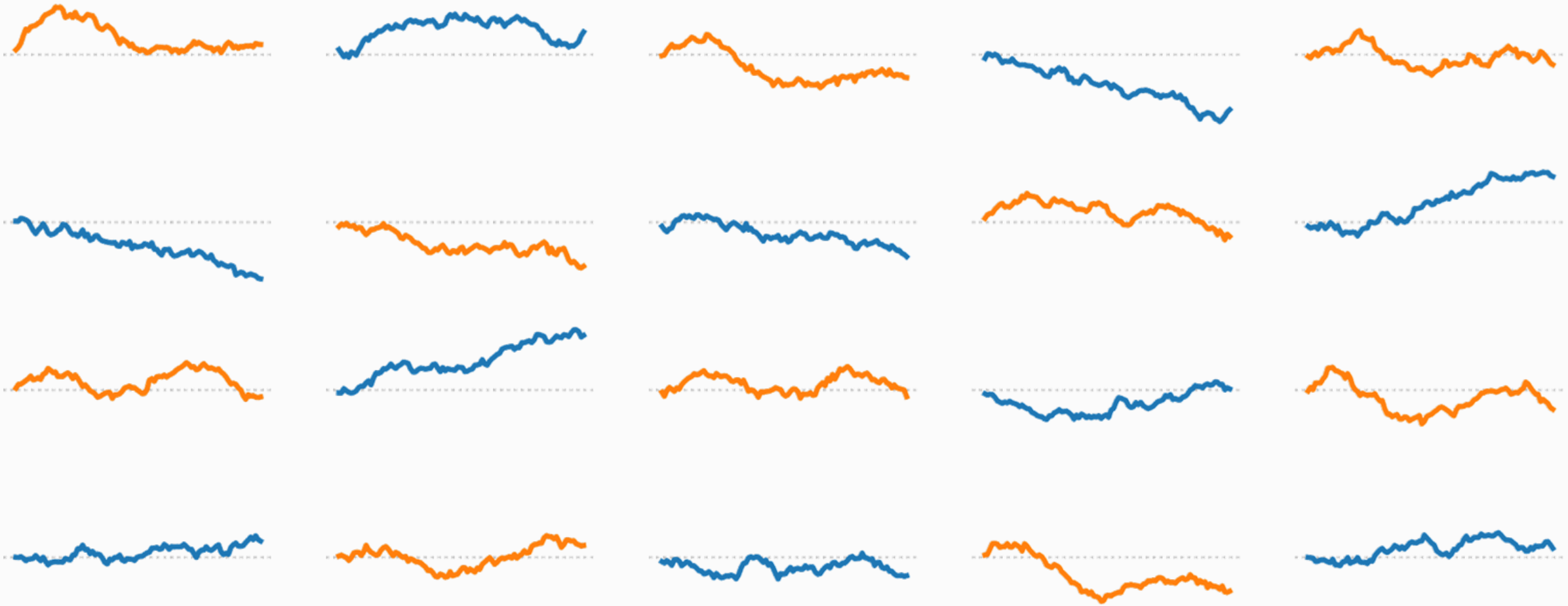
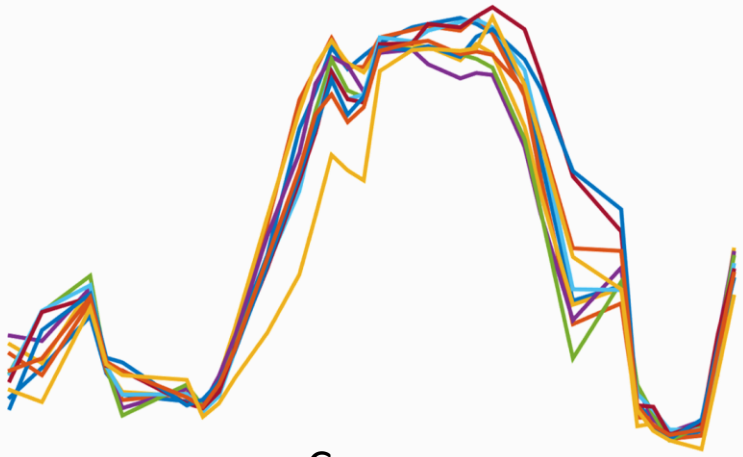
Convolutional Kernels for Efficient and Effective Time Series Analytics

Geoff Webb
Monash University
http://i.giwebb.com
Work with A. Dempster, C. Bergmeir, F. Petitjean, D. Schmidt, C. W. Tan

Time series classification

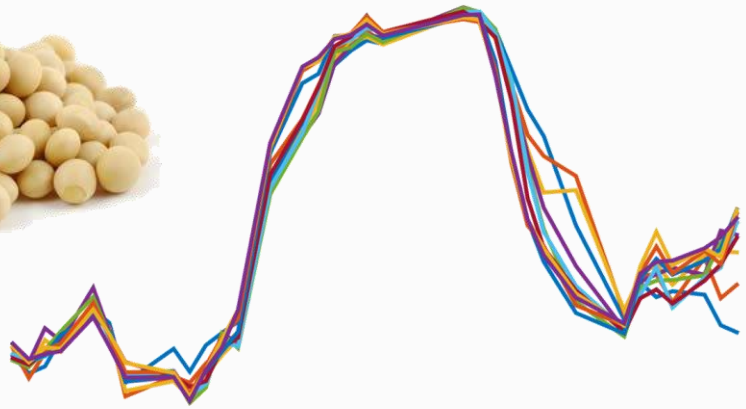


Land use mapping from earth observations

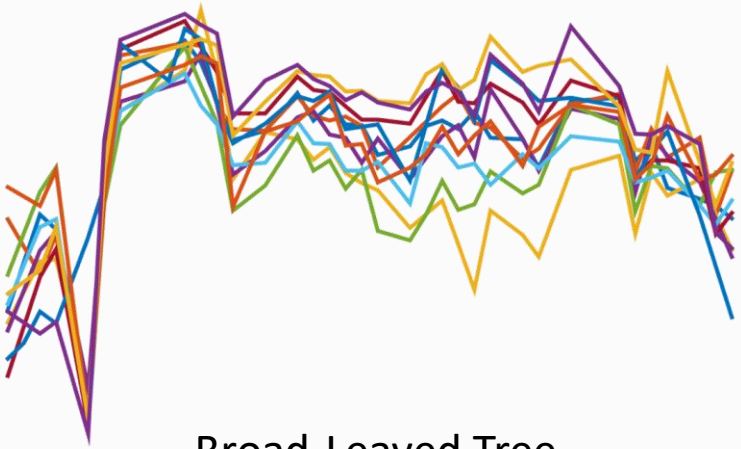


Corn

Wheat



Soybean



Broad-Leaved Tree

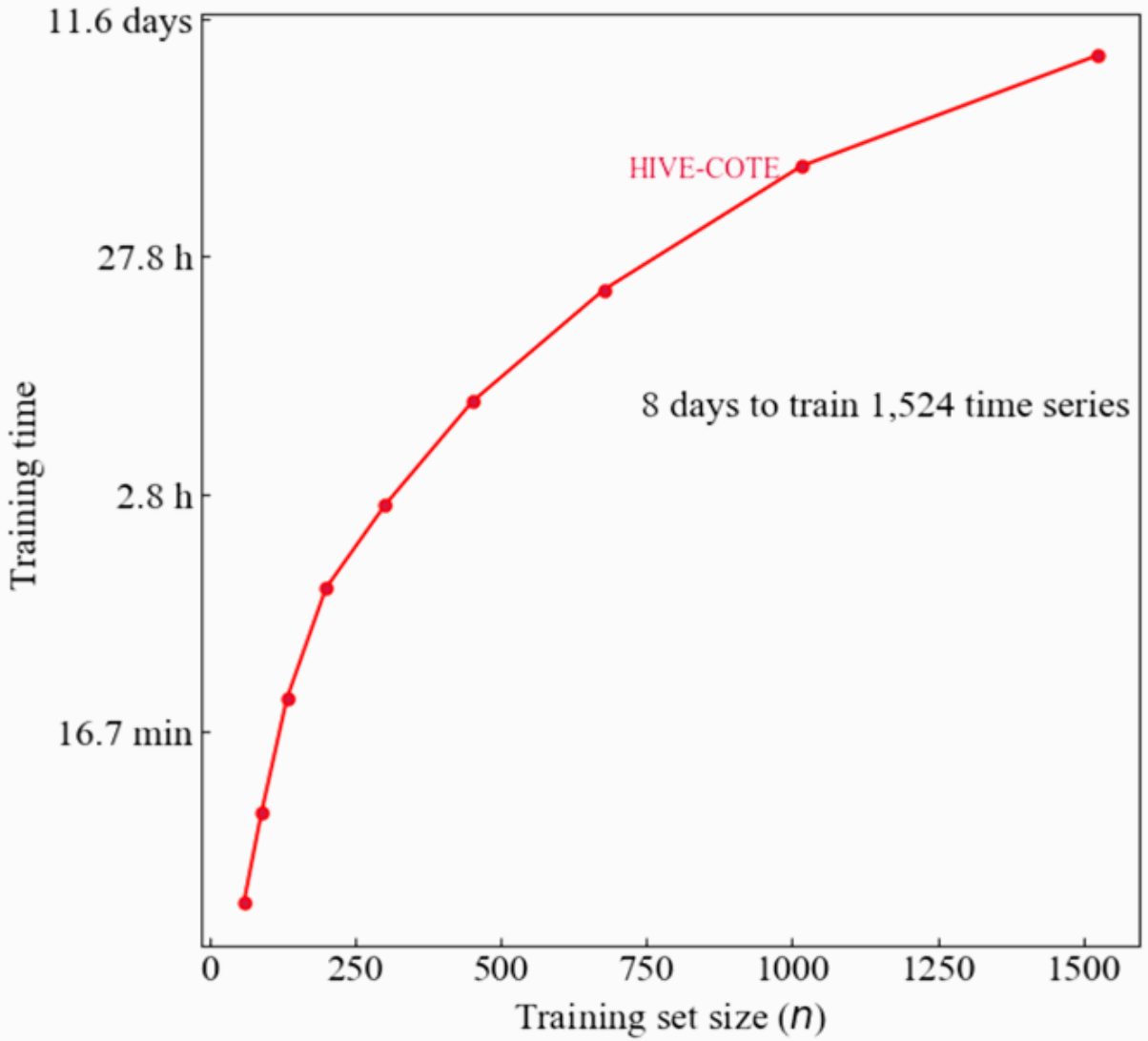
Many aspects of a series may be relevant

- From EXACTLY THE SAME time series:
  - Frequency
    - Racing vs normal pulse
  - Variance in frequency
    - Irregular vs normal heartbeat
  - Amplitude
    - Strong vs weak pulse
  - Local pattern
    - Fault in valve vs normal from shape of peak in ECG
  - Global pattern
    - Declining or improving heartbeat

Many specialized techniques

- Nearest neighbor with specialized similarity measures
- Shapelets
- Dictionary
- Interval statistics
- Deep learning
- Ensembles

However, circa 2019 the most accurate classifiers did not scale

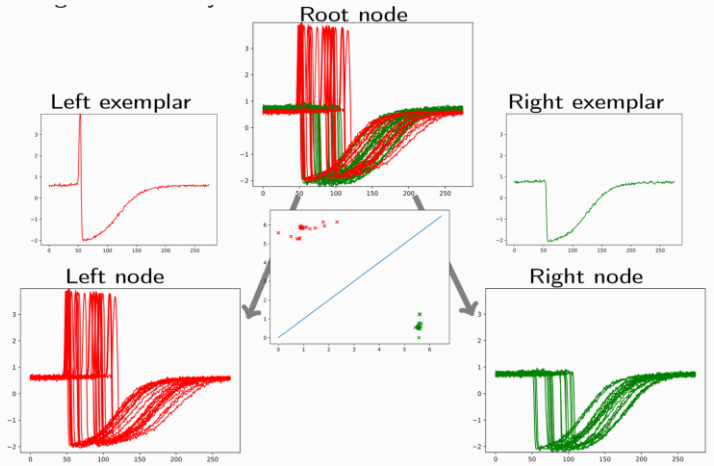


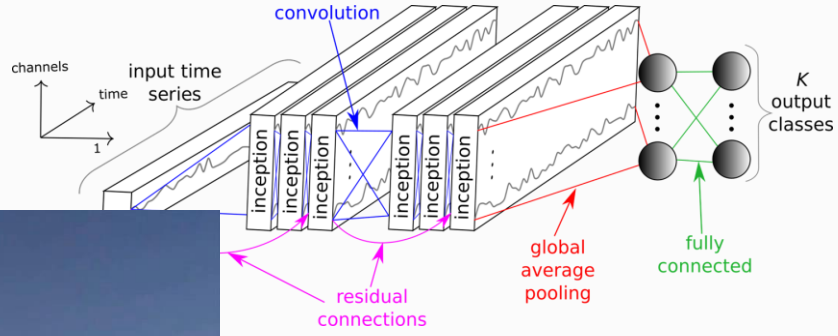
200+ years to learn from 1 million series!

Our accurate and scalable TSCs

- Tree-based: *Proximity Forest* and *TS-CHIEF*
- Deep Learning: *InceptionTime*
- A revolution:

ROCKET





ROCKET

2020 – Data Mining and Knowledge Discovery

Data Mining and Knowledge Discovery
https://doi.org/10.1007/s10618-020-00701-z

ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels

Angus Dempster[1] · François Petitjean[1] · Geoffrey I. Webb[1]

Received: 28 October 2019 / Accepted: 18 June 2020
© The Author(s), under exclusive licence to Springer Science+Business Media LLC, part of Springer Nature 2020

Abstract
Most methods for time series classification that attain state-of-the-art accuracy have high computational complexity, requiring significant training time even for smaller datasets, and are intractable for larger datasets. Additionally, many existing methods focus on a single type of feature such as shape or frequency. Building on the recent success of convolutional neural networks for time series classification, we show that simple linear classifiers using random convolutional kernels achieve state-of-the-art accuracy with a fraction of the computational expense of existing methods. Using this method, it is possible to train and test a classifier on all 85 'bake off' datasets in the UCR archive in < 2 h, and it is possible to train a classifier on a large dataset of more than one million time series in approximately 1 h.

Keywords Scalable · Time series classification · Random · Convolution

1 Introduction

Most methods for time series classification that attain state-of-the-art accuracy have high computational complexity, requiring significant training time even for smaller datasets, and simply do not scale to large datasets. This has motivated the development of more scalable methods such as Proximity Forest (Lucas et al. 2019), TS-CHIEF

Responsible editor: Aristides Gionis, Carlotta Domeniconi, Eyke Hüllermeier, Ira Assent.

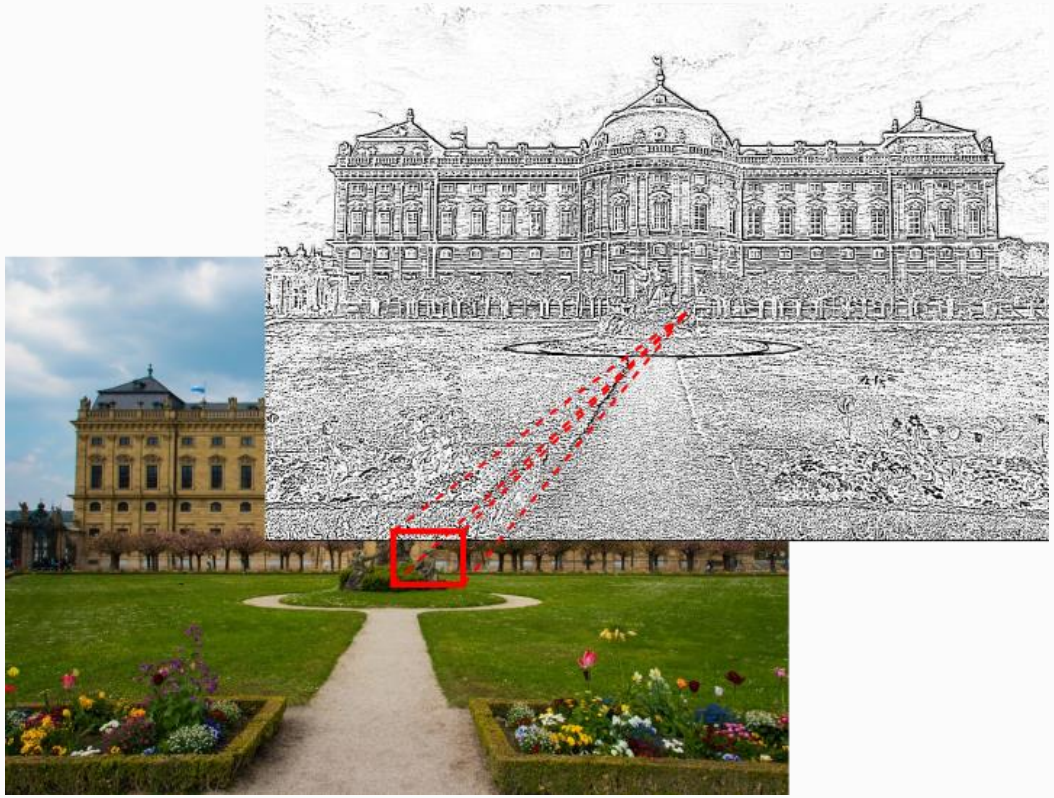
Angus Dempster
angus.dempster1@monash.edu
François Petitjean
francois.petitjean@monash.edu
Geoffrey I. Webb
geoff.webb@monash.edu

[1] Faculty of Information Technology, Monash University, Melbourne, Australia

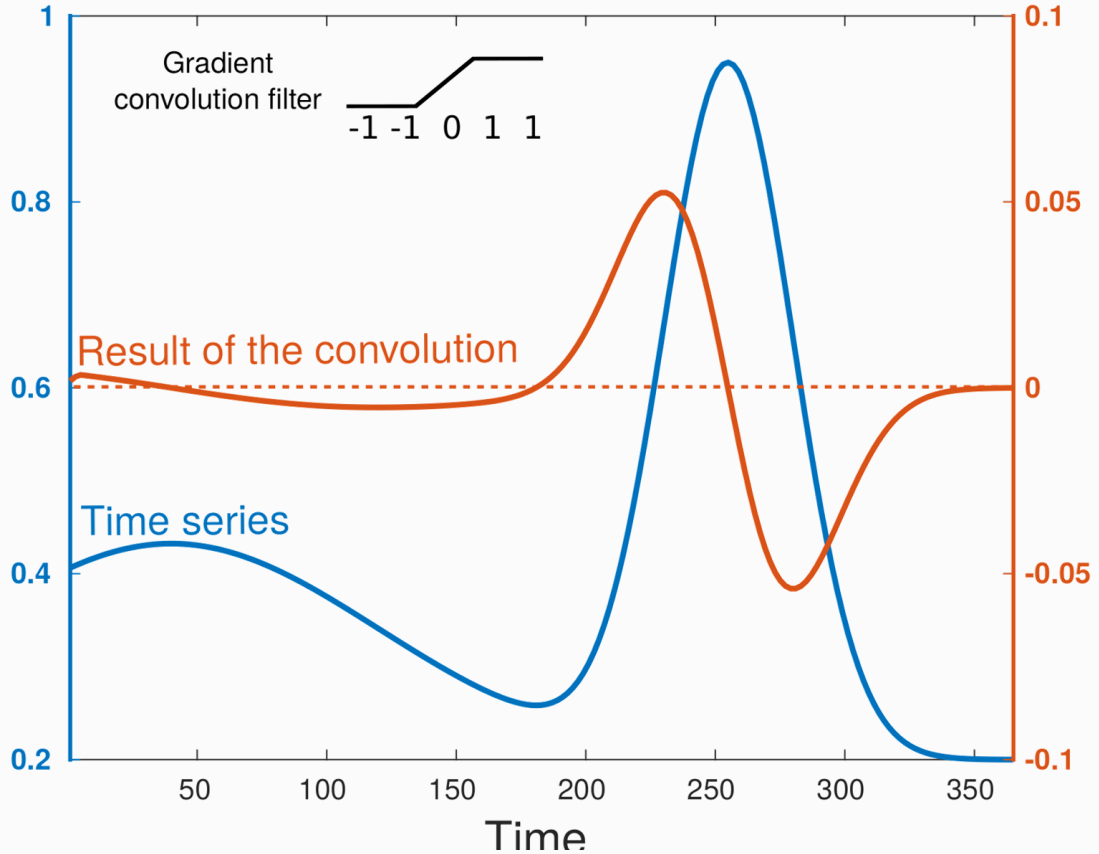
Published online: 13 July 2020 Springer

A. Dempster, F. Petitjean, G. Webb (2020). ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*. **34**(5): 1454 - 1495.
A. Dempster, D. Schmidt, G.Webb (2021) MINIROCKET: A Very Fast (Almost) Deterministic Transform for Time Series Classification. *Proceedings of the 27th SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 248-257, 2021.

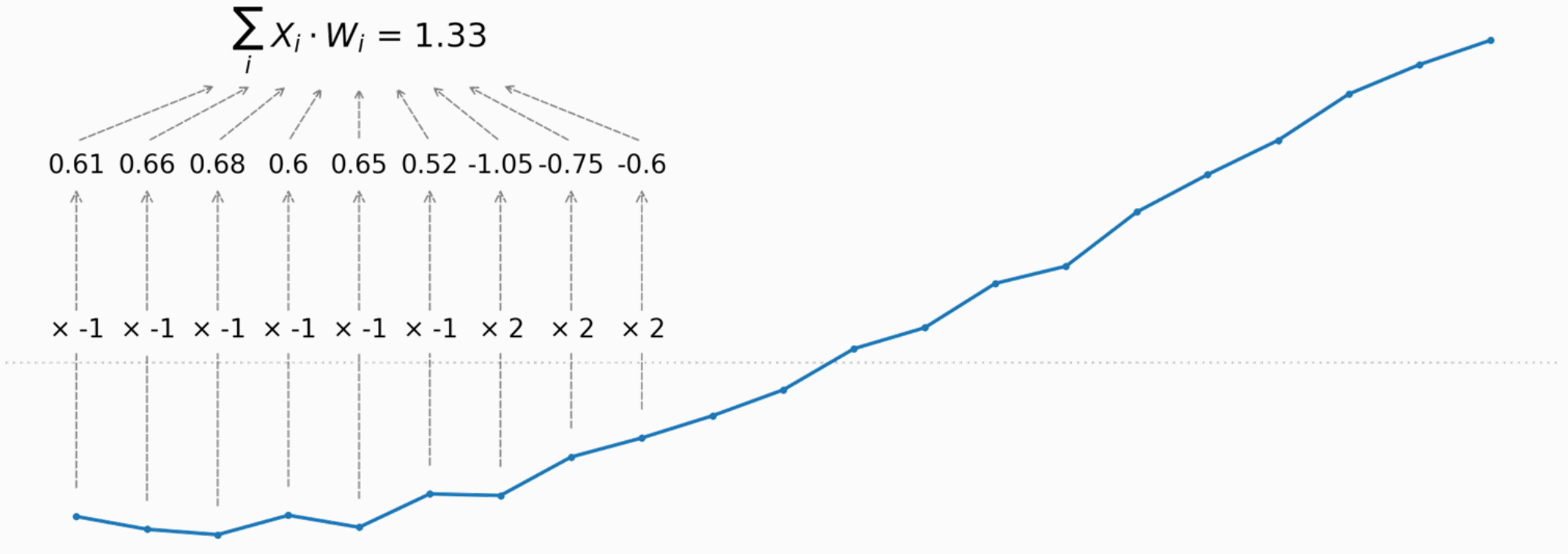
Convolution on images vs. time series



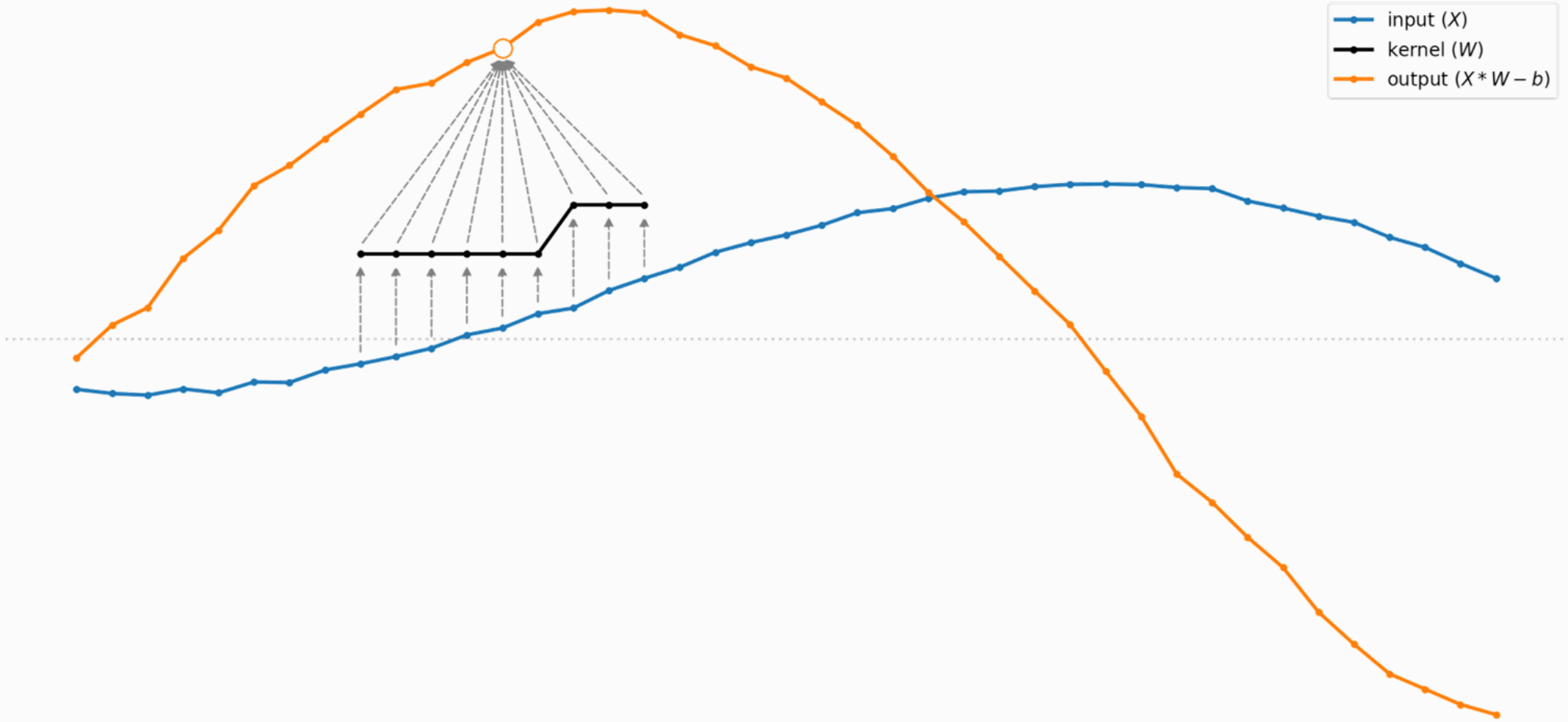
The result of applying an edge detection convolution on an image



Time Series Convolutions

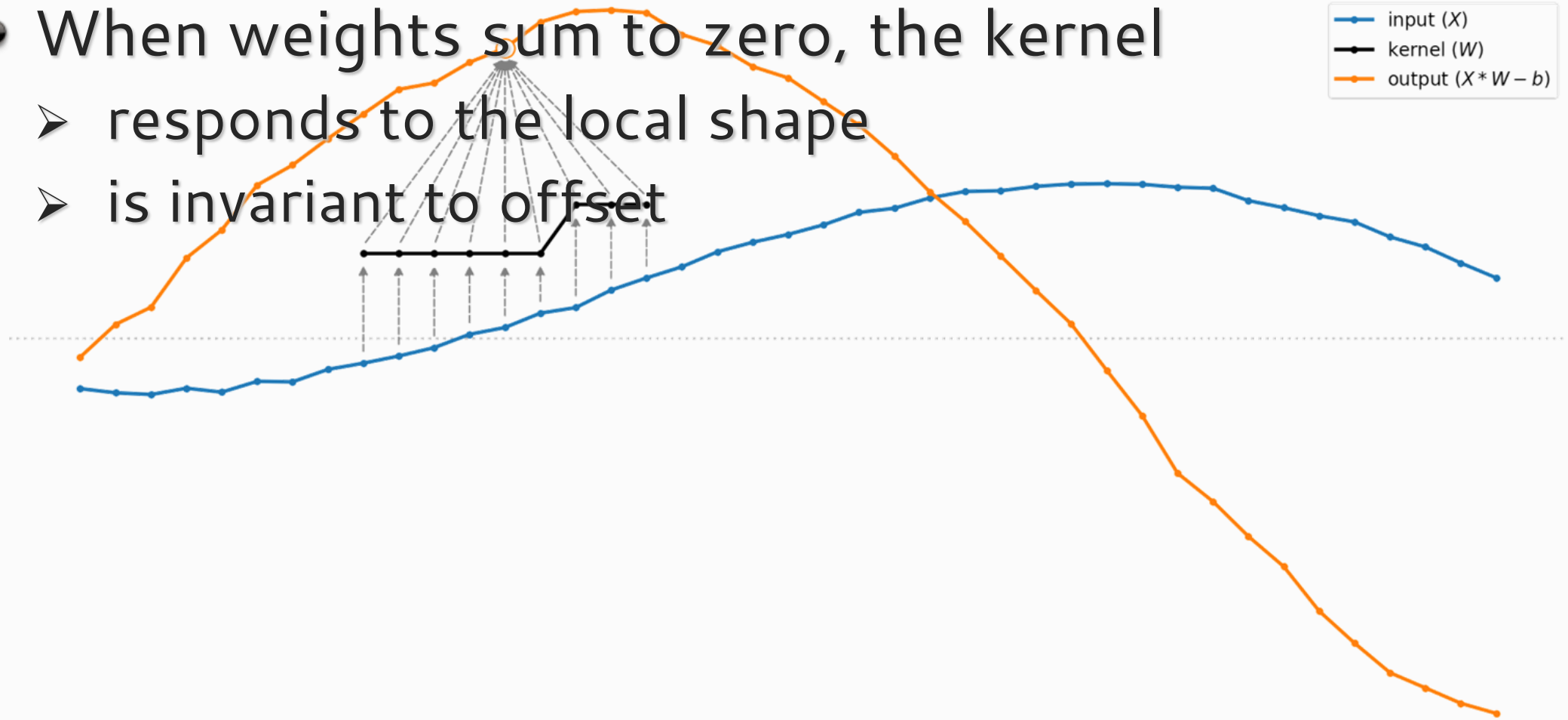


Time Series Convolutions

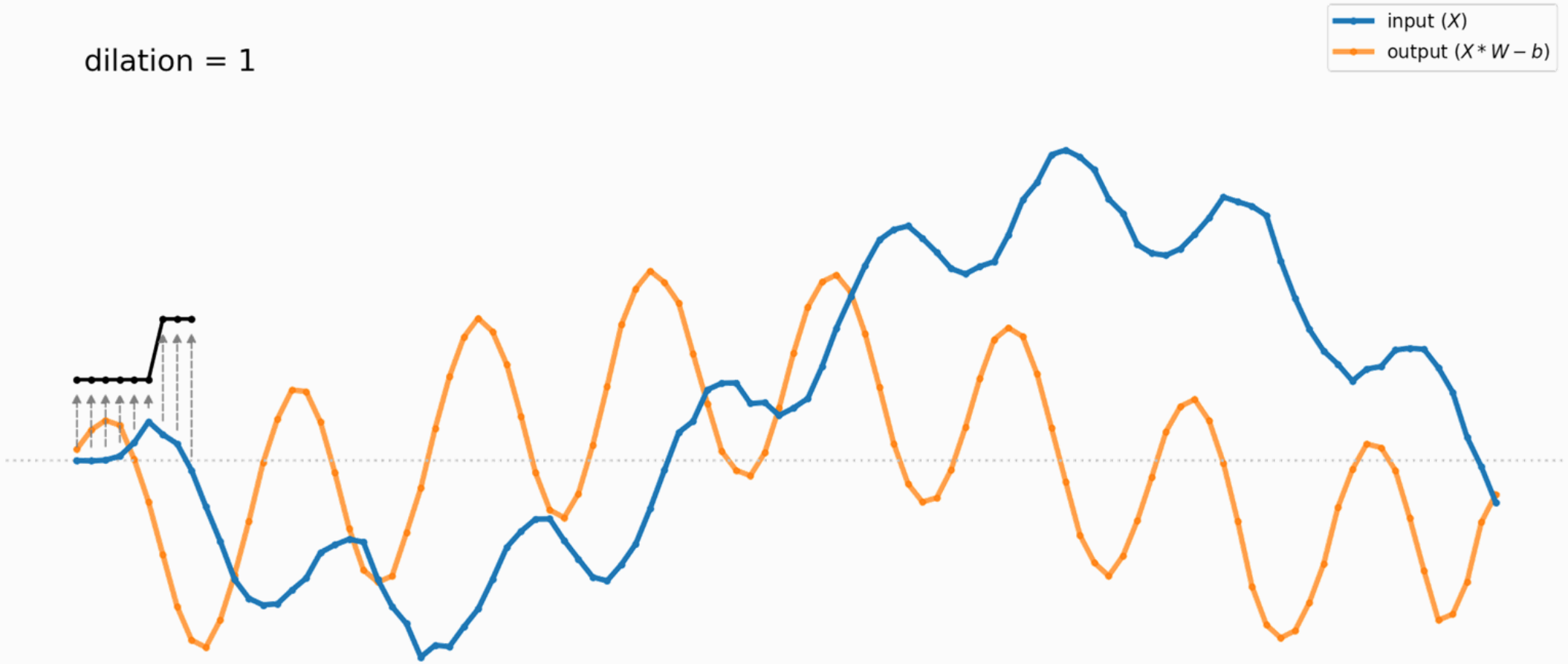


Sum to zero

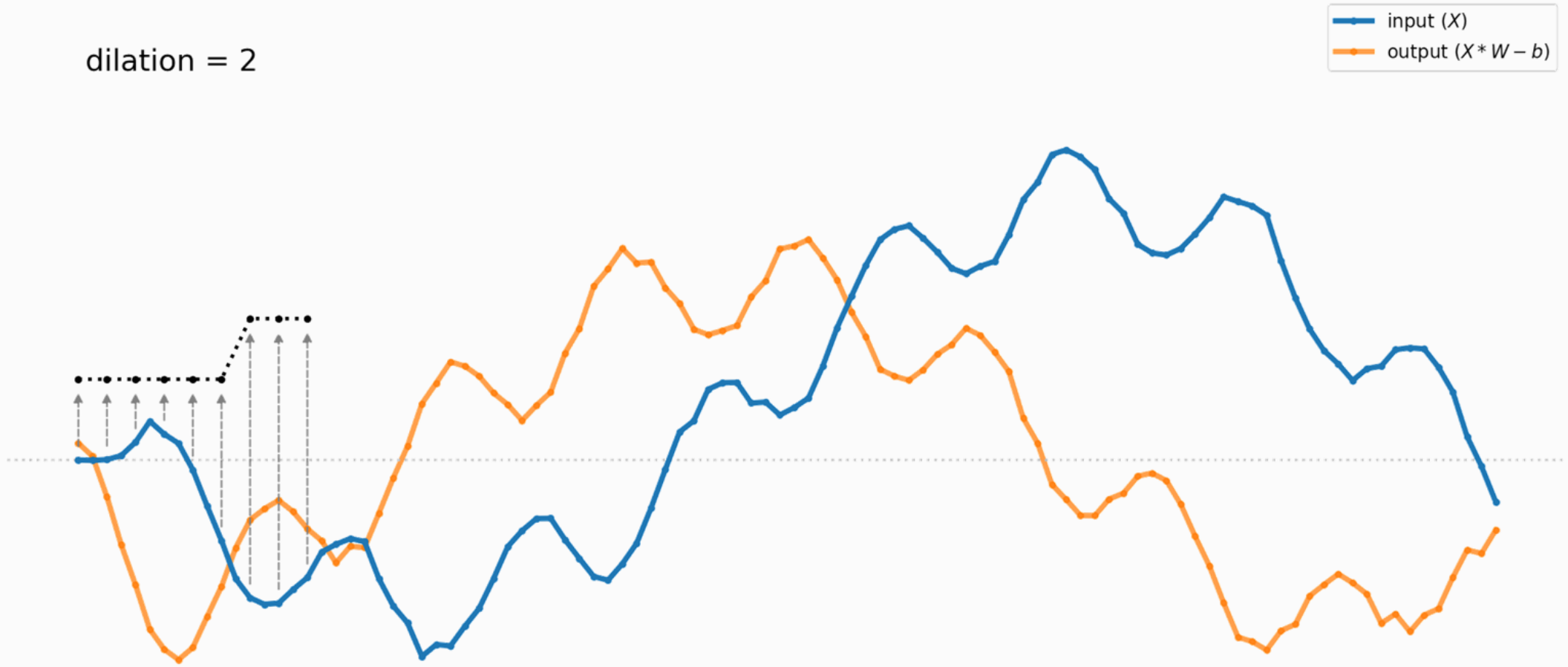
- When weights sum to zero, the kernel
  - responds to the local shape
  - is invariant to offset



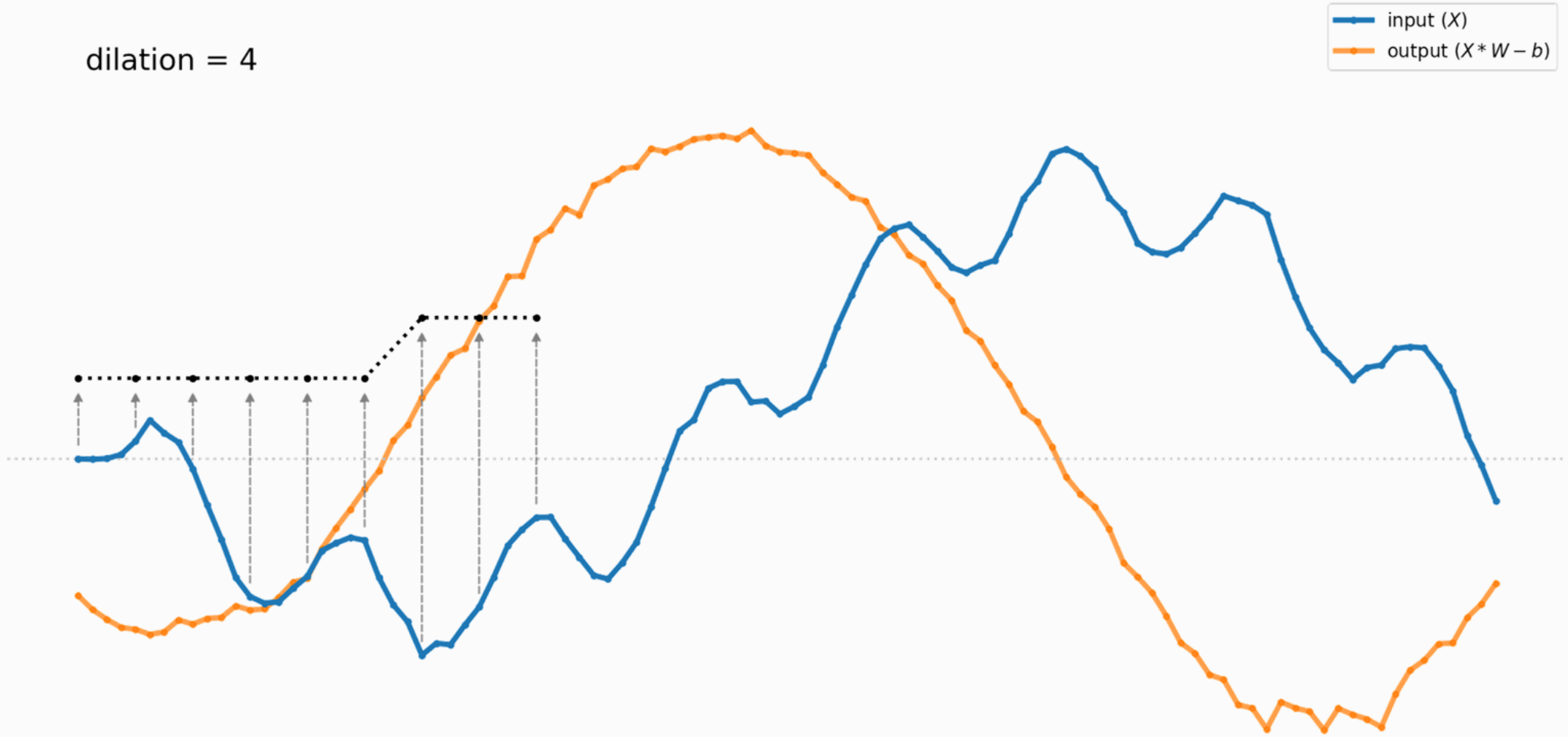
Dilation



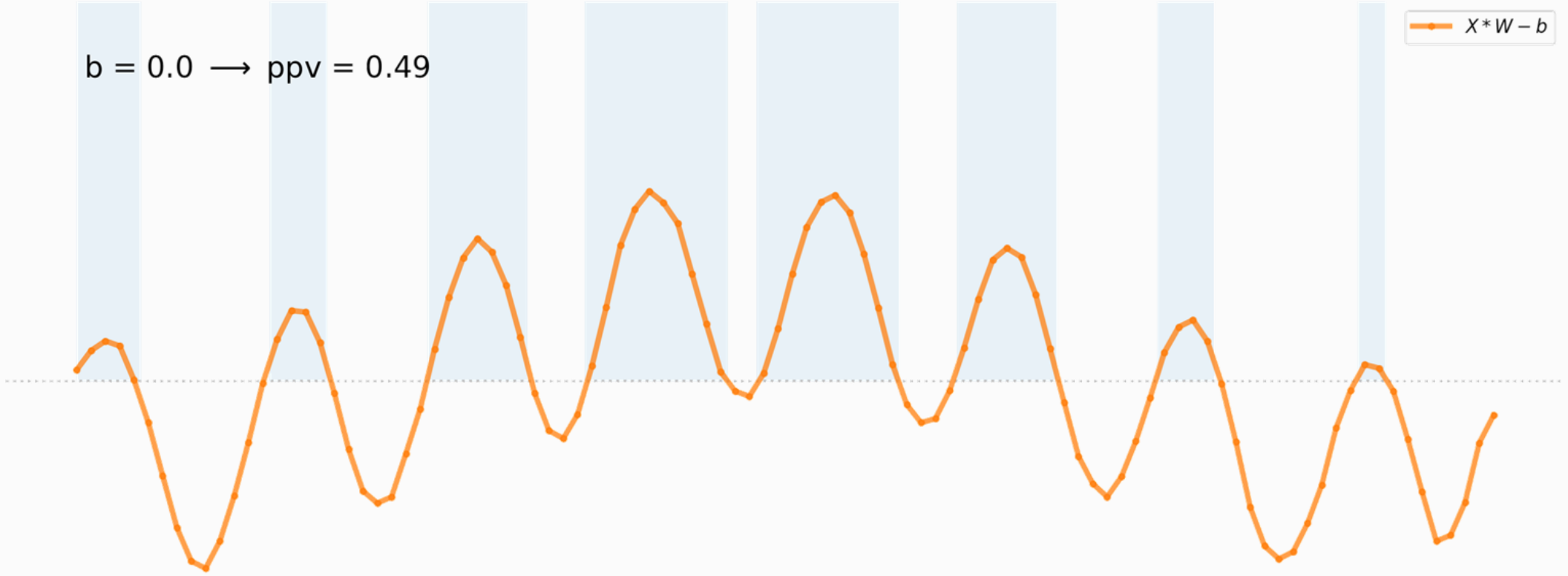
Dilation



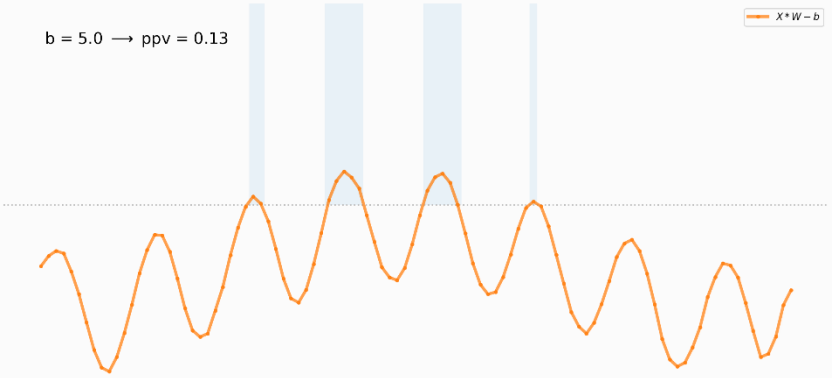
Dilation

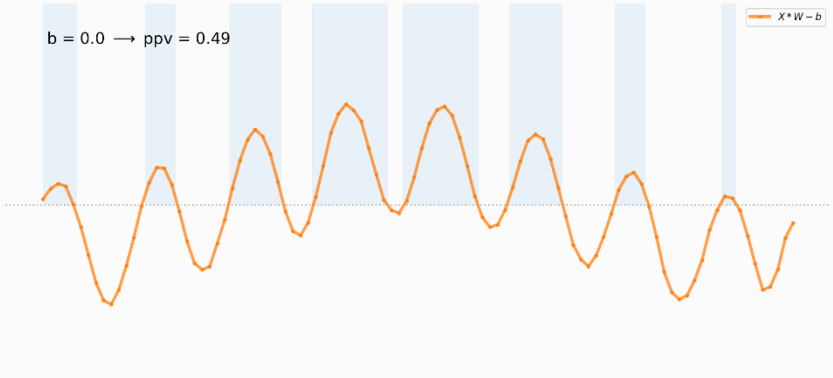


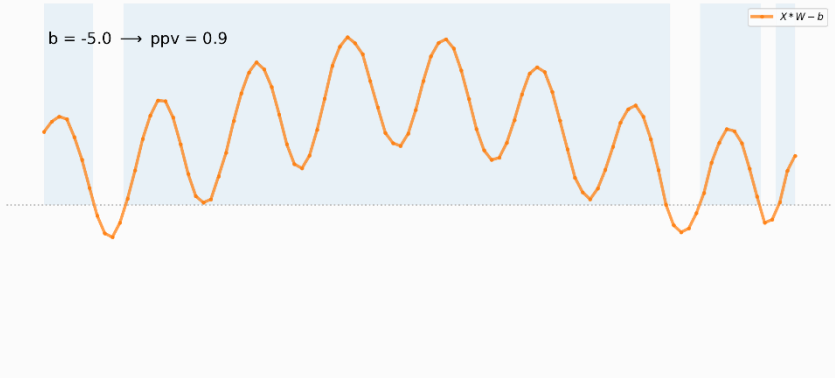
Proportion of Positive Values (PPV)



PPV and Bias







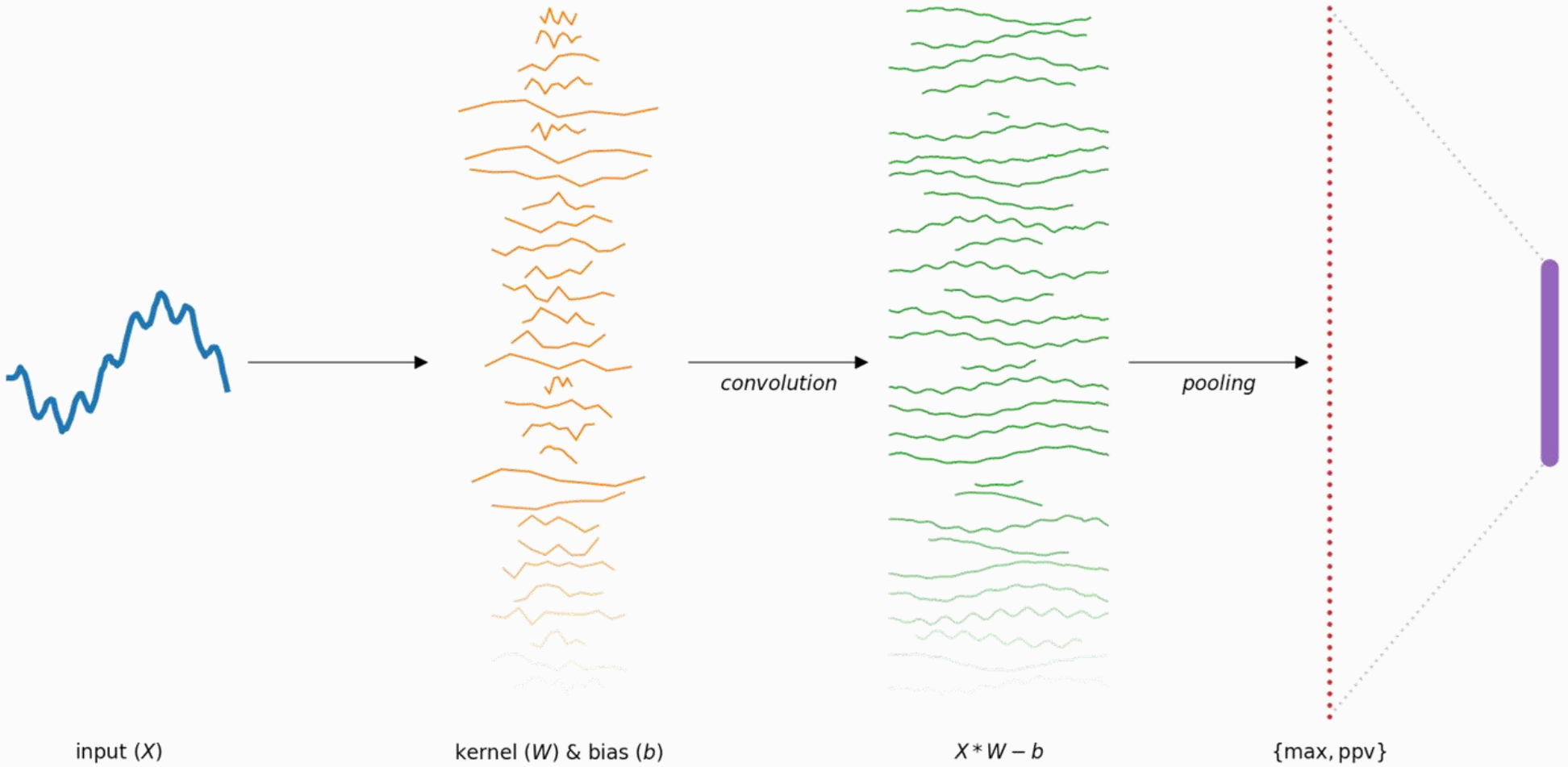
Exploit power of convolutional filters

- Convolutional filters can capture many different types of feature of a time series
  - Frequency, Amplitude, Local pattern, Global pattern
- Generate a large number – 10,000
- Learn simple linear classifier
  - Logistic regression when number of examples is large
  - Ridge regression when number of examples is small
    - Stronger regularisation
    - Faster for small sample size

ROCKET: Random choices per kernel

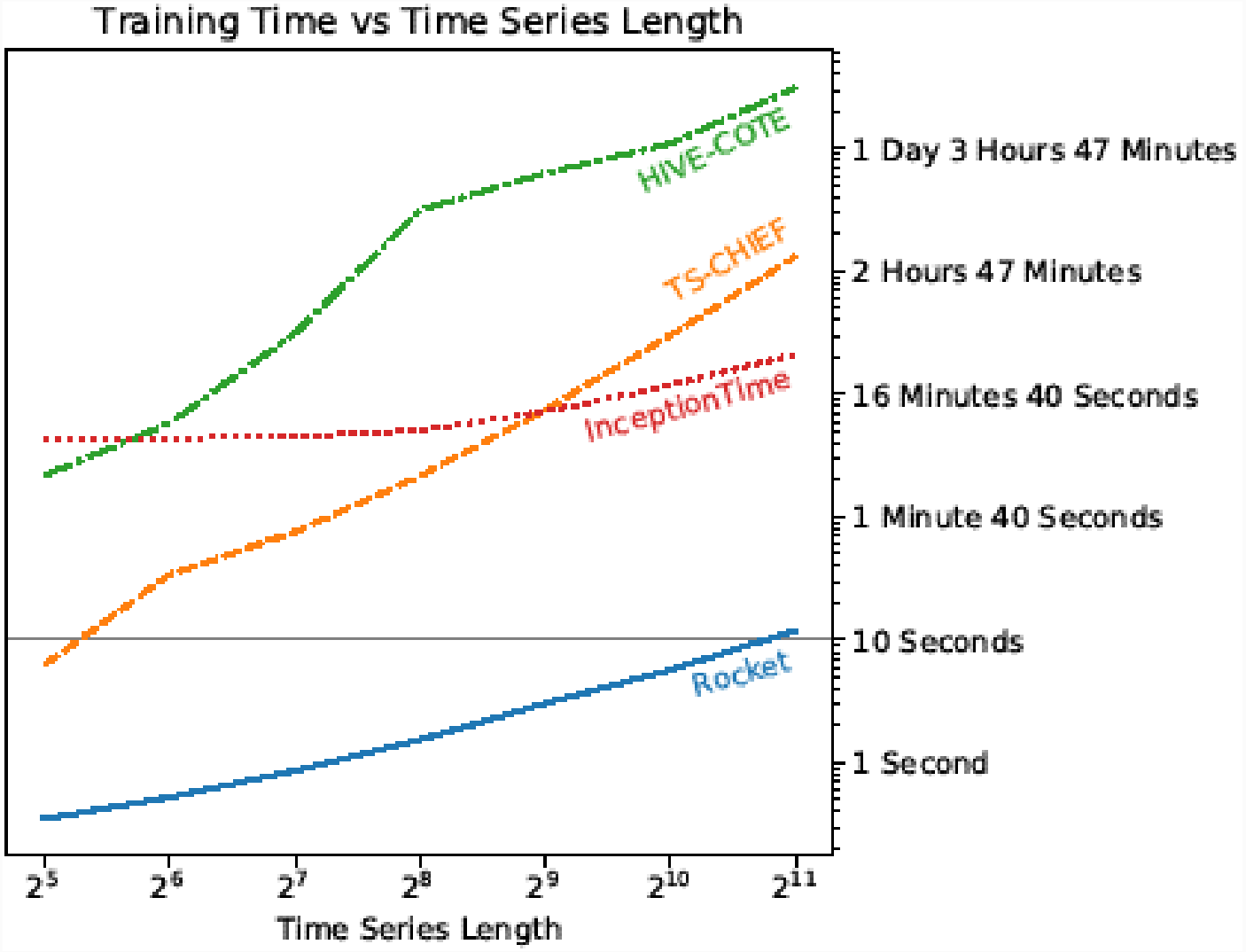
- Length: {7, 9, 11}
- Weights: $N(0, 1)$, then normalized to sum to zero
- Bias: $U(-1, 1)$
- Dilation: $2^{U(0, A)}$, where A limits kernel to series length
- Padding: {T, F}
- Pooling operators: PPV + Max

ROCKET



Time relative to number of training examples (Satellite dataset)

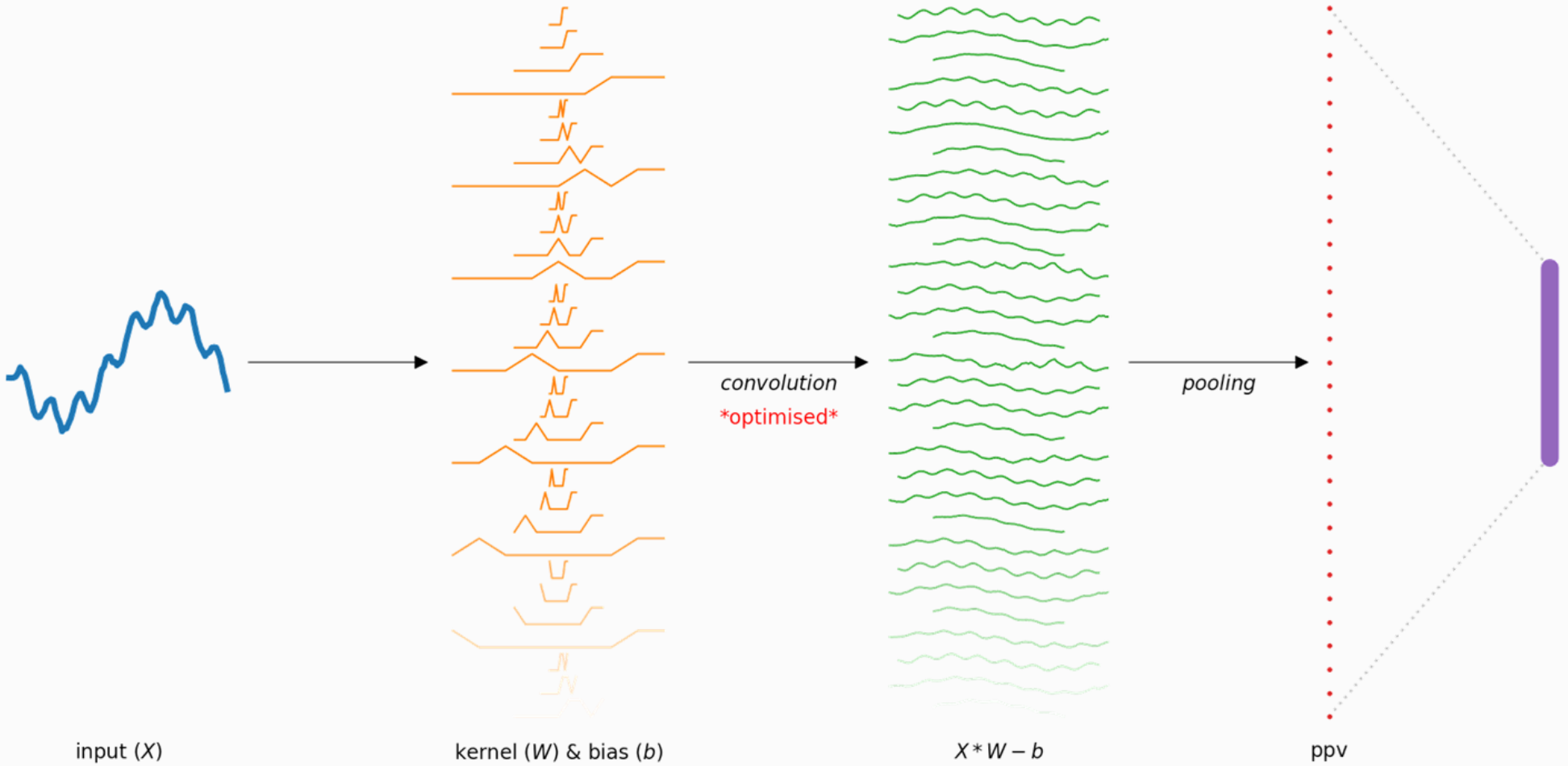
Time relative to series length (Inlineskate dataset)



MiniRocket: Fixed choices per kernel

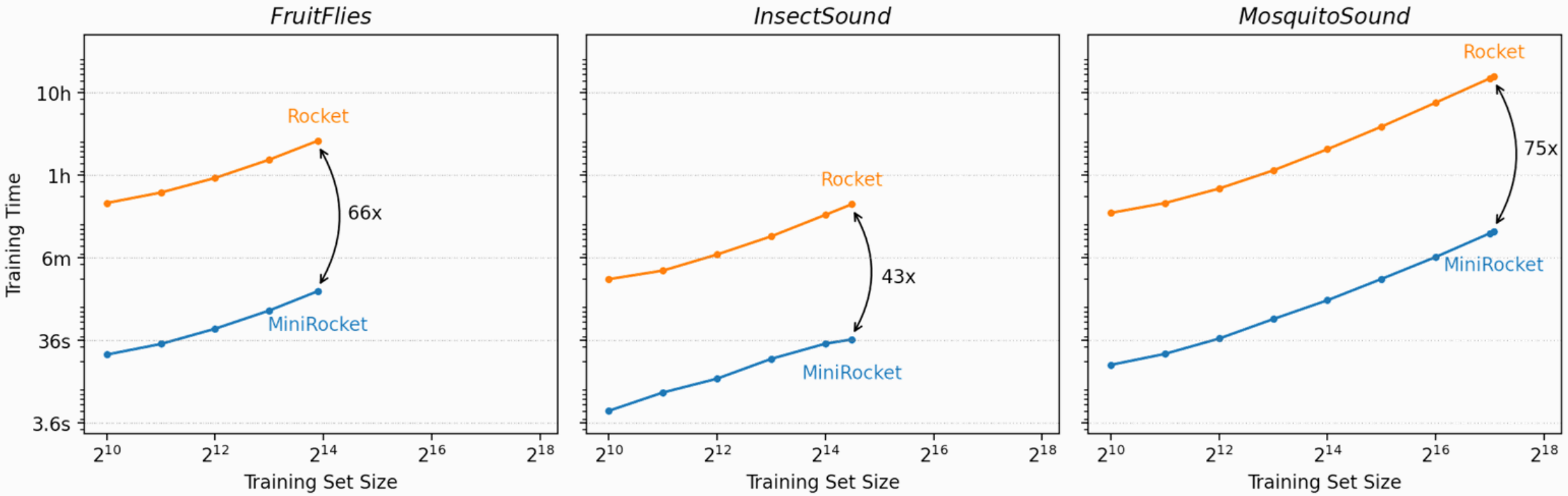
- Length: $\{7, 9, 11\}$
- Weights: $N(0, 1)$ $\{-1, 2\}$, such that sum to zero
- Bias: $U(-1, 1)$ from convolution output
- Dilation: $2^{U(0, A)}$ fixed (relative to input length)
- Padding: {T, F}
- Pooling: PPV + Max

MiniRocket


convolution
optimised
pooling
input (X)
kernel (W) & bias (b)
X * W − b
ppv

Scalability: MiniRocket vs Rocket

restricted to 1 CPU core



17,259 x 5,000

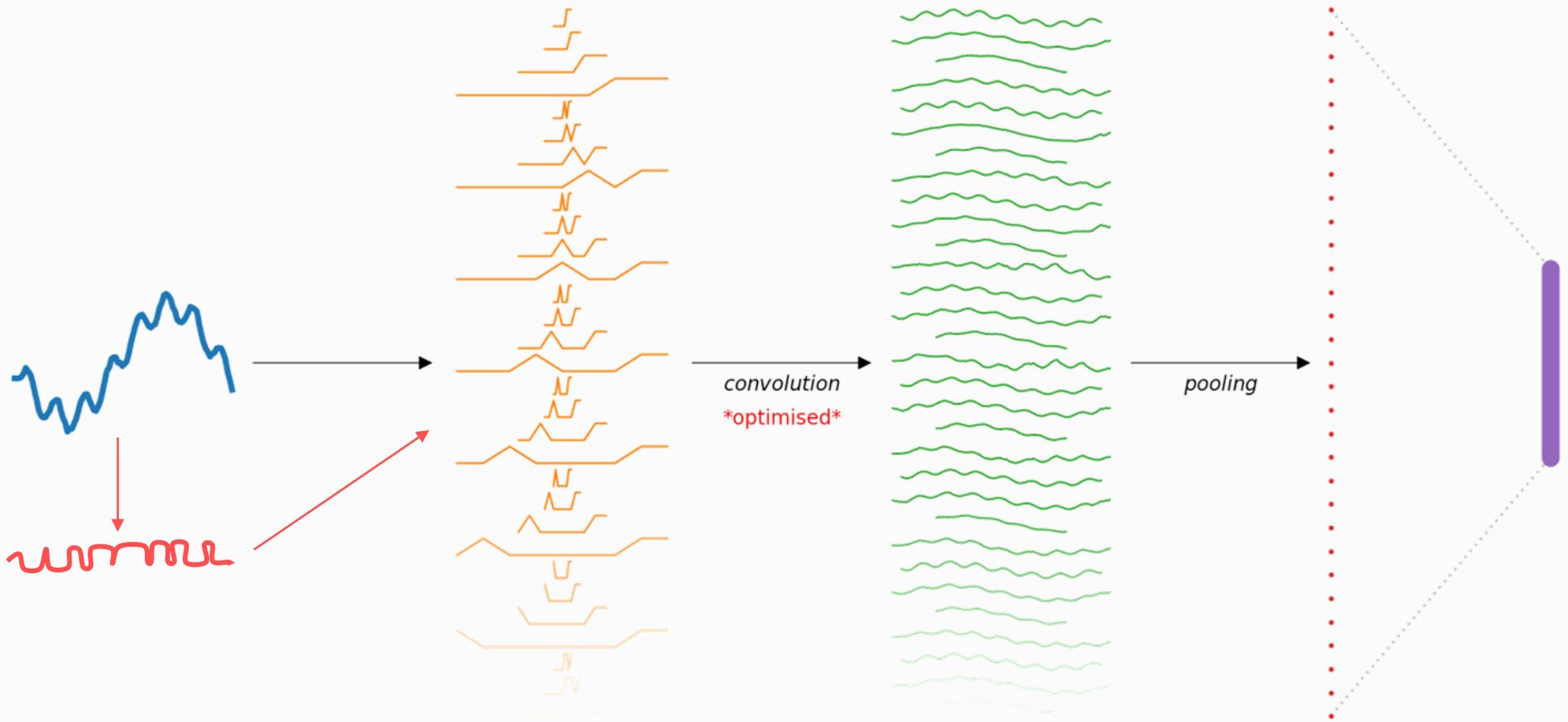
25,000 x 600

139,780 x 3,750

num. training examples

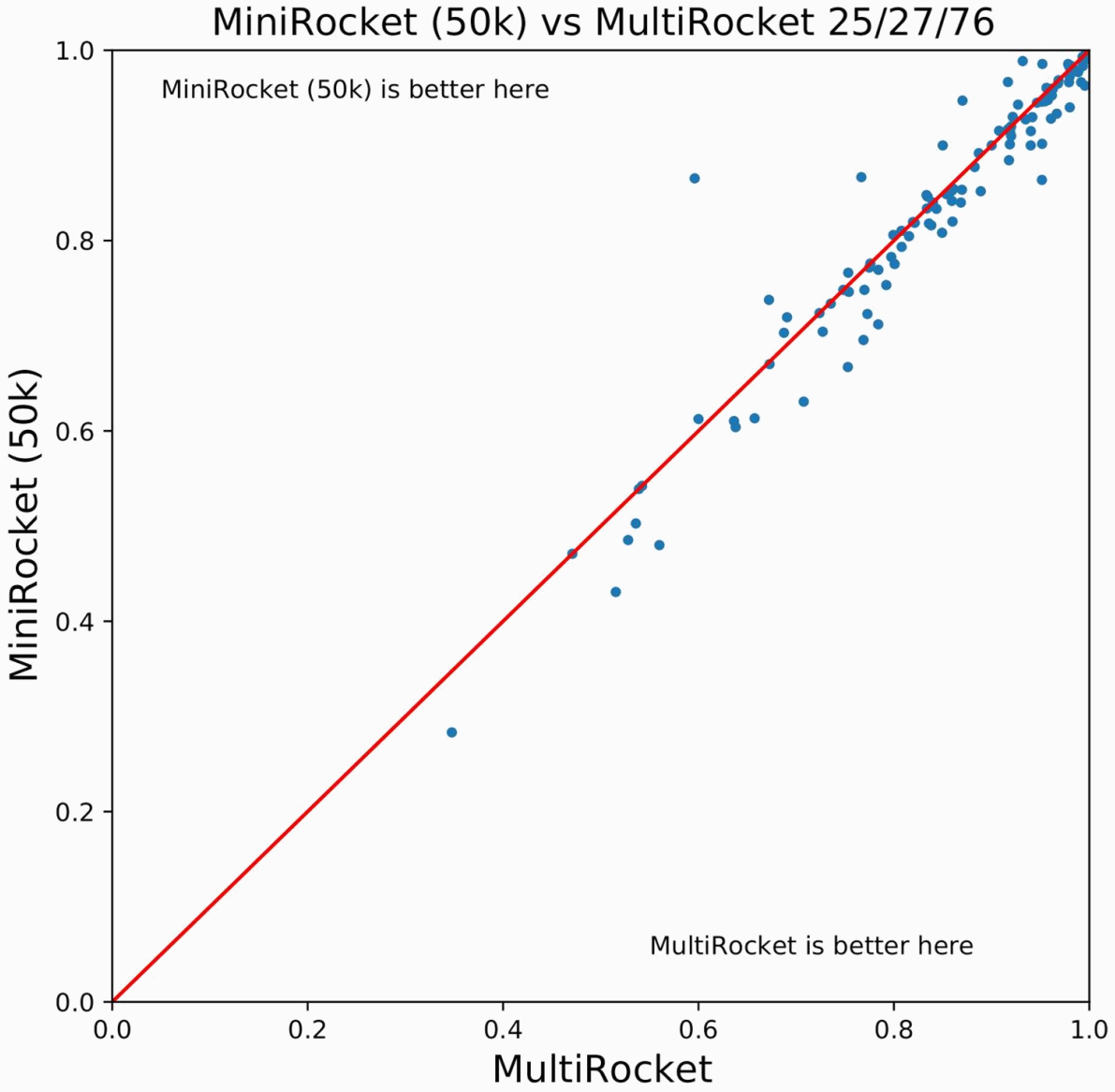
time series length

MultiRocket

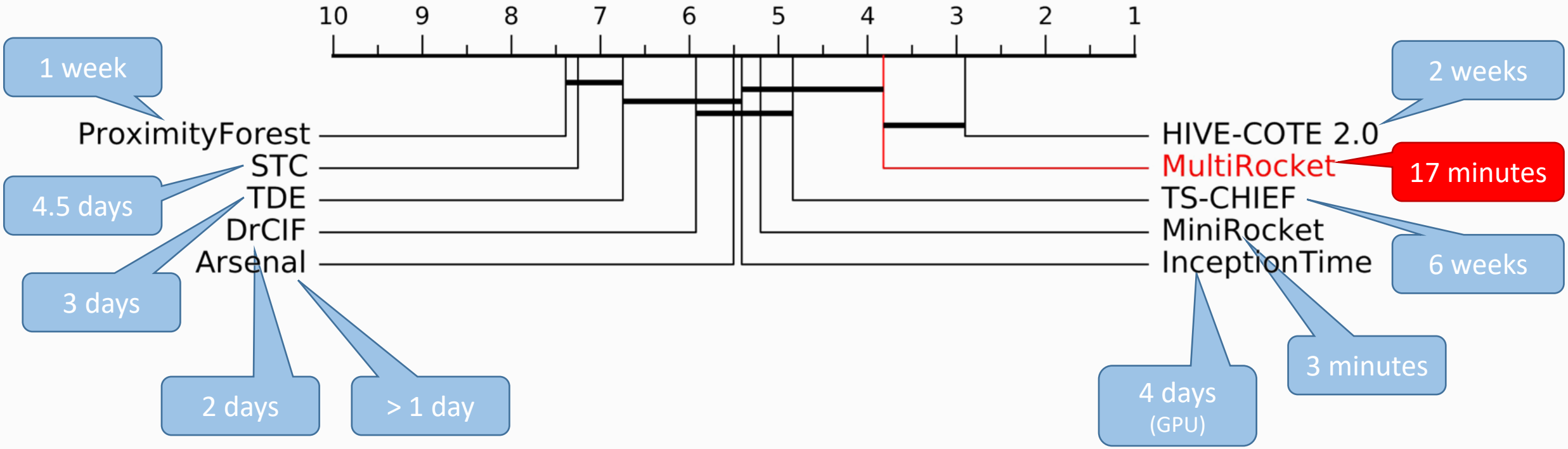


ppv (how many)
Mean Positive Value (how strong)
Mean Positive Value Index (where)
Longest Stretch of Positive Values (how dispersed)

MultiRocket vs MiniRocket (accuracy)



Benchmark accuracy (UCR repository)



An independent assessment

'ROCKET is the best ranked and by far the fastest classifier and would be our recommendation as the default choice for Multivariate Time Series Classification problems.'

Ruiz, A.P., Flynn, M., Large, J. *et al.* The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery* (2020).

An independent assessment

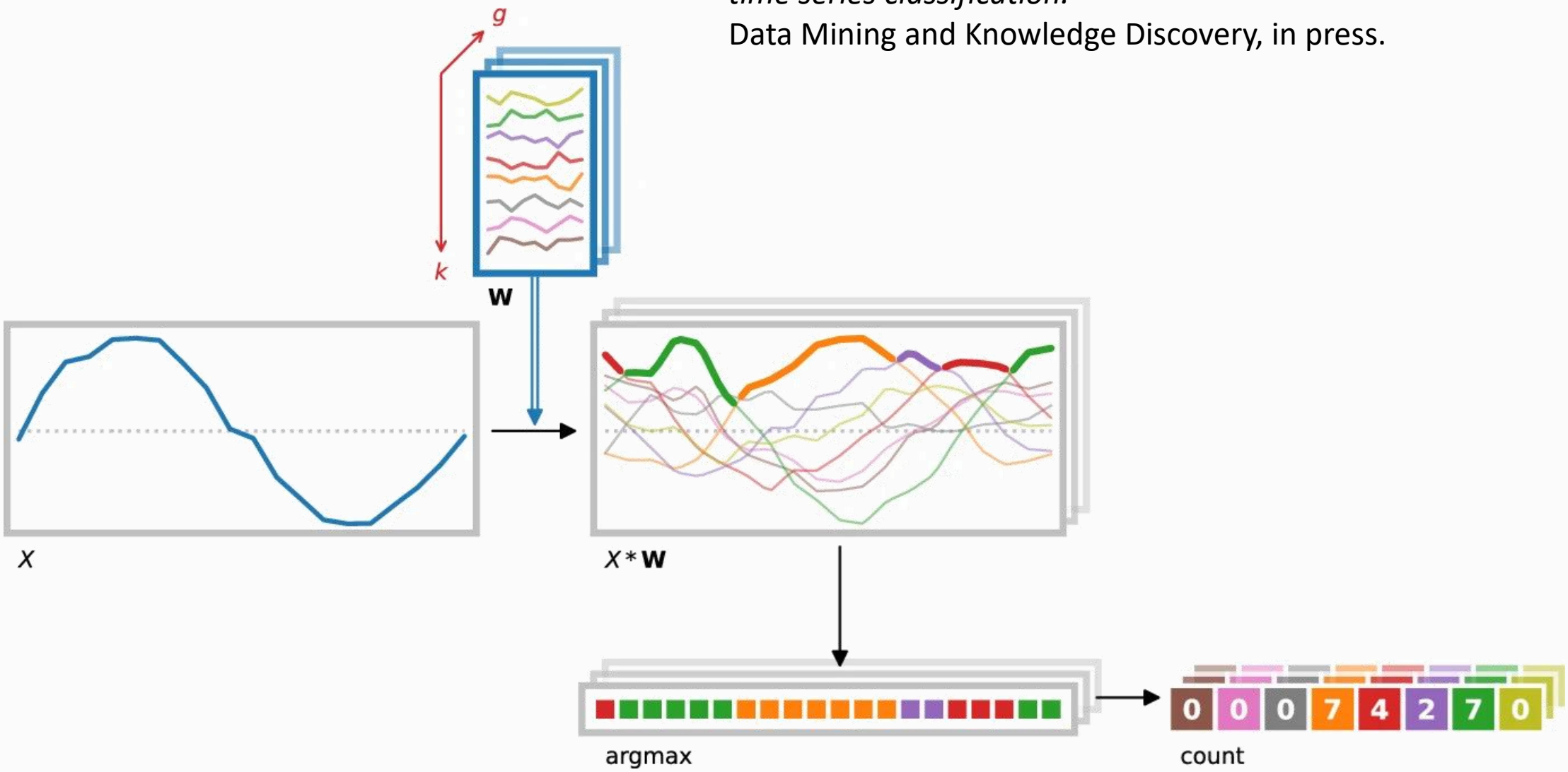
'***There are currently four algorithms with reasonable claim to being state of the art for TSC*** *based on experimentation on the recently expanded UCR archive (Dau et al. 2019). These are: the deep learning approach called* ***InceptionTime*** *(Fawaz et al. 2020); the tree based Time Series Combination of Heterogeneous and Integrated Embedding Forest (****TS-CHIEF****) (Shifaz et al. 2020); the Random Convolutional Kernel Transform (****ROCKET****) (Dempster et al. 2020); and the heterogeneous meta-ensemble Hierarchical Vote Collective of Transformation-based Ensembles (****HIVE-COTE****) (Lines et al. 2018), the latest version of which is called HIVE-COTE version 1.0 (HC1) (Bagnall et al. 2020).*'

'**ROCKET is a very fast classifier that has state-of-the-art accuracy, and we believe it is the most important recent development in the field**.'

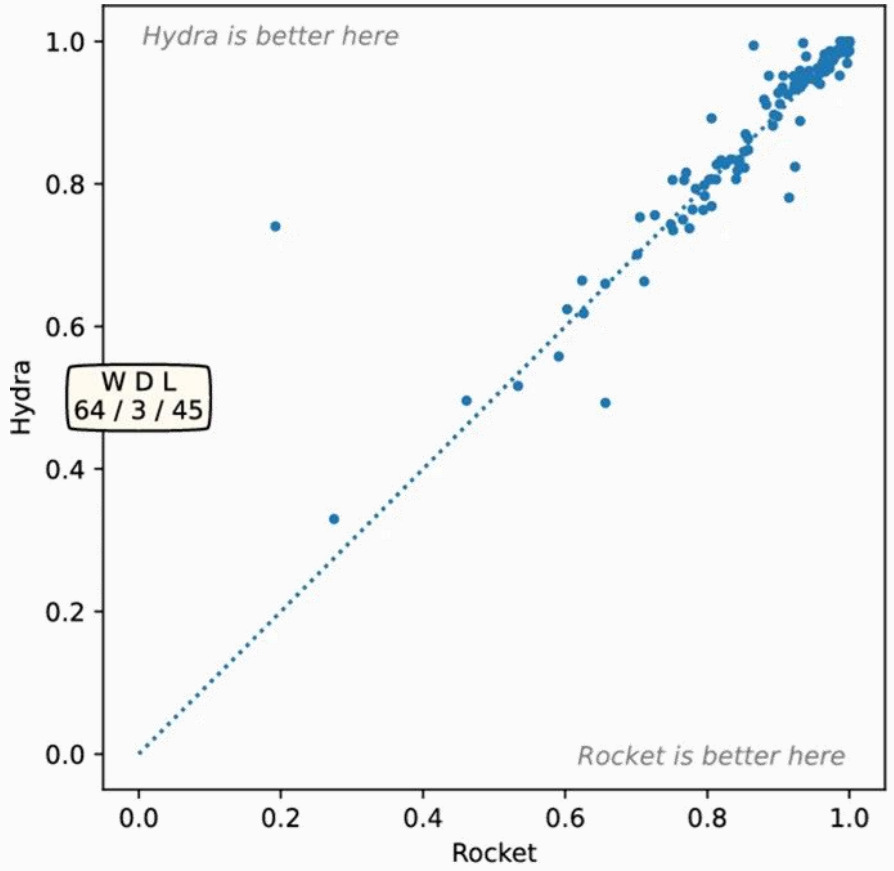
Middlehurst, M., *et al.* HIVE-COTE 2.0: a new meta ensemble for time series classification. *Machine Learning* **110,** 3211–3243 (2021).

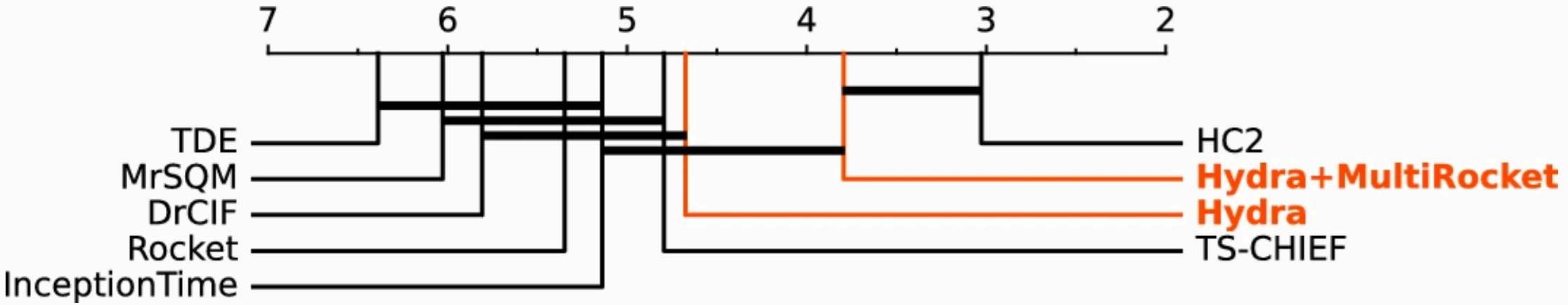
Hydra

Dempster, A., Schmidt, D. F., & Webb, G. I.
HYDRA: competing convolutional kernels for fast and accurate time series classification.
Data Mining and Knowledge Discovery, in press.



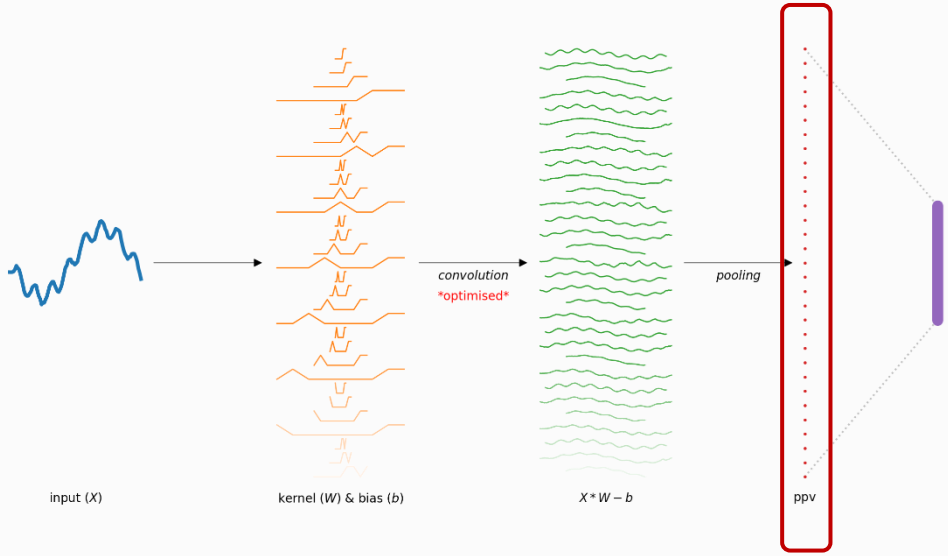
Hydra


Hydra is better here
Rocket is better here
W D L
64 / 3 / 45
Hydra
Rocket
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


7
6
5
4
3
2
TDE
MrSQM
DrCIF
Rocket
InceptionTime
HC2
Hydra+MultiRocket
Hydra
TS-CHIEF

Research opportunities

- Additional pooling operators



- Employ other forms of kernel
- Use ROCKET features in other time series analytics
  - Forecasting, regression, clustering, anomaly detection, …
- Use non-linear classifiers
- Apply to other data types

Conclusions

- **Rocket** revolutionized time series classification
  - state of the art accuracy
  - many orders of magnitude less computation
- **MiniRocket** achieves another order of magnitude speed up
  - shows that stochasticity does not directly contribute to accuracy
- **MultiRocket** provides substantial gain in accuracy for modest computation
- **Hydra** provides a further consistent gain in accuracy for modest computation
- We believe in reproducible research:
  - **Rocket** → https://github.com/angus924/rocket
  - **MiniRocket** → https://github.com/angus924/minirocket
  - **MultiRocket** → https://github.com/ChangWeiTan/MultiRocket
  - **Hydra** → https://github.com/angus924/hydra

Thank you!

Angus Dempster

Geoff Webb

Francois Petitjean

Daniel Schmidt

Chang Wei Tan

http://i.giwebb.com/

Images and video licensed from Shutterstock and Pexel