(Multiparameter) Iterated sums

J. Diehl (University of Greifswald)
(joint works with K. Ebrahimi-Fard, N. Tapia, L. Schmitz)

March 13th '24

Overview

Detecting patterns in time series

Detecting patterns in images

Toy example 'Consecutive'

Classification problem. Both classes are noise + pattern.

Class A The pattern $[2, -3, 16]$ is added somewhere, consecutively, to white noise.

Class B A random, non-identity, permutation of $[2, -3, 16]$ is added somewhere, consecutively, to white noise.

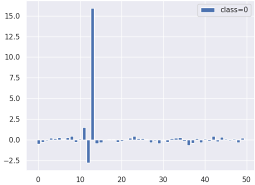


Fig: Class A

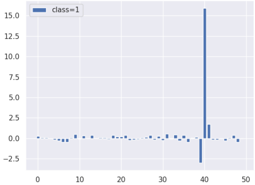


Fig: Class B

A fully connected NN can, of course, deal with this (variance of white noise = 0.01)

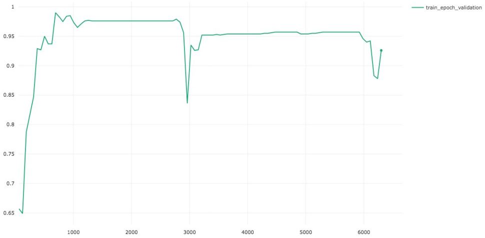


Fig: FCN (Params: 13762)

A fully connected NN can, of course, deal with this (variance of white noise $= 0.01$)

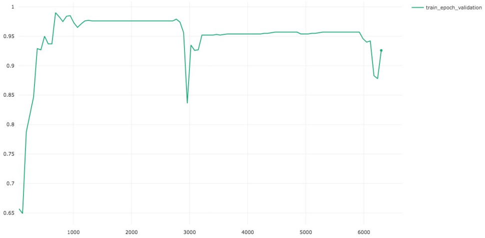
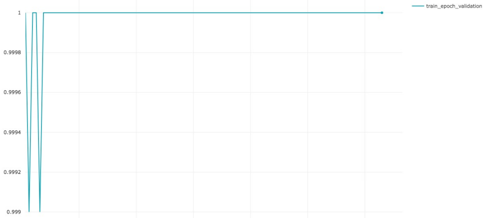


Fig: FCN (Params: 13762)

The data suggest that a CNN should do very well, and indeed



Reminder on Convolutional Neural Networks (CNNs) (stride 1, kernel size 3, maxpooling everything):

Input: $x_i, i = 1, \dots, N$

Output:

$$\max_{i=1,\dots,N-3} \mathrm{ReLU}(\alpha_0 + \alpha_1 x_i + \alpha_2 x_{i+1} + \alpha_3 x_{i+2})$$

or more generally

$$\max_{i=1,\dots,N-3} \phi(x_i, x_{i+1}, x_{i+2}).$$

`https://www.dropbox.com/s/dtp0rwghlj6vwe6/1D-consecutive.gif?dl=0`

The advantage of the inductive bias of CNNs on this problem becomes apparent when increasing variance (here $= 1.0$)

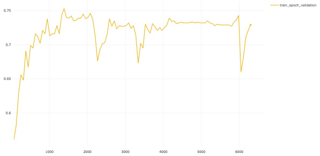


Fig: FCN (Params: 13762)

Fig: CNN (Params: 68)

Toy example 'Nonconsecutive'

Classification problem. Both classes are noise + pattern.

Class A The pattern $[2, -3, 16]$ is added at random timepoints $t_1 < t_2 < t_3$, to white noise.

Class B A random, non-identity, permutation of $[2, -3, 16]$ is added at random timepoints $t_1 < t_2 < t_3$, to white noise.

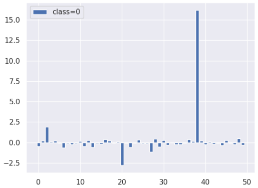


Fig: Class A

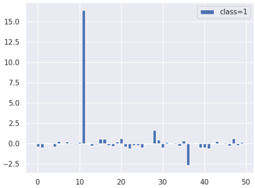


Fig: Class B

Fully connected does alright,

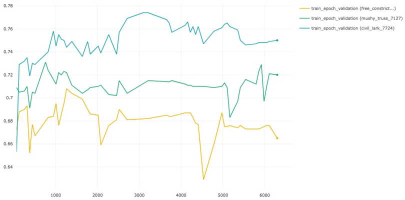


Fig: FCN, N=100 (Params: 13762)

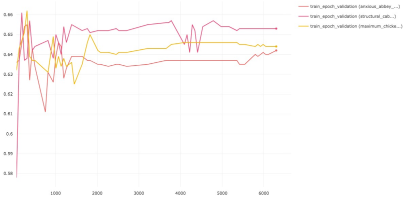


Fig: FCN, N=300 (Params: 13762)

Unsurprisingly the CNN from before completely fails

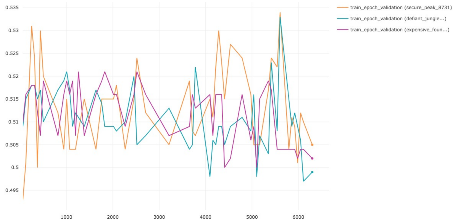


Fig: CNN (Params: 68)

Can we adapt the CNN architecture to this problem?

We had,

$$\max_{i=1,\dots,N-3} \phi(x_i, x_{i+1}, x_{i+2}),$$

which could only detect consecutive patterns.

We want to detect the pattern at any three ordered time points, so we are lead to

$$\max_{1 \le i_1 < i_2 < i_3 \le N} \phi(x_{i_1}, x_{i_2}, x_{i_3}).$$

https://www.dropbox.com/s/plg7n4sc5bvvbhl/1D-nonconsecutive.gif?dl=0

Conceptually: ✓

Algorithmically: this involves $\mathcal{O}(\binom{N}{3})$ operations ↯

But! This

$$\max_{1\le i_1<i_2<i_3\le N} \phi(x_{i_1}, x_{i_2}, x_{i_3}).$$

looks suspiciously like a sum,

$$\sum_{1\le i_1<i_2<i_3\le N} z_{i_1} z_{i_2} z_{i_3}.$$

The latter is, of course, an iterated sum

$$\sum_{1\le i_3\le N} \left(\sum_{1\le i_2<i_3} \left(\sum_{1\le i_1<i_2} z_{i_1} \right) z_{i_2} \right) z_{i_3}.$$

Conceptually: ✓

Algorithmically: this involves $\mathcal{O}(\binom{N}{3})$ operations ⚡

But! This

$$\max_{1 \le i_1 < i_2 < i_3 \le N} \phi(x_{i_1}, x_{i_2}, x_{i_3}).$$

looks suspiciously like a sum,

$$\sum_{1 \le i_1 < i_2 < i_3 \le N} z_{i_1} z_{i_2} z_{i_3}.$$

The latter is, of course, an iterated sum

$$\sum_{1 \le i_3 \le N} \left(\sum_{1 \le i_2 < i_3} \left(\sum_{1 \le i_1 < i_2} z_{i_1} \right) z_{i_2} \right) z_{i_3}.$$

$$\sum_{1 \le i_3 \le N} \left(\sum_{1 \le i_2 < i_3} \left(\sum_{1 \le i_1 < i_2} z_{i_1} \right) z_{i_2} \right) z_{i_3}.$$

Can be calulated as follows. First

$$Q_{i_2} := \sum_{1 \le i_1 < i_2} z_{i_1} \qquad \forall 1 \le i_2 \le N,$$

at cost $\mathcal{O}(N)$. Then

$$P_{i_3} := \sum_{1 \le i_2 < i_3} Q_{i_2} z_{i_2} \qquad \forall 1 \le i_3 \le N,$$

at cost $\mathcal{O}(N)$. And then the final sum, also at cost $\mathcal{O}(N)$, giving a total cost of $\mathcal{O}(3N) = \mathcal{O}(N)$, $\underline{\text{instead of } \mathcal{O}(N^3)}$!

Recap: we calculated

$$\sum_{1 \le i_1 < i_2 < i_3 \le N} z_{i_1} z_{i_2} z_{i_3}.$$

with ostensibly $\mathcal{O}(N^3)$ complexity in $\mathcal{O}(N)$ time. We used: associativity and distributivity.

We were lead here by

$$\max_{1 \le i_1 < i_2 < i_3 \le N} \phi(x_{i_1}, x_{i_2}, x_{i_3}). \qquad (*)$$

First, a negative result on that

Proposition

There are functions ϕ such that $(*)$ has true cost $\mathcal{O}(N^3)$.

But, $(\max, +)$ forms a commutative semiring (associative + distributive). Therefore

$$\max_{1 \leq i_1 < i_2 < i_3 \leq N} (\alpha_0 + \alpha_1 x_{i_1} + \alpha_2 x_{i_2} + \alpha_3 x_{i_3}),$$

is calculable with cost $\mathcal{O}(N)$.

Corollary (DEFT '20)

Let $(\mathbb{S}, \oplus_{\mathbb{S}}, \odot_{\mathbb{S}}, \mathbf{0}_{\mathbb{S}}, \mathbf{1}_{\mathbb{S}})$ *be a commutative semiring. Then*

$$\bigoplus_{\mathbb{S}}{}_{1 \leq i_1 < \cdots < i_k \leq N} z_{i_1}^{\odot_{\mathbb{S}} \alpha_1} \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} z_{i_k}^{\odot_{\mathbb{S}} \alpha_k},$$

is calculable in $\mathcal{O}(N)$*-time.*

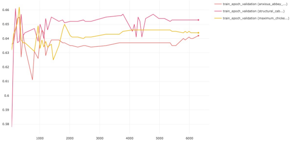


Fig: FCN (Params: 13762)

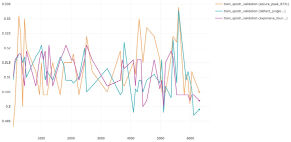


Fig: CNN (Params: 68)

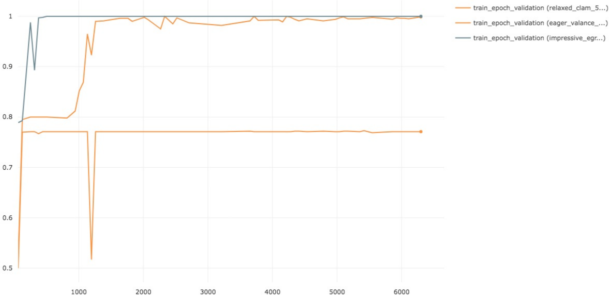


Fig: ISS (Params: 342)

Examples

- Over the ring $\mathbb{R}$

$$\sum_{i_1<\cdots<i_k} z_{i_1}^{\alpha_1} \ldots z_{i_k}^{\alpha_k},$$

$\rightsquigarrow$ iterated-sums signature (quasisymmetric functions)

This has a long history.

- Graham '13 "Sparse arrays of signatures for ...".
- Lyons, Ni, Oberhauser '14 "A feature set for streams ..."
- various works by L Jin et al '15 on Chinese character recognition.
- Kiraly, Oberhauser '16 "Kernels for sequentially ordered data".
- Lyons, Oberhauser '17 "Sketching the order of events".
- D '13, D,Reizenstein '19 on invariant features.
- D,Ebrahimi-Fard,Tapia '19 "Time warping invariants".
- Kidger, Bonnier, Arribas, Salvi, Lyons '19 "Deep Signature Transforms".
- Toth, Bonnier, Oberhauser '20 "Seq2Tens".

In these works it progressively emerged that it is helpful to learn the signature-type features.

Paraphrasing

$$\rightsquigarrow \sum_{i_1<\cdots<i_k} f_{\theta_1}(z_{i_1}) \cdot \cdots \cdot f_{\theta_k}(z_{i_k}).$$

with $f_\theta : \mathbb{R}^d \to \mathbb{R}$.

We propose to boil this down to the bare minimum needed, namely

distributivity and associativity,

to arrive at a richer set of features.

$$\rightsquigarrow \bigoplus_{i_1<\cdots<i_k}{}_{\mathbb{S}} f_{\theta_1}(z_{i_1}) \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} f_{\theta_k}(z_{i_k}),$$

with $f_\theta : \mathbb{R}^d \to \mathbb{S}$.

In these works it progressively emerged that it is helpful to learn the signature-type features.

Paraphrasing

$$\rightsquigarrow \sum_{i_1 < \cdots < i_k} f_{\theta_1}(z_{i_1}) \cdot \cdots \cdot f_{\theta_k}(z_{i_k}).$$

with $f_\theta : \mathbb{R}^d \to \mathbb{R}$.

We propose to boil this down to the bare minimum needed, namely

distributivity and associativity,

to arrive at a richer set of features.

$$\rightsquigarrow \bigoplus_{i_1 < \cdots < i_k}{}_{\mathbb{S}} f_{\theta_1}(z_{i_1}) \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} f_{\theta_k}(z_{i_k}),$$

with $f_\theta : \mathbb{R}^d \to \mathbb{S}$.

Examples

- Over the arctic semiring

$$\max_{i_1<\cdots<i_k} \{\alpha_1 \cdot z_{i_1} + \cdots + \alpha_k \cdot z_{i_k}\}$$

$\rightsquigarrow$ arctic-sums signature
(arctic quasisymmetric expressions [DEFT '20])

Leaving the strict setting of arctic-sums, we can do a learnable version:

$$\rightsquigarrow \max_{i_1<i_2<i_3} \left\{ f_{\theta_1}(z_{i_1}) + f_{\theta_2}(z_{i_2}) + f_{\theta_3}(z_{i_3}) \right\},$$

Summary

Instead of CNNs

https://www.dropbox.com/s/dtp0rwghlj6vwe6/1D-consecutive.gif?dl=0

we 'scan' over all ordered time points

https://www.dropbox.com/s/dtp0rwghlj6vwe6/1D-consecutive.gif?dl=0

And we can do so in linear time.

Algebraic setting

For $z_1, z_2, \dots \in \mathbb{S}$, $s < t$, we define a collection of values in $\mathbb{S}$, indexed by words in the alphabet $\mathbb{N}$,

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \right\rangle := \bigoplus_{\mathbb{S}}{}_{s<i_1<\dots<i_k<t+1} z_{i_1}^{\odot_{\mathbb{S}} w_1} \odot_{\mathbb{S}} \dots \odot_{\mathbb{S}} z_{i_k}^{\odot_{\mathbb{S}} w_k}.$$

For example

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), 537 \right\rangle = \bigoplus_{\mathbb{S}}{}_{s<i_1<\dots<i_3<t+1} z_{i_1}^{\odot_{\mathbb{S}} 5} \odot_{\mathbb{S}} z_{i_2}^{\odot_{\mathbb{S}} 3} \odot_{\mathbb{S}} z_{i_3}^{\odot_{\mathbb{S}} 7}$$

which in min-plus equals

$$\min_{s<i_1<i_2<i_3<t+1} \{5 \cdot z_{i_1} + 3 \cdot z_{i_2} + 7 \cdot z_{i_3}\}.$$

$\mathrm{ISS}^{\mathbb{S}}_{s,t}(z)$ is an element of $\mathbb{S}\langle\langle\mathbb{N}\rangle\rangle$, the space of formal, infinite sums of words (in the alphabet $\mathbb{N}$) with coefficients in $\mathbb{S}$:

$$\mathrm{ISS}^{\mathbb{S}}_{s,t}(z) = \sum_{w} c_w \; w,$$

with

$$c_w := \bigoplus_{\substack{\mathbb{S} \\ s<i_1<\dots<i_k<t+1}} z_{i_1}^{\odot_{\mathbb{S}} w_1} \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} z_{i_k}^{\odot_{\mathbb{S}} w_k}.$$

Theorem (DEFT '20)

1. *(Quasi-shuffle identity)*

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), u \right\rangle = \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \star u \right\rangle$$

2. *(Chen's identity) For $s < t < u$,*

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,u}(z), w \right\rangle = \bigoplus_{\substack{\mathbb{S} \\ w' \cdot w'' = w}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w' \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{t,u}(z), w'' \right\rangle$$

3. $\mathrm{ISS}^{\mathbb{S}}_{0,\infty}(z)$ *is invariant to inserting* $\mathbf{0}_{\mathbb{S}}$ *into* z.

Quasi-shuffle:

$32 \star 4 = 324 + 36 + 342 + 72 + 432$

Theorem (DEFT '20)

1. *(Quasi-shuffle identity)*

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), u \right\rangle = \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \star u \right\rangle$$

2. *(Chen's identity) For* $s < t < u$,

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,u}(z), w \right\rangle = \bigoplus_{w' \cdot w'' = w}{}_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w' \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{t,u}(z), w'' \right\rangle$$

3. $\mathrm{ISS}^{\mathbb{S}}_{0,\infty}(z)$ *is invariant to inserting* $\mathbf{0}_{\mathbb{S}}$ *into* z.

Quasi-shuffle:

$32 \star 4 = 324 + 36 + 342 + 72 + 432$

Theorem (DEFT '20)

1. *(Quasi-shuffle identity)*

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), u \right\rangle = \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \star u \right\rangle$$

2. *(Chen's identity) For $s < t < u$,*

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,u}(z), w \right\rangle = \bigoplus_{w' \cdot w'' = w}{}_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w' \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{t,u}(z), w'' \right\rangle$$

3. $\mathrm{ISS}^{\mathbb{S}}_{0,\infty}(z)$ *is invariant to inserting* $\mathbf{0}_{\mathbb{S}}$ *into* z.

Concatenation:

$32 \cdot 4 = 324$

Detecting patterns in time series

Detecting patterns in images

A CNN works like this:

`https://www.dropbox.com/s/ni6eqtl33e9s707/2D-consecutive.gif?dl=0`

We want:

`https://www.dropbox.com/s/9c3fs7ycocdb0a2/2D-nonconsecutive.gif?dl=0`

Multiparameter iterated sums

For example

$$\left\langle \mathrm{ISS}(z), \begin{bmatrix} 5 & 1 \\ 0 & 3 \end{bmatrix} \right\rangle = \sum_{i_1 < i_2; j_1 < j_2} z^5_{i_1,j_1} z^1_{i_1,j_2} z^0_{i_2,j_1} z^3_{i_2,j_2},$$

or over $(\max, +)$

$$\left\langle \mathrm{ISS}(z), \begin{bmatrix} 5 & 1 \\ 0 & 3 \end{bmatrix} \right\rangle$$
$$= \max_{i_1 < i_2; j_1 < j_2} \left(5 z_{i_1,j_1} + 1 z_{i_1,j_2} + 0 z_{i_2,j_1} + 3 z_{i_2,j_2}\right).$$

Algebraic setting

- quasi-shuffle structure ✓
- tridendriform structure ✓
- ennea (quadriform) structure ✓
- Chen's identity / efficient calculation (blackboard) ?

Thank you!

Two phd positions this year:

- **May 2024** - Counting permutation and chirotope patterns: Algorithms, algebra, and applications
- (**October 2024** - AI and Augmented Reality; funding pending)

Please share with interested students.