integrated with **SINDy** (as compared to ridge regression and the greedy algorithm).

Methods

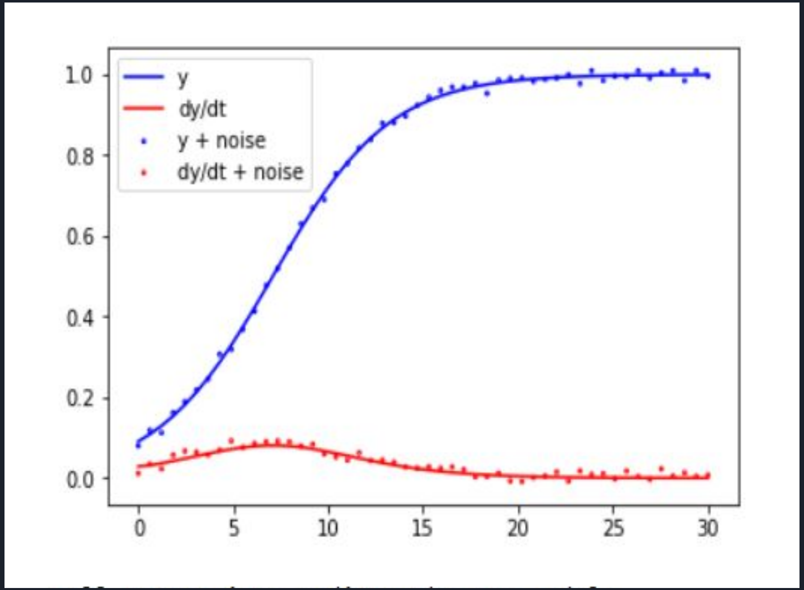
- Simulate data to test SINDy's performance when underlying truths are known
- Assume logistic growth model

$$y = \frac{e^{kt}}{A+e^{kt}}$$

$$\frac{dy}{dt} = ky(1 - \frac{y}{k})$$

- Small amount of gaussian noise

$$\text{Data} = y + \text{Noise}$$



Methods

- We attempt to replicate published methodology using ridge regression with best subsets selection

- We perform all regression in Python using Scikit Learn
    - Optimal hyperparameter and tolerance threshold for ridge regression chosen through 5 fold cross-validation

- After finishing sparse regression, best subsets selection is performed on the selected library terms using AIC/BIC

https://github.com/Xieyangxinyu/SINDy

Methods

- Note that the information criterions have the same general form as the sparse regression algorithms - a loss function with a penalty term

$$\{\text{Loss}\} + \{\text{Penalty}\}$$

$$\text{AIC} = -\frac{2}{N}\log\text{lik} + 2\frac{d}{N}$$

$$\text{BIC} = -2\log\text{lik} + d\log N$$

$$\hat{\xi} = \underset{\xi}{\text{argmin}}\left(||\frac{dy}{dt} - \Theta\xi||_2^2 + \lambda||\xi||\right)$$

$$\hat{\xi} = \underset{\xi}{\text{argmin}}\left(||\frac{dy}{dt} - \Theta\xi||_2^2 + \lambda||\xi||_2^2\right)$$

Methods

Relaxed Lasso

- Use cross-validation to estimate the initial penalty parameter for the lasso
- Apply a second penalty parameter to the selected set of predictor
- Since the variables in the second step have less “competition” from noise variables, cross-validation will tend to pick a smaller value for λ, and hence their coefficients will be shrunken less than those in the initial estimate.

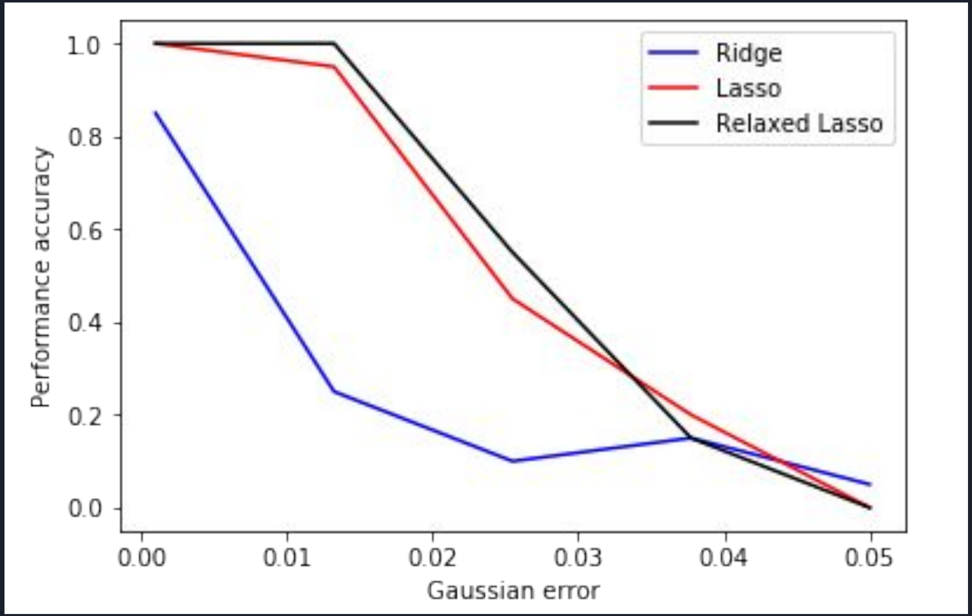
Results

Θ	Ridge CV	Lasso CV	Elastic Net CV	Relaxed Lasso
$u^0, u^1, u^2, u^3, u^4, e^u$	2.86 s	195 ms	221 ms	322 ms
$u^0, u^1, u^2, u^3, u^4, u^5, u^6, u^7$	2.86 s	151 ms	176 ms	306 ms

Results

$$u^0, u^1, u^2, u^3, u^4, u^5, u^6, u^7$$

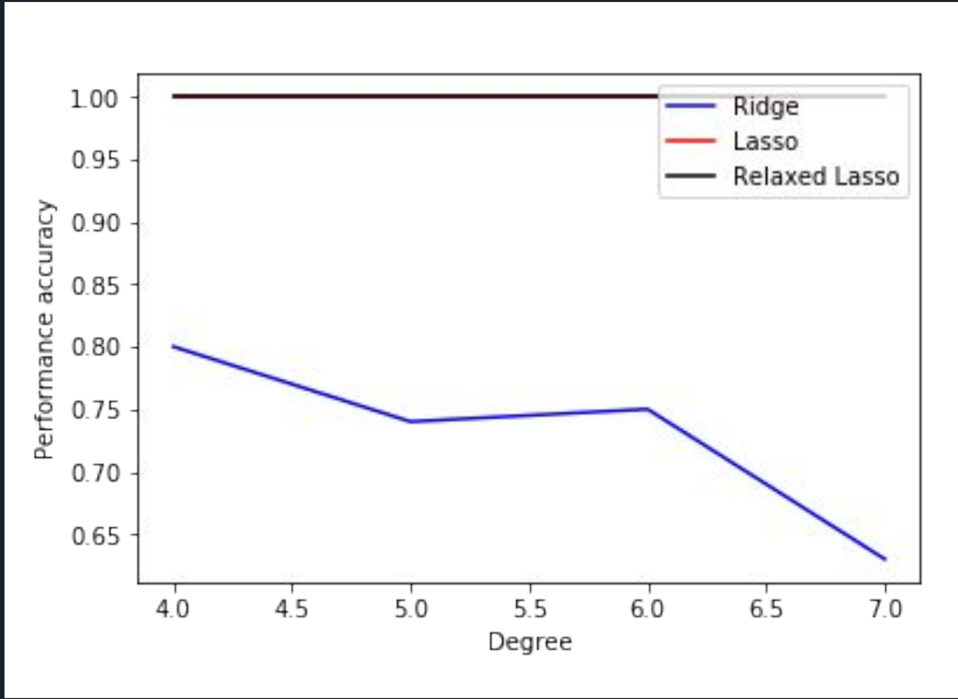
Performance accuracy with increase in the deviance of the Gaussian error



Results

$$u^0, u^1, u^2, u^3, u^4, u^5, u^6, u^7$$

- Relaxed Lasso remains robust throughout as we expand our library from degree 4 to degree 7
- Ridge gets worse

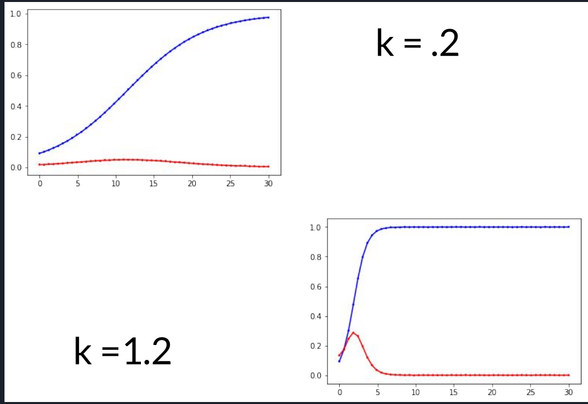


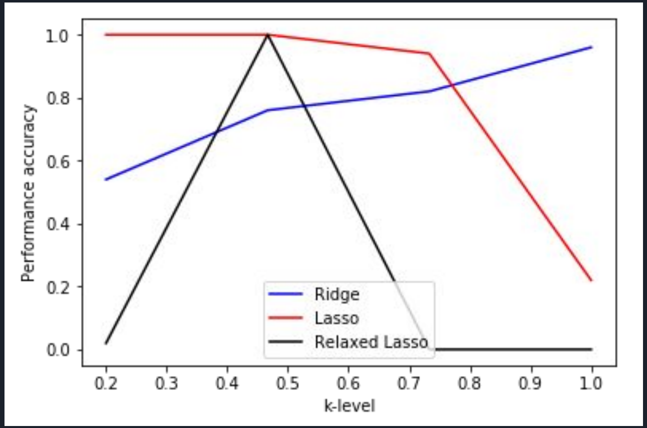
Results

$$u^0, u^1, u^2, u^3, u^4, u^5, u^6, u^7$$

- k represents the range of the data collected
- Suppose we only collect a fixed range of data, the performance of the SINDy model is dependent on the true value of k

$$y = \frac{e^{kt}}{A+e^{kt}}$$

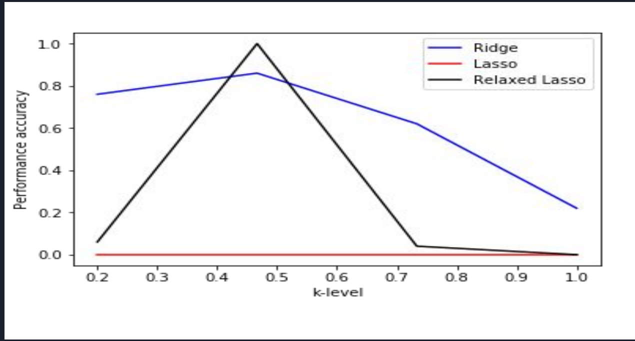


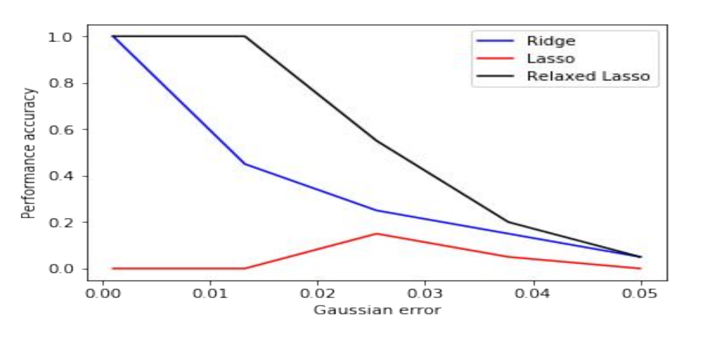


Limitations

- LASSO performs well on high degree polynomial terms but poorly for small order terms. Ridge regression performs well on small order terms but poorly on high degree polynomial terms.
- LASSO performs poorly for exponential terms. Ridge regression performs well, but not as well as relaxed LASSO.

$$u^0, u^1, u^2, u^3, u^4, e^u$$





Limitations

- PCA was surprisingly uninformative
  - One principal component, comprising all non-constant terms in the library, accounted for 98 percent of variation

Alternative Loss and Penalty Functions

Squared Loss:

$$L(y, \hat{f}(X)) = (y - \hat{f}(X))^2$$

Absolute Loss:

$$L(y, \hat{f}(X)) = |y - \hat{f}(X)|$$

Deviance:

$$\begin{aligned} L(G, \hat{p}(X)) &= -2 \sum_{k=1}^{K} I(G = k) \log \hat{p_k}(X) \\ &= -2 \log \hat{p_G}(X) \end{aligned}$$

Hinge Loss:

$$L(y, \hat{f}(X)) = \max\left(0, 1 - y \times \hat{f}(X)\right)$$

Elastic Net Penalty:

$$\min_{\omega \in R^p} \left(\Sigma_{i=1}^{n} l(y_i, \omega^T \times x_i) + \lambda \left[\alpha ||\omega||_1 + \frac{1-\alpha}{2} ||\omega||_2^2 \right] \right)$$

Number of Beta's Penalties

$$\min_{\omega \in R^p} \frac{1}{2} ||y - X\omega||^2 + \lambda_0 ||\omega||_0$$

Conclusion

SINDy is useful for inferring ODE form from data

- Best result: LASSO with cross-validation
- Relaxed LASSO has similar results
- Ridge regression unsuccessful for some libraries

Acknowledgements

Thomas Gehrmann

Dr. Mansoor Haider

Dr. David Banks

John Nardini

Pulong Ma

Wenjia Wang

Xinyi Li

Nikolas Bravo

Fellow SAMSI UGs!

References

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *113*(15), 3932-3937.

Boyd, S. P., & Vandenberghe, L. (2018). *Convex optimization*.

Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis*, *52*(1), 374-393.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Second ed.). New York, NY: Springer New York. doi:10.1007/978-0-387-84858-7

Kutner, M. H. (2005). *Applied linear statistical models* (5th ed.). Boston: McGraw-Hill Irwin.

Mangan, N., Brunton, S., Proctor, J. and Kutz, J. (2016). Inferring Biological Networks by Sparse Identification of Nonlinear Dynamics. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 2(1), pp.52-63.

Mangan, N. M., Kutz, J. N., Brunton, S. L., & Proctor, J. L. (2017). Model selection for dynamical systems via sparse regression and information criteria. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *473*(2204), 20170009.

Scikit-Learn: Machine learning in Python - scikit-learn 0.16.1 documentation. Retrieved from https://scikit-learn.org/

Suparman, & Rusiman, M. S. (2018). Bootstrap-based model selection in subset polynomial regression. *International Journal of Advances in Intelligent Informatics, 4*(2), 87-94. doi:http://dx.doi.org.ezproxy.lib.utexas.edu/10.26555/ijain.v4i2.173

Questions?