Dynamics and Model Discovery

AMATH563: Homework 2

Dan Gorringe
dngrrng@uw.edu

May 7, 2020

Sec. I. Introduction and Overview

From a limited dataset of 20th century esoteric Canadian wildlife populations, several models are derived. Using Dynamic Mode Decomposition, time-embedding, Lotka-Voltera fitting, and SINDy models are compared using information metrics. The report ends with a brief furore into a beautiful chemical reaction - an amazing case study of DMD utility.

The findings come from a Python Notebook that can be found on my Github[1].

Sec. II. Theoretical Background

Dynamic Mode Decomposition

Dynamic mode decomposition, DMD, is able to find oscillatory modes to fit a time series. Utilising singular value decomposition, and the strongest weighted modes an approximation with few nodes is possible for periodic data.

Standard DMD uses individual instances of observations to produce the oscillatory modes. A super-cool extension is to use time-embeddings as data to learn from. Modes found form a relation between data histories. This can require much larger computations, but is able to find useful DMD fits for classically non cosinusoidal.

SINDy

In order to model data, a dynamical system can be constructed. Sparse Identification of Nonlinear Dynamics, SINDy, uses model selection techniques as discussed in the previous homework to find minimal solutions from a library of potential nonlinear dynamical systems.

Information measures of fit

From Information theory Kullback-Leibler, KL, divergence is borrowed, to compare utility of predicted models. KL-divergence is measure of error using one probability distribution to model another. Unfortunately it is difficult to empirically calculate when learning one distribution for another which is unknown in full.

Akaike Information Criteria, AIC, and Bayesian Information Criteria, BIC, incorporate an approximation of divergence and a factor for model complexity. AIC and BIC respectively are useful in comparing different models for the same system - in independence they mean very little.

Sec III. Algorithm Implementation and Development

AIC and BIC Implementation

To simplify matters, model errors are assumed to be normally distributed - which makes it possible to calculate AIC and BIC measures with residual sum of squares, RSS.

Both are a function of the number of model variables, k, which is a measure of complexity, and n, the number of data points fit.

$$\mathrm{AIC} = 2k + n\ln(\mathrm{RSS})$$

$$\mathrm{BIC} = k\ln(n) + n\ln\left(\frac{\mathrm{RSS}}{n}\right)$$

Video Convolution

The homework data file is 174.9mb - this is insane. For puny python notebooks run on 2015 Macbook Pros to be able to sufficiently fit models the video resolution had to be downsampled. A simple mean average three square filter was applied.

Data Interpolation

The given data has 30 points, there have been 25 different Village People performers, both are simply not enough. To fit a worth-while model more data points are needed for the Snowhow Hare and Canada Lynx between 1845 and 1903.

The interpolation of the dataset does not sit incredibly well - in interpreting the models that follow this has to be remembered. There should be no defining behaviour at a scale of less than 2 years, else the interpolated data is taking precedent over the true dynamics being modelled.

A cubic interpolation is used to find 1024 points.

Sec IV. Computational Results

Sec V. Summary and Conclusions

Question 1: Develop a DMD model to forecast the future population states

As is good practise, before running DMD the SVD eigenvalues are observed to determine a suitable truncation. Find the trivial plot in Figure 1

[1] https://github.com/DanGorringe/AMATH563_inferringComplexSystems

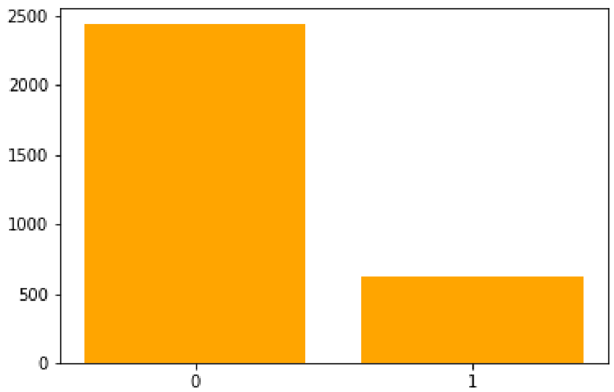

Figure 1: Both eigenvalues returned from SVD of data for Question 1.

It is found that using a truncation of two will likely suffice. DMD is then run, and the time series is predicted and plotted in Figure 2.

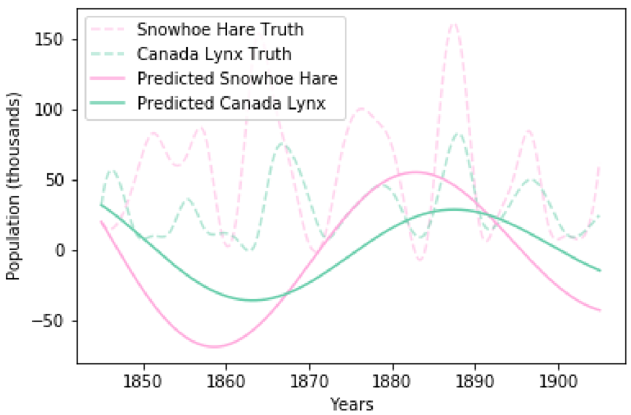

Figure 2: Plot of DMD model upon the training data.

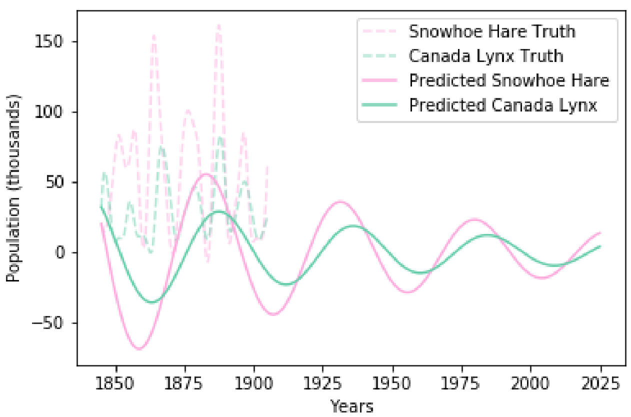

Figure 3: Using the same DMD from Figure 2 to predict into the future.

Figure 3 continues into the future. DMD does not extrapolate well - a model for Snare and Lynx population tending to equilibria is undesirable. DMD is unable to usefully model in this manner as the dynamics are not suitably sinusoidal.

Question 2: Do a time-delay DMD model to produce a forecast and compare with regular DMD. Determine if it is likely that there are latent variables.

Using a time delay, lets you find more intricate behaviour. Though model simplicity is sacrificed, there is a difficult decision in how much embedding to do. After some investigation I decided upon embedding 150 previous timestamps.

By obligation the SVD is plotted, see Figure 4. It is found that the first 8 modes contribute significantly, and hence can use this number for truncation in the following DMD.

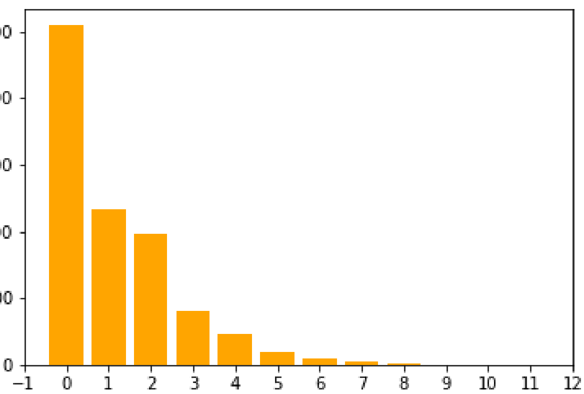

Figure 4: Plotting only the first 12 eigenvalues of the time delayed Hankel.

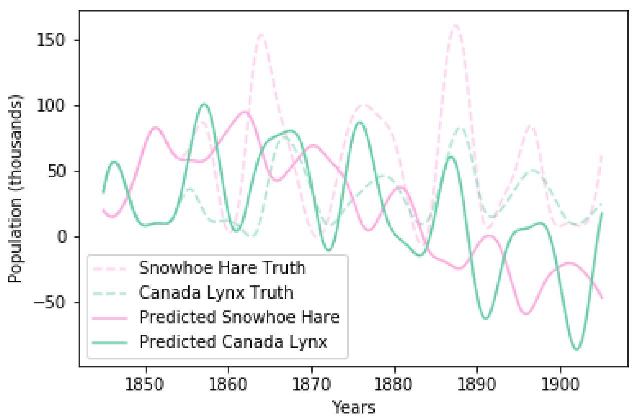

Figure 5: Plot of DMD model upon the training data.

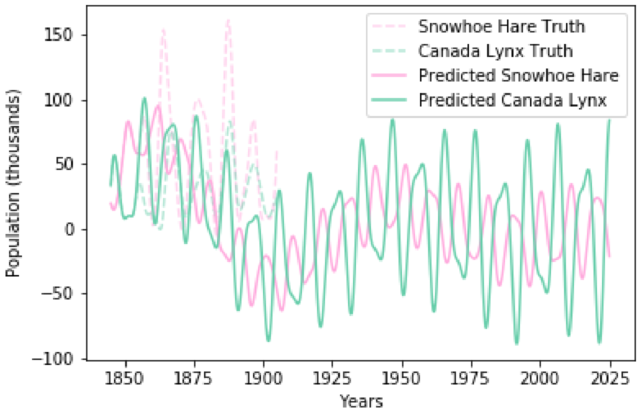

Figure 6: Using the same DMD from Figure 5 to predict into the future.

Figure 6 shows that a relationship between Hare and Lynx is found - and the model makes oscillatory predictions in the future. This would suggest that the model is doing what we want.

Per the question - A time delay is better than the solution from problem 1 because is manages to find seemingly characteristic periodicity. Latent variables presumably exist, from the Sigma plot, they have been exploited by truncating.

Question 3: Empirical Predator-Prey models such as Lotka-Volterra are commonly used to model such phenomenon. Consider the model $\dot{\mathbf{x}} = (\mathbf{b} - \mathbf{p}\mathbf{y})\mathbf{x}$ and $\dot{\mathbf{y}} = (\mathbf{r}\mathbf{x} - \mathbf{d})\mathbf{y}$. Use the data to fit values of b, p, r and d.

After a derivative is found for your data, this becomes a simple model fitting exercise. A central difference method is used to approximate. Fancy improvements could use Numpy's derivative function - it would be interesting to see the impact of derivatives on information metrics.

On fitting a model, one has all the parameters to produce derivatives for all time. Figure 7 shows a relatively well fit model.

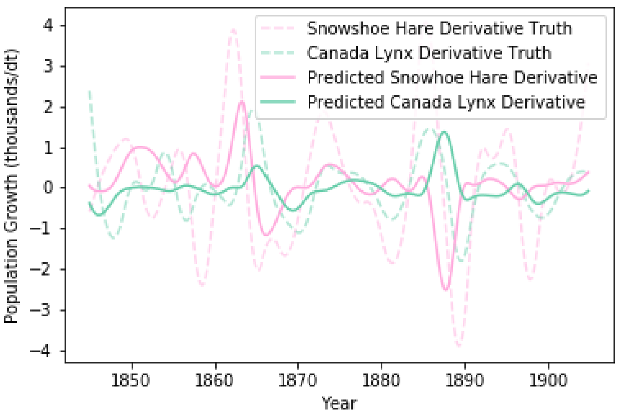

Figure 7: The derivative model predictions versus the learning derivatives.

To compute the population over time I simply incremented initial conditions by the predicted derivative. Better methods exist, extra time to implement does not.

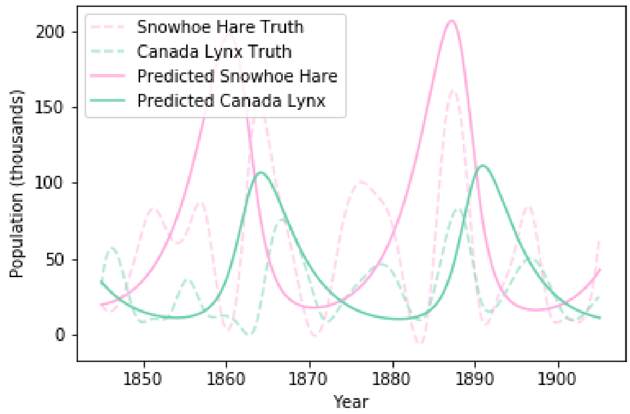

Figure 8: The derivative model predictions versus the learning derivatives.

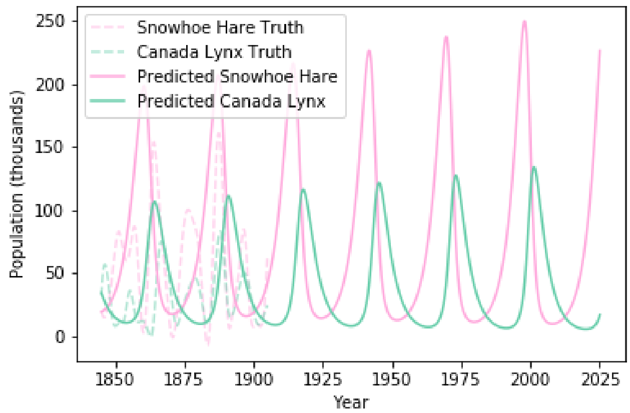

Figure 9: The derivative model predictions versus the learning derivatives.

It's difficult to say that the Lotka-Voltera model found accurately follows the interpolated data given. However it certainly behaves in exactly the manner we'd expect, and therefore it's future predictions are likely some of the most reliable.

Question 4: Find the best fit nonlinear, dynamical systems model to the data using sparse regression.

In the first of university I had a group project to investigate wolf and elk populations - a predator-prey data modelling project. The library you can use can be fantastical - though from this previous project I have grounded myself.

Without further research the only trig-like behaviour in species I know is seasonal. The underlying data is from every two years, so it would be improper to introduce a sub one year term. I have no faith in Lynx nor Hares, so dare not try out exponential terms either. In conclusion I am boring and have only create a library using polynomials, for the following plots up to cubics are used.

Lasso, like in the previous homework, is used in order to minimise the number of coefficients and hence terms in the library used. Figure 10 is proof of sparsity, few terms are present.

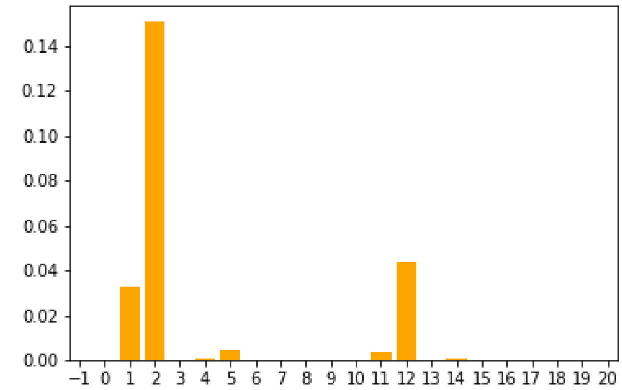

Figure 10: The absolute value of coefficients int he nonlinear dynamical system model of up to cubic term polynomials.

The nonlinear system found does not seem to be very good, see Figure

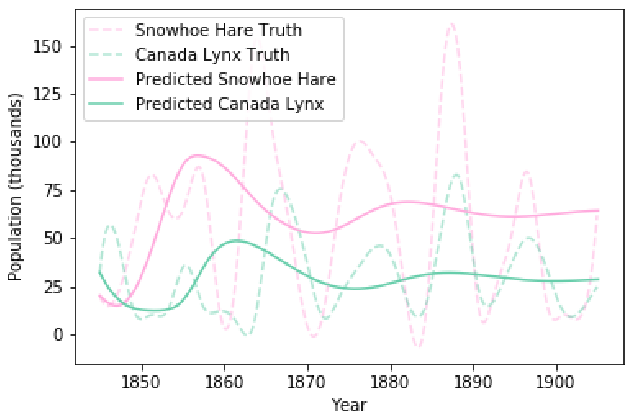

Figure 11: The derivative model predictions versus the learning derivatives.

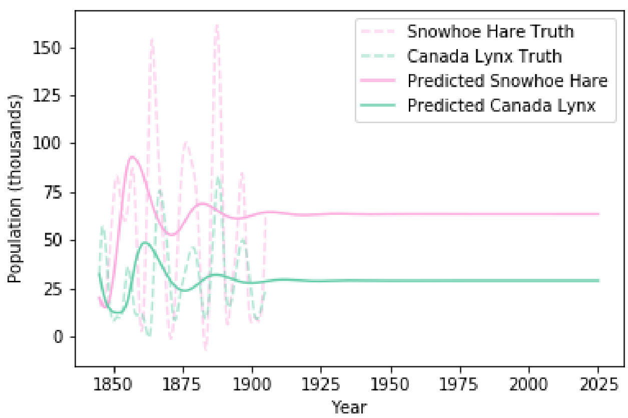

Figure 12: The model predictions versus the learning population.

A clear conclusion is possible - that this is not a model we'd choose to keep around by choice.

Question 5: Compute the KL divergence of the best model fit to the data between all the above models.

Before computing the KL-divergence, each model needs an associated probability distribution. Using a similar method to the course text, probability density functions are created.

Then using the `scipy` entropy routine, KL divergence is computed for each of the models.

DMD	4.12
DMD (time delayed)	1.38
Lotka-Voltera	3.13
Non-linear	3.13

The time-delayed DMD performs best, whilst the pitiful sine waves from regular DMD produce the worst probability density function to model the function created using the training data.

Problem 6: Retain three of your best fit models and compare their AIC and BIC scores.

Each model had their AIC and BIC calculated.

Model	AIC	BIC
DMD	15068	7980
DMD (time delayed)	15080	8021
Lotka-Voltera	16049	8954
Non-linear	14934	7877

I do not have confidence that this are entirely accurate, but it's currently 3am and timezones aren't helping with deadlines. I was expecting the time delayed DMD to perform the best - I would go as far to say these are inversely proportional to what I expected.

Problem ?: Play around with a massive video file.

Despite the ominous huge odely formatted data, this problem turned out to be pretty simple to implement and incredibly satisfying. Having, last quarter, just finished a project investigating stability of Autoencoders I had functions to convert to GIFS to hand.

To decrease the compute required the images were scaled down by a factor of 3.

The obligatory peak at SVD sigmas shows that there are many many important modes. I will win no awards for kidding anybody that I only care about recreating the Belousoz-Zhabotinsky reaction to enough accuracy to model a visualiser for the Rain Song by Led Zeppelin. That is to say that I picked 30 modes, because that roughly looked right from a few other plots. See Figure 14 as proof I've kinda done what I said.

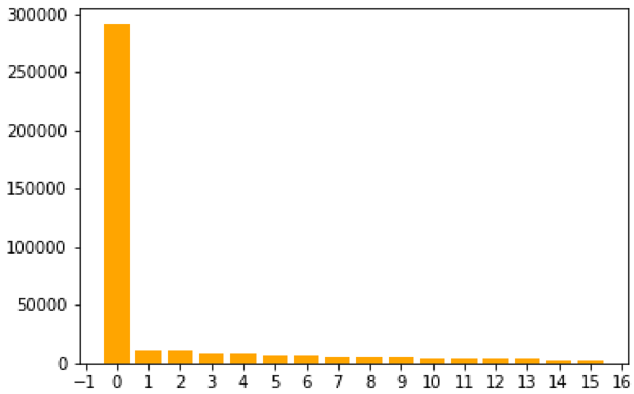

Figure 14: Obligatory eigenvalue plot for BZ data.

Using the classical DMD method worked staggeringly well. My intuition is telling me that cosines are able to to model 2d circles sent to vectors better than they have any right to. On consideration of this thought, it does also look that in the predicted version we see more perfect circles formed.

Projecting video onto paper is a problem as old as time, with a page limit a montage will have to suffice, see Figure 13.

It would be super fun to create a colour map, and get these integrated with the tempo from Spotify's API. I'd also like to see how well DMD works on videos of other pretty random quasi-noise, like a flame or CIV V end-slate graphs.

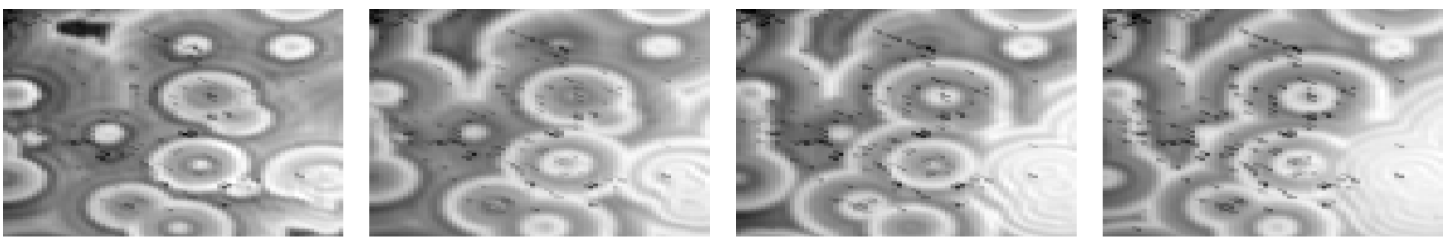

Figure 13: Stills for the predicted video.