Antal Jakovác
Department of Computational Sciences

Uncovering Hidden Laws in Time Series

MT Kurbucz
P Pósfay
A Telcs
TS Biró

Understanding the Diversity of Financial Risk
Machine Learning for Prediction and Optimisation
Budapest, Hungary, November 24, 2023

Introduction

Motto:

πάντα ῥεῖ

(Ἡράκλειτος)

panta rhei: everything flows

(Heraclitus)

Introduction

Motto:

πάντα ῥεῖ

(Ἡράκλειτος)

panta rhei: everything flows

(Heraclitus)

We get all information in time, but …

nothing remains the same in time:

(example: a pen)

- move position
- change shape
- change matter content
- change microstate or quantum state

What we refer to as “a pen” is in fact a set of states.

Introduction

What defines this set? → not an easy question: we use *intelligence* to answer it

Consider visual information:

- the image is represented by pixels: all possible images Ω
- pen images form a subset $P \subset \Omega$
- **possible way**: seek the indicator function $I_P(x) = \begin{cases} 1 \text{ if } x \in P \\ 0 \text{ if } x \notin P \end{cases}$
- we need a universal function approximator: e.g. neural networks
- **technically:**
  - trial function with a lot of parameters $f(x; \alpha_1 \ldots \alpha_n)$
  - adjust parameters on examples until $f(x; \alpha_1^* \ldots \alpha_n^*) \approx I_P(x)$ in the training set.

Introduction

Neural networks provided a lot of magnificent achievements

- classification (dog breeds, faces, birdsong, flowers, etc.)
- text generation (chatGPT, Bing, Bard, Vicuna23B, etc.)
- image generation (midjourney, Dall-E, Dreamstudio, etc.)
- autonomous cars / AI driving assistants
- etc.

in human psychology:

System I:
fast, automatic, intuitive way of thinking

… they are not apt for some problems

- pen or not pen (they need balanced datasets)
- generalization (“catastrophic forgetting”)
- error control (prone to adversarial attacks)
- planning

we would need **System II**:
slow, conscious, deliberate way of thinking

Persistent concepts

Start from the beginning: all information come to us as time series

- everything flows, but we want to ensure our existence for a longer time period
- we have to do predictions
- for prediction we need quantities that are the same in the future as in the past

We need to find **conserved quantities** (laws) in the observed time series!

- These laws are inherent properties of the time series, no need to teach (unsupervised → supervised learning, when the label is part of the time series)

Persistent concepts: science

All of our concepts come from conservation laws!

In science:

- strong intermolecular forces → solid states → **objects** with solid contour are important
  lived we in a gaseous environment, the ”object” concept would be useless
- complicated mechanical motions (e.g. double pendulum) → persistent relation between **force** and **acceleration** (Newton law) $ma - F = 0$
- gas molecules take different configurations, but what matters macroscopically are the (quasi) conserved quantities: **volume, temperature, particle number**, etc
- to characterize animals we consider persistent properties → **species, breeds**
  taxonomy collects these properties in a hierarchical system

Persistent concepts: finance

All of our concepts come from conservation laws:

Finance: fluctuating prices $S(t)$ make predictions difficult;
goal to find some persistent concepts

- price distribution is more stable $P(S, \dot{S}, \ldots)$ → **market model**
- to achieve persistency we combine assets → **hedging, indexes**
- lack of arbitrage, extensive hedging → **risk neutral** market, **martingales** and **pricing**
- for better predictive power we need to find more conserved quantities!

Persistent concepts: hierarchy

Remarks:

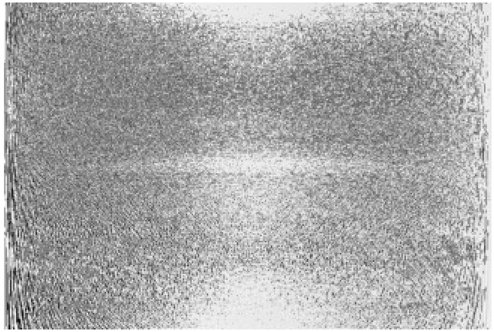
- **hierarchy of concepts**: if we observe different laws, we can relate them by coexistence or consecutiveness → new laws
(e.g. “fire – hot = 0” or “[rain now] – [wet street later] = 0”)
- humans tend to establish these relations easily → not always correct, needs later refinement! (e.g. “[wet street now]-[rain earlier]=0” not necessarily true)
- **context**: we may have a lot of persistent concepts, but not all are important changing context → change relevant concepts (renormalization group)
- **number of relevant concepts**: few in science, can be very large in a realistic environment (e.g. in image recognition)

Persistent concepts: representation

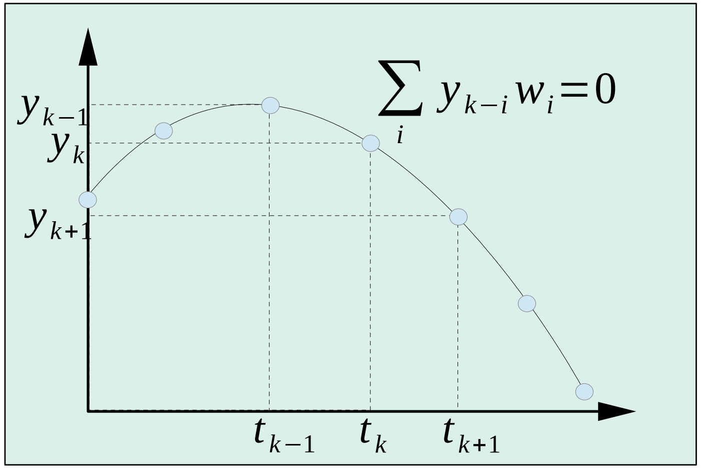
mathematically a law can be expressed as

$$f(x)=0\,,\ \ \forall\ \ x\in X$$

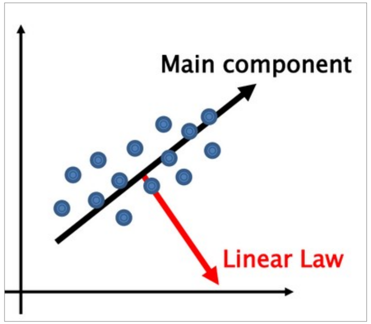
- problem with the representation: assume digital input, i.e. $x\in X=\{0,1\}^N,\ \ |X|=2^N$
- then the number of possible $X\rightarrow\{0,1\}$ functions is $2^{|X|}$: **impossible to cover**!
- we *single out some functional space*: shapes, textures $\rightarrow$ humans are good
- mathematically simplest: **linear laws** $\rightarrow$ we are not prepared to recognize it



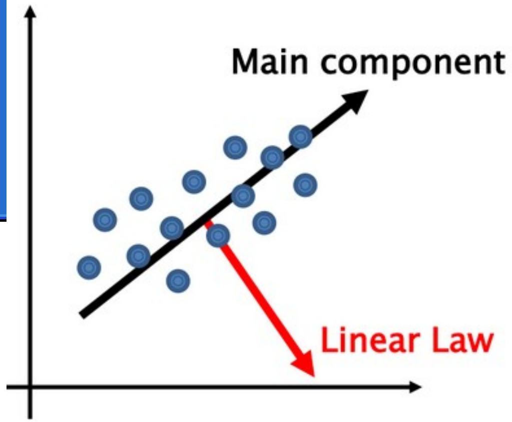
Linear laws in time series



- simplest form: linear relation
- **Mathematical procedure**:
  - Measure time series data: $y_k = y(t = k\Delta t)$
  - Embedding: $Y_{ki} = y((k-i)\Delta t), \quad i = 1 \dots l$
  - Linear map with parameters: $F(Y_k) = \sum_i Y_{ki} w_i = (Yw)_k$
  - Linear law: $F(Y_k)=$minimal for all *k* $\Rightarrow$ $\sum_k |(Yw)_k|^2 = w^T (Y^T Y) w =$minimal
  - Condition $|w| = 1$ leads to an eigenvalue problem: $Y^T Y w = \lambda w$
  - We need the ***smallest*** eigenvalue (in contrary to PCA)
  - example: $y = \sin(\omega t + \phi)$ $\Leftrightarrow$ $y_{k+1} - 2 y_k \cos(\omega \Delta t) + y_{k-1} = 0$



Linear laws in time series

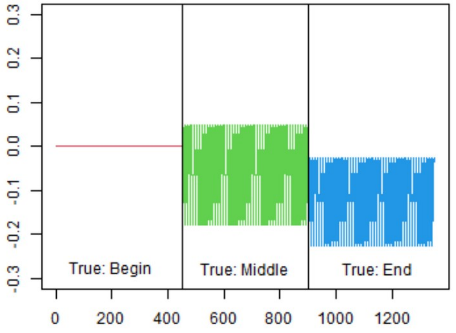


- Linear laws are usually not exact
- We use them as feature transformation $y_i \rightarrow \xi_k = (Y\, w)_k$
  - if y is in a class, and w is the law of the class: new features are small
  - if y is in a class, and w is the law of another class: new features are large
- Collect linear laws for different classes, compute transformed features
- Use standard classifiers (KNN, RF, SVM) on the new feature set.

Application: AReM database

MT Kurbucz, P Pósfay, AJ, Scientific Reports 12 (1), 18026

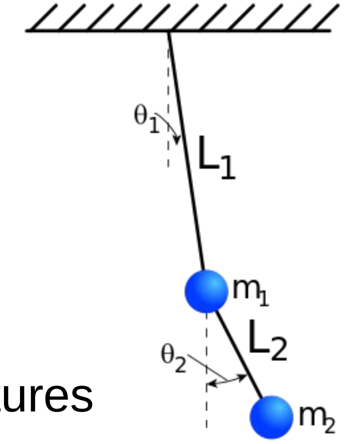
- Activity Recognition system based on Multisensor data fusion (AReM) Data Set
  - 7 motion classes (bending, lying, cycling,etc.)
  - 3 sensor data → 6 features (mean and variance)
  - 88 time series (instances), 480 values in each



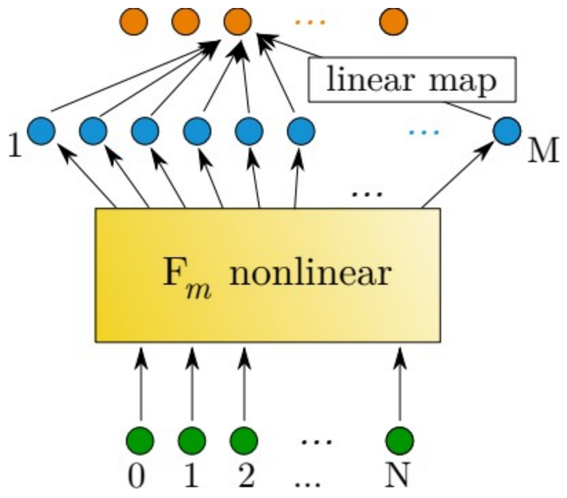
- Mehtod: LLT (Linear Law based feature Transformation)
  - Determine the laws for each instances and channels in the training sets
  - Apply them to the test series, take temporal average/variance → features
  - Train a classifier on the results (KNN, DT, SVM)
  - KNN provides **error-free classification**

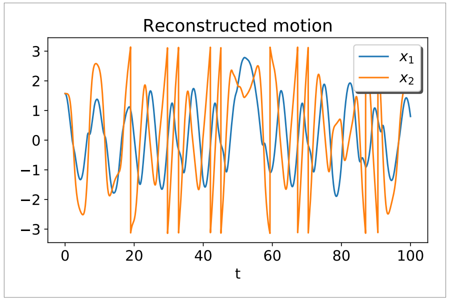
Nonlinear laws

AJ, MT Kurbucz, P Pósfay, New Journal of Physics 24 (7), 073021



- Generalization: input are not directly the embedded data, but pre-trained features
- F_m can be represented by (deep) neural network
- Extreme learning: the exact form of F_m does not matter
- Reconstruction of mechanical motions: 3-leg embedding (discrete Newton-equations)
- Chaoticity, stability → recursion to reconstruct motion

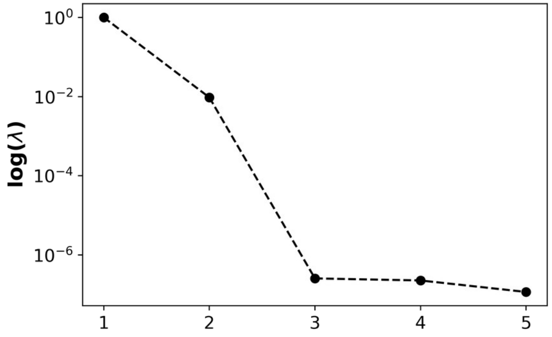


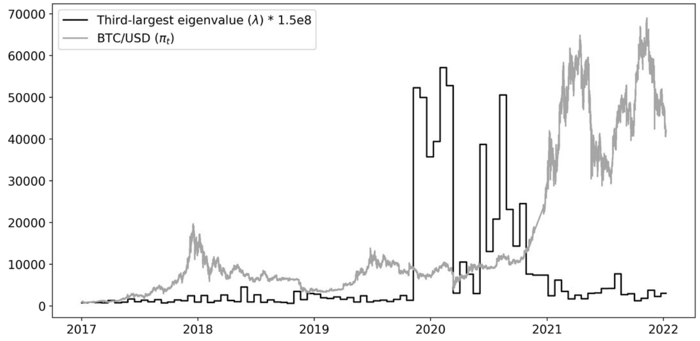


Stochastic processes

MT Kurbucz, P Pósfay, A Jakovác, arXiv preprint arXiv:2201.09790

- Markov chains: stochastic process where $P^{(n+1)}(x)=\sum_y T^{(n)}_{x,y} P^{(n)}(y)$ $\Rightarrow$ $P^{(n+1)}=T\,P^{(n)}$
- In equilibrium (steady state) no *n* dependence, for equilibrium distribution: $P=T\,P$
- 2-variable correlation functions: $\langle f(x_n, x_{n+k})\rangle = \mathrm{Tr}(F\,T^k)$ where $F_{xy}=f(x,y)P(x)$
- These satisfy linear laws: $\sum_k \langle f(x_n, x_{n+k})\rangle w_k = 0$ if $\sum_k w_k T^k = 0$ characteristic polynomial
- Dimensionality of the Markov process can be determined from the laws





Predicting cryptocurrency price movement

MT Kurbucz, P Pósfay, A Jakovác, arXiv:2305.04884

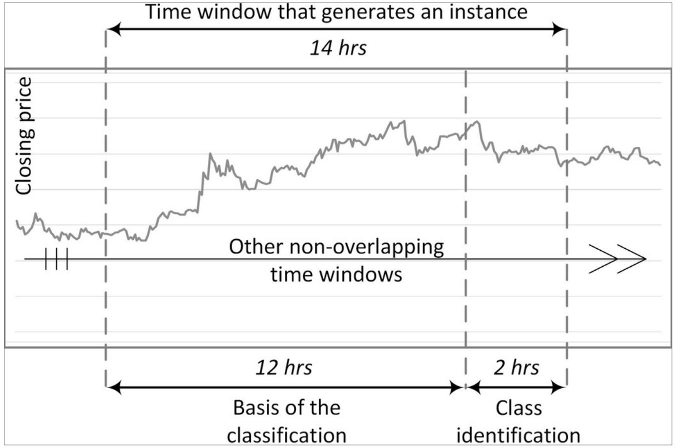
- database: cryptocurrency 1min prices from January 1st 2019- October 22nd 2022
- prediction: 12hrs → price after 2hrs larger/smaller (0/1)
- data: k=720 data (12*60), 6 features (price quotes), n instances (non-overlapping)
- collect linear laws for training instances for both classes
- for test data apply all collected laws, take temporal mean
- use standard classifiers on transformed feature space

(a) Original feature space

	BTC	ETH	BNB	XRP
Ensemble	56.5	55.8	52.3	56.2
KNN	57.0	55.8	57.5	57.3
DT	57.0	54.2	50.8	52.4
SVM	59.6	56.1	57.5	54.5

(b) Feature space transformed by LLT

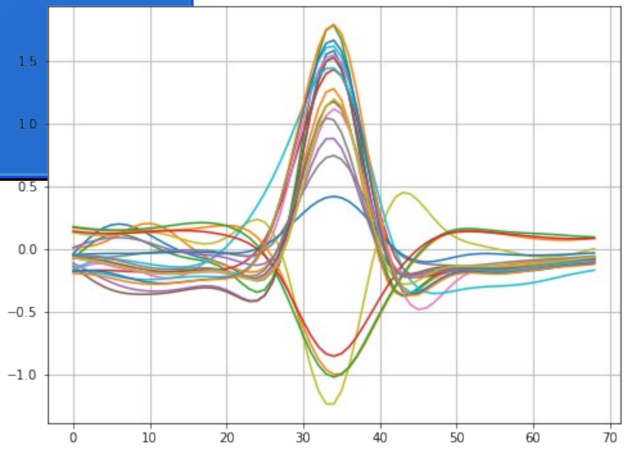
	BTC	ETH	BNB	XRP
Ensemble	75.2	80.8	70.4	79.5
KNN	84.3	82.0	77.6	81.4
DT	65.9	73.6	60.8	67.5
SVM	65.9	64.3	58.8	62.0



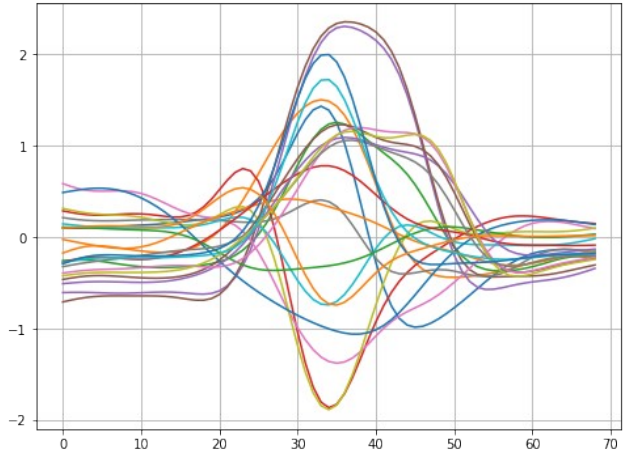
Conclusions

- we need to maintain ourselves (give predictions) in an ever flowing environment
- thus we need to find persistent phenomena (laws): these are our **concepts**
- we need a functional space to seek laws: simplest is the **linear space**
- **LLT: Linear Law based feature Transformation:** find laws for different classes, and transform the time series with them → *very powerful for certain class of problems*

Application: ECG analysis



- Goal: classify heart beats into normal and ectopic
- ECG signal: cleaning, standardizing
- Mehtod: prepare test, validation and training sets
  - Find linear laws for the QRS complex (11 leg embedding, universal laws)
  - Train a classifier on the results (KNN, RF, SVM)
  - Results depend on several factors, best result SVM: 94.3% (close to state-of-art results)
  - More data could help to improve accuracy
- Can be used in a non-annotated dataset (self annotation)



The end