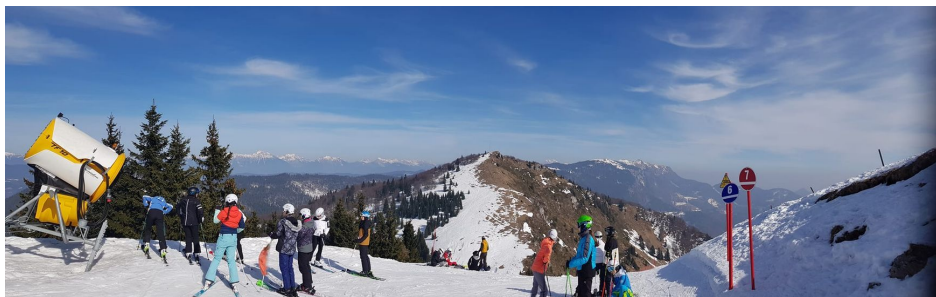# Data Augmentation Does Not Necessarily Beat a Smart Algorithm

Krisztian Buza

Institute Jozef Stefan
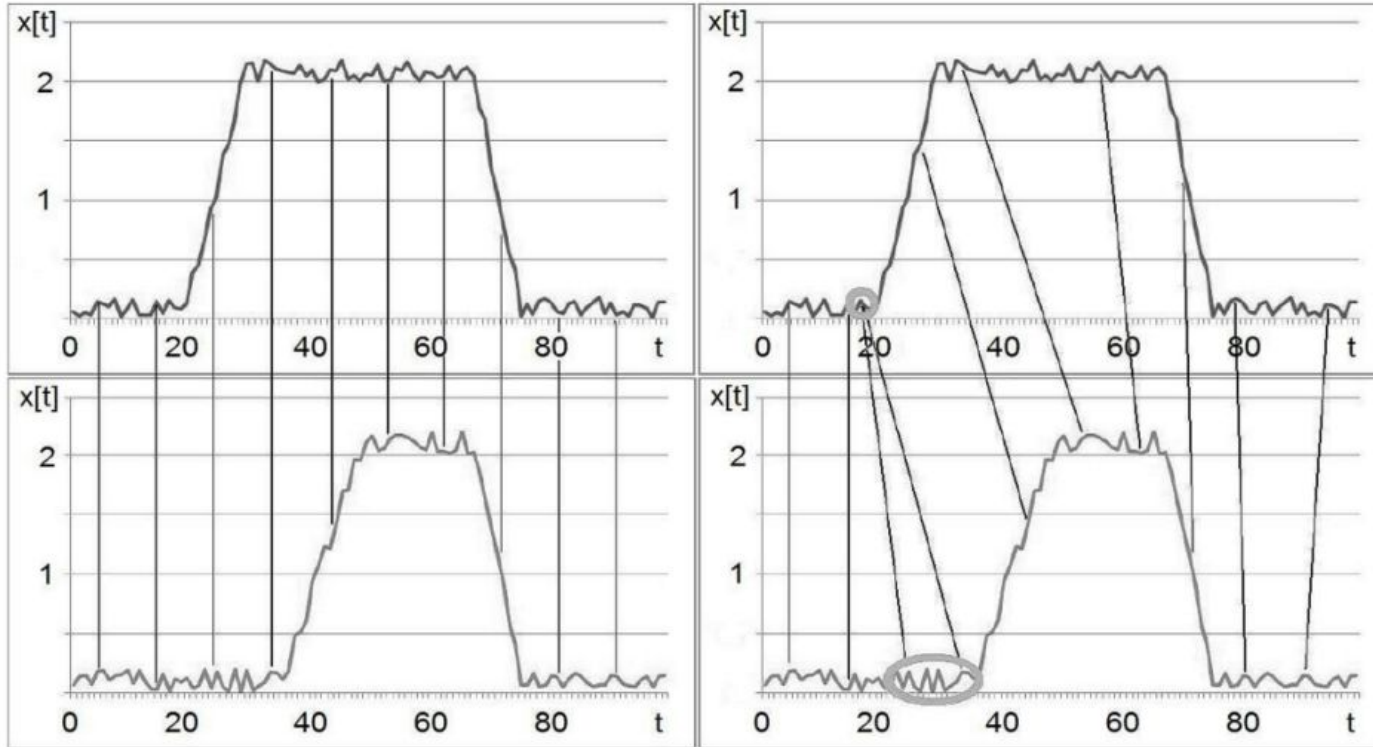Artificial Intelligence Laboratory
Jamova 39, 1000 Ljubljana, Slovenia

BioIntelligence Group
Department of Mathematics-Informatics
Sapientia Hungarian University of Transylvania
Targu Mures, Romania
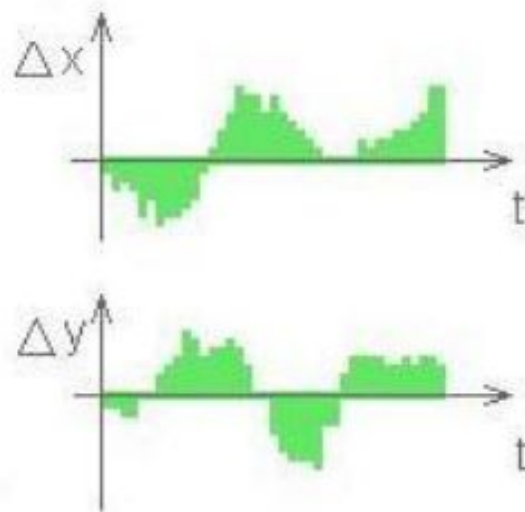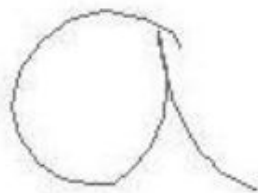
buza@biointelligence.hu
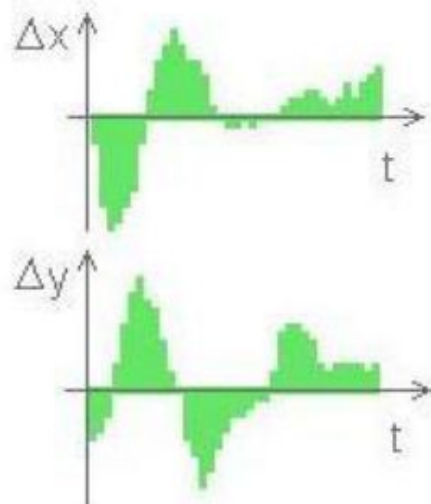
Dynamic Time Warping

Dynamic Time Warping in Time Series Classification

See also:
H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition
IEEE transactions on acoustics, speech, and signal processing 26.1 (1978)

Handwriting Recognition

Dynamic Image Warping
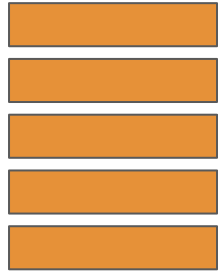
Dynamic Image Warping

intensity values

250	249	251	127	…
251	248	252	126	…
253	247	136	103	…
251	249	127	98	…
…	…	…	…	…

Dynamic Image Warping

intensity values

250	249	251	127	…
251	248	252	126	…
253	247	136	103	…
251	249	127	98	…
…	…	…	…	…

Dynamic Image Warping

intensity values

250	249	251	127	...
251	248	252	126	...
253	247	136	103	...
251	249	127	98	...
...	...	...	...	...

Dynamic Image Warping

intensity values

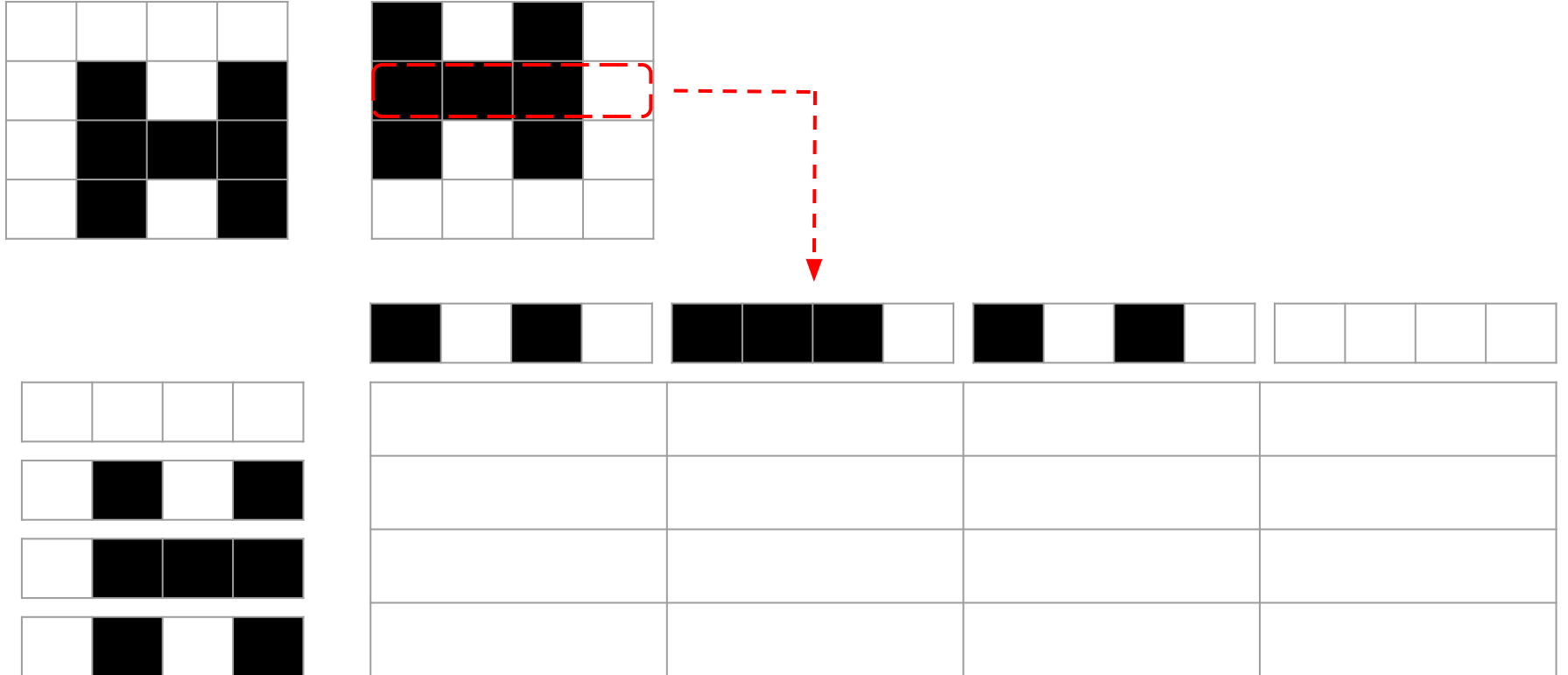| 250 | 249 | 251 | 127 | ... |
251	248	252	126	...
253	247	136	103	...
251	249	127	98	...
...	...	...	...	...

DTW = Dynamic Time Warping

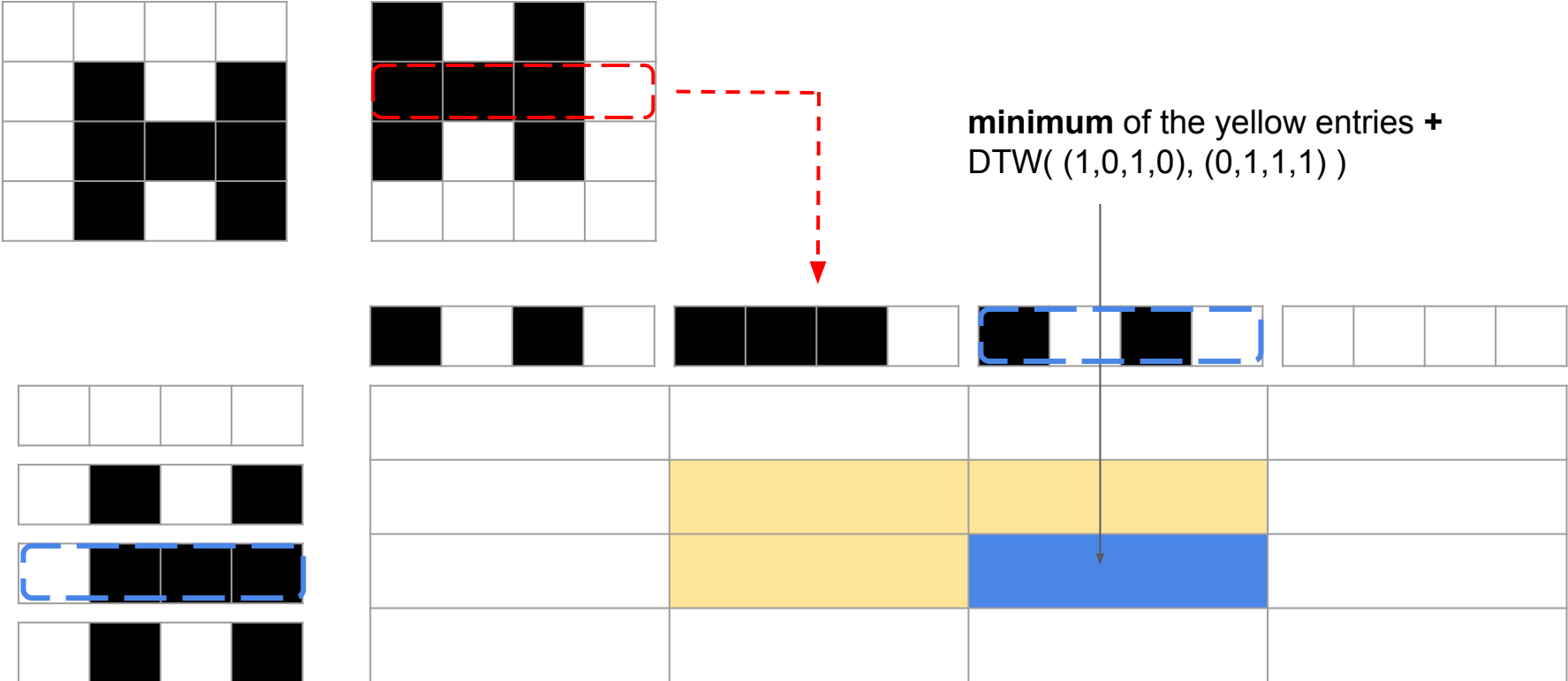DIW = Dynamic Image Warping
     = DTW with DTW as inner distance

Dynamic Image Warping - Example
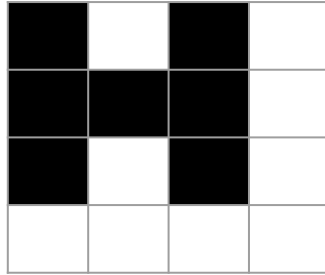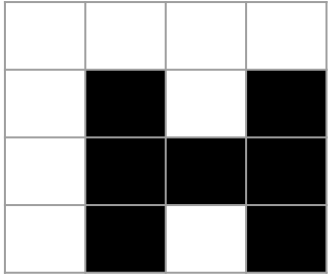
Dynamic Image Warping - Example

Dynamic Image Warping - Example

DTW((1,0,1,0), (0,0,0,0))

Dynamic Image Warping - Example

minimum of the yellow entries **+**
DTW((1,0,1,0), (0,1,1,1))
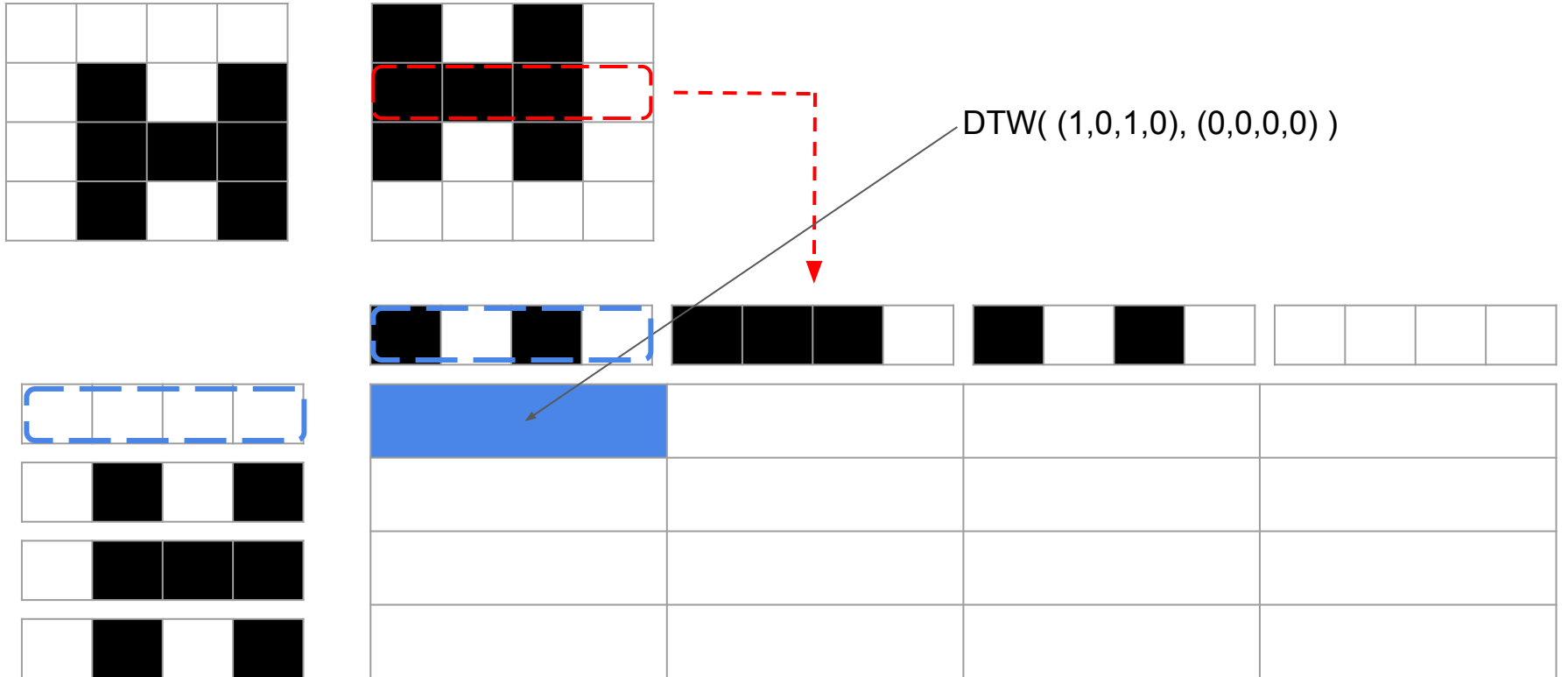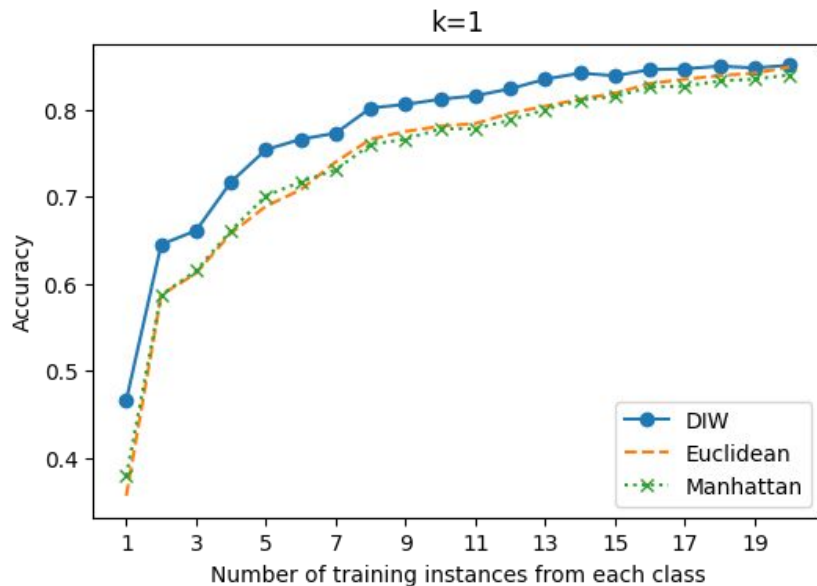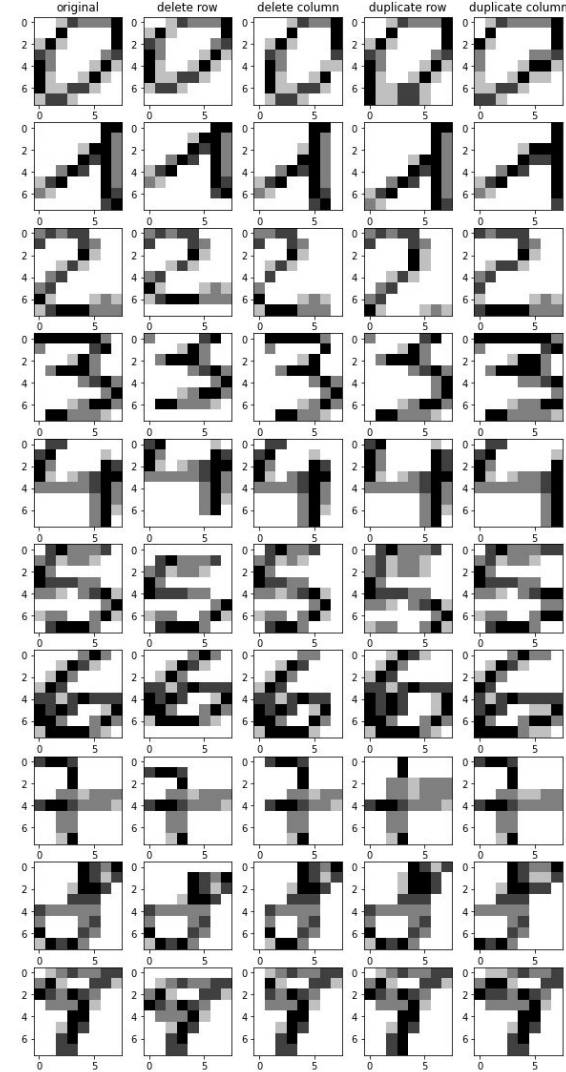
Experiments

Experiment 1 - Accuracy as function of training instances

- We implemented DIW in Cython in order to combine Python's rapid prototyping with the efficiency of C
- Nearest neighbor classification on the Semeion dataset (handwritten digits) from the UCI repository

- See also:
  https://github.com/kr7/diw/blob/main/DIW.ipynb

Experiment 2 - Data Augmentation

k=1

DIW
Euclidean
Manhattan

Accuracy

Number of augmented training instances from each class

See also:
https://github.com/kr7/diw/blob/main/DIW-augmentation.ipynb

original   delete row   delete column   duplicate row   duplicate column
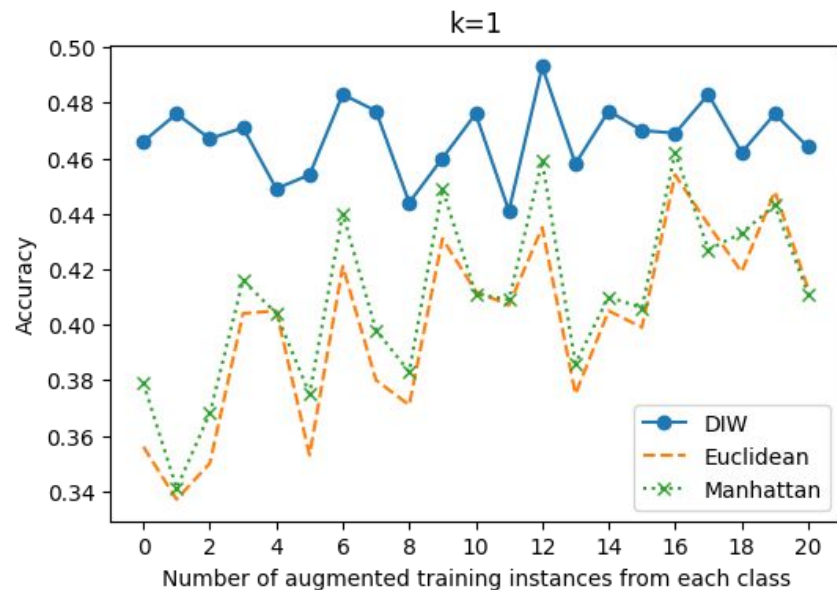
Conclusions and Outlook

Conclusions & Outlook

- Extension of dynamic time warping for images

- It may be particularly useful in cases when limited amount of training instances are available, e.g. rare diseases, user authentication

- Data augmentation does not necessarily beat a smart algorithm

- DIW may be used a block of convolutional neural networks

https://github.com/kr7/diw

The first European Statistics Awards for Web Intelligence

Eurostat is pleased to announce the **Web Intelligence Competition** as part of the **European Statistics Awards** Program. The competition aims at stimulating innovation in the area of Web Intelligence for European statistics.

The Web Intelligence **DEDUPLICATION CHALLENGE** will focus on identifying potential duplicate job postings on websites as a basic condition to produce high quality statistics from online job advertisements.

TIMELINE

The competition will be launched in the second half of **December 2022**.

Registration for participation will be open until **1 March, 2023**.

The deadline for submissions will be **31 March 2023** or **16 April 2023** depending on the chosen award. Teams can compete for three different awards.

Link Prediction in Bipartite Graphs and Its Applications to Drug-Target Interaction Prediction

https://www.youtube.com/watch?v=qq2PxN8IYYQ

http://www.biointelligence.hu/dti