FINAL PROJECT PRESENTATION

By
6231325521 Thanapat Trachu
6231340921 Pongsapak Pulthasthan

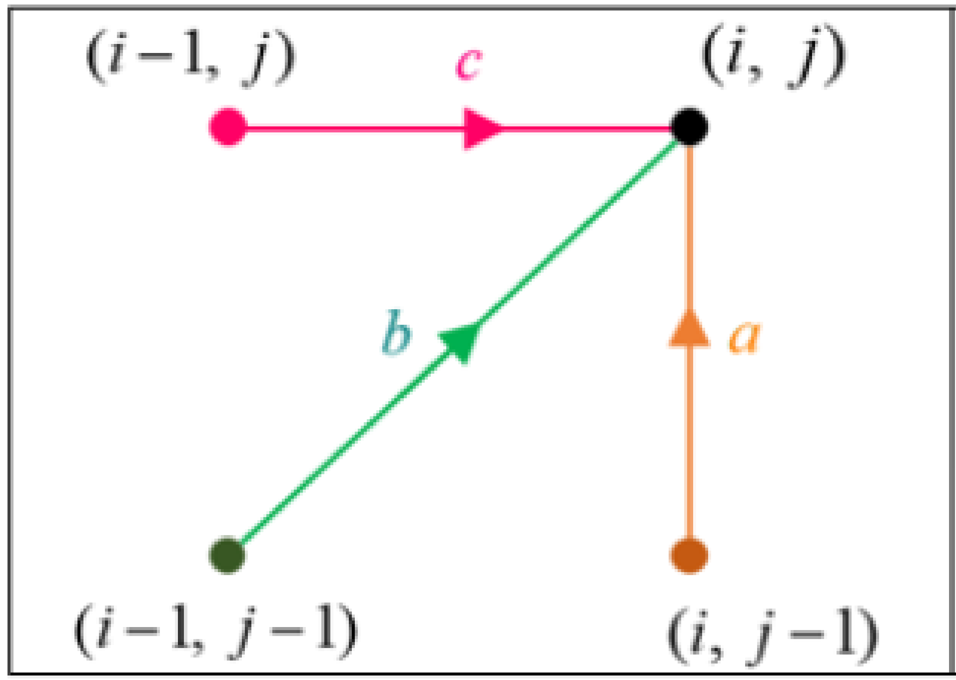
2110430 Time Series Mining and Knowledge Discovery (2022/2)

OUTLINE

1. Examining the effects of weight and slope constraint condition in Dynamic Time Warping on 1-nearest-neighbor classification accuracy
2. Shape Averaging Method for multiple time series sequences

1.) EXPERIMENT ON VARIOUS DTW

1.1) WEIGHTS



$$g(i,j) = \min \begin{cases} g(i, j-1) + a \cdot d(i, j) \\ g(i-1, j-1) + b \cdot d(i, j) \\ g(i-1, j) + c \cdot d(i, j) \end{cases}$$

Symmetric Weight

$a = c$

Example

$(a, b, c) = (1, 1, 1), (1, 2, 1)$

Asymmetric Weight

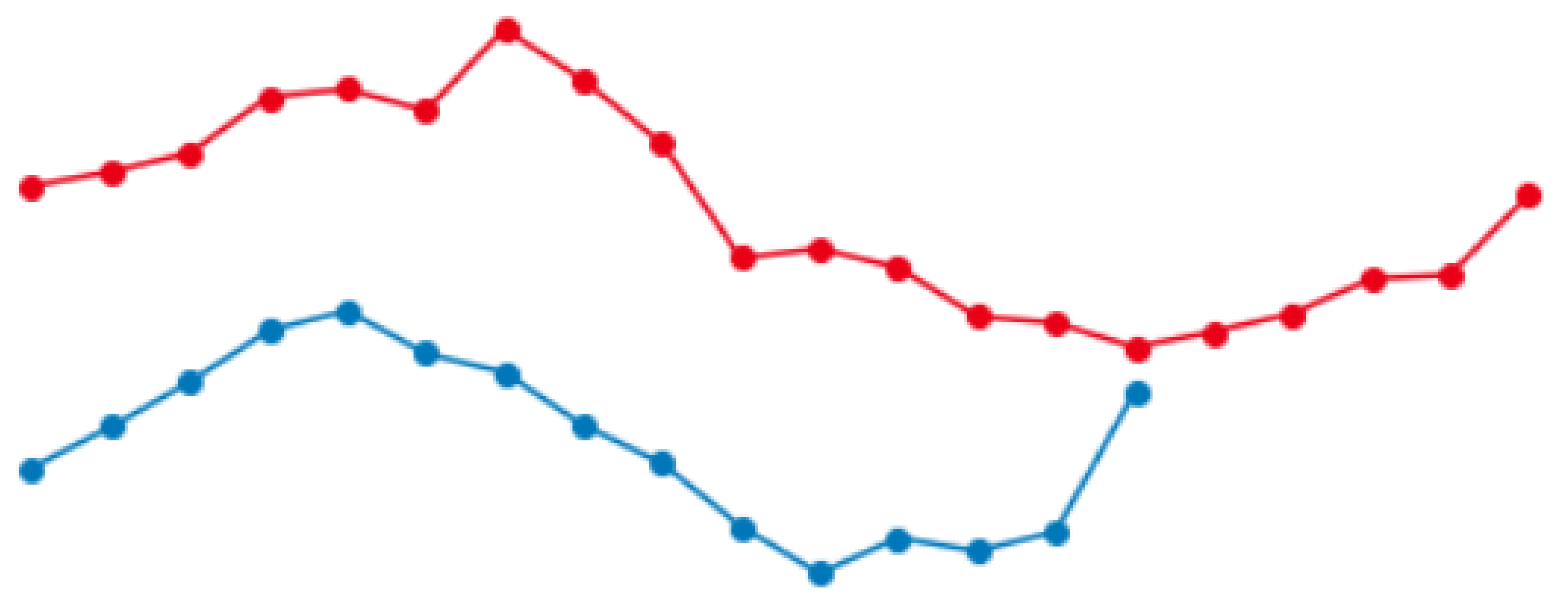
$a \neq c$

Example

$(a, b, c) = (1, 1, 0), (0, 1, 1)$

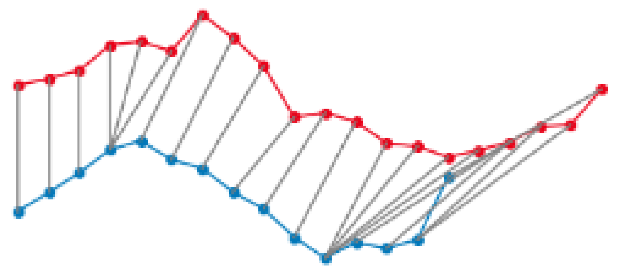
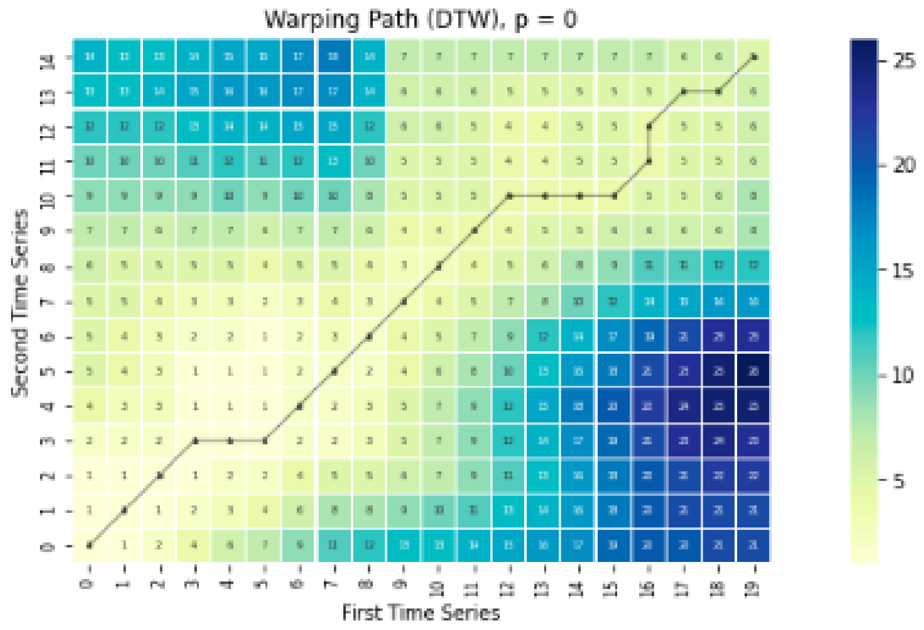
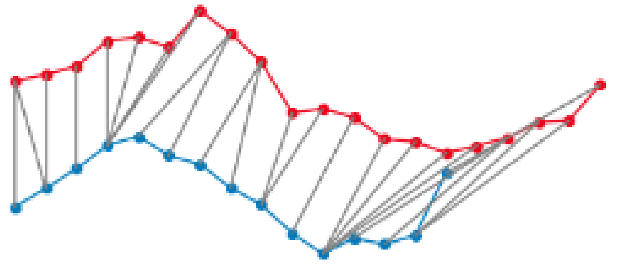
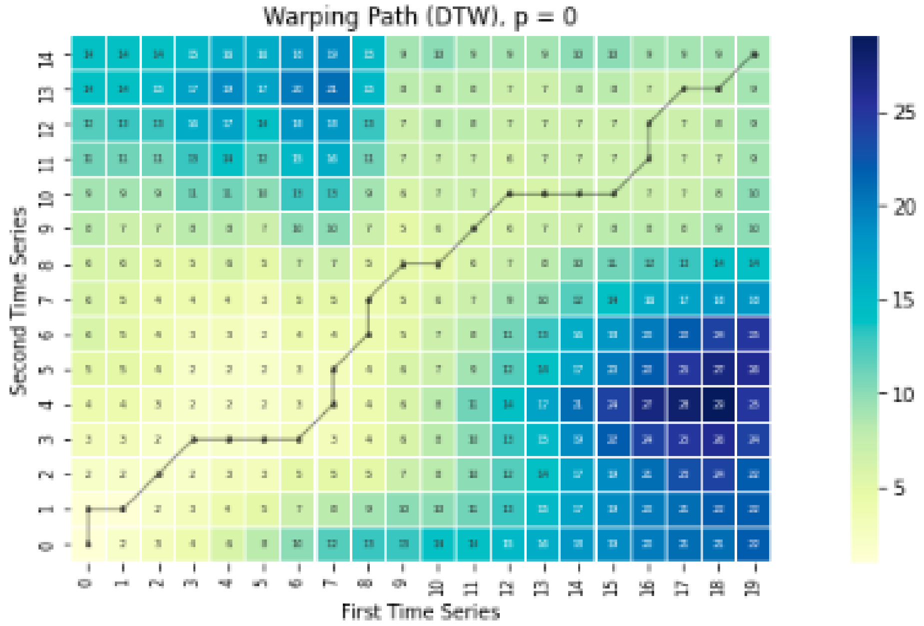
MOCK UP TIME SERIES

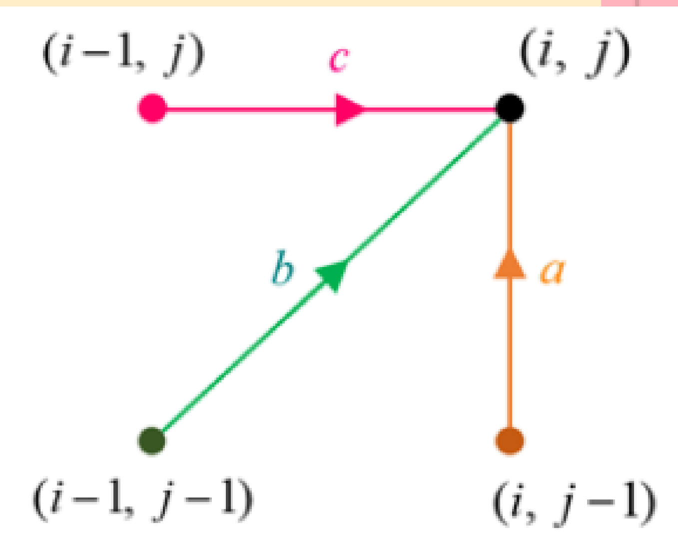
Two mock-up time series are generated (simple sinusodial waves with gaussian noise) and calculate DTW using different weights and observe the alignment obtained from calculation.



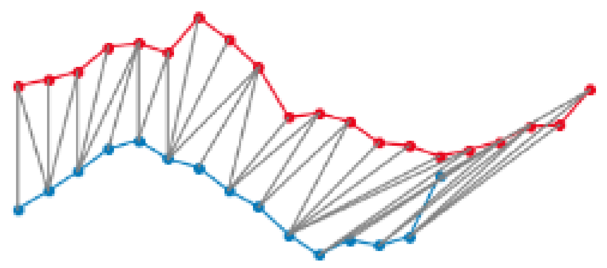
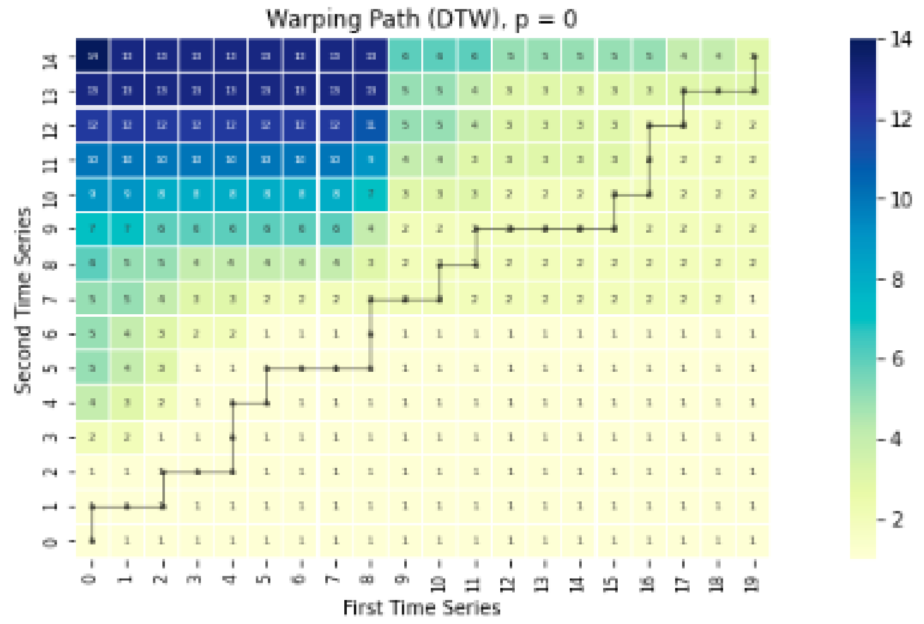
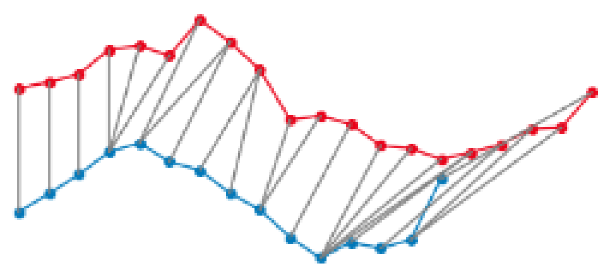
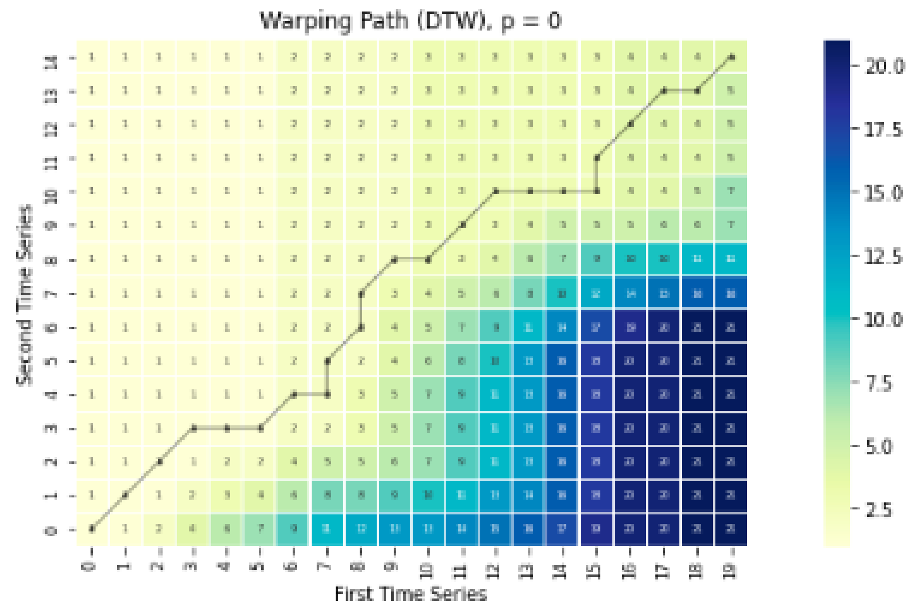
Two Mock-up Generated Time Series (A, B)

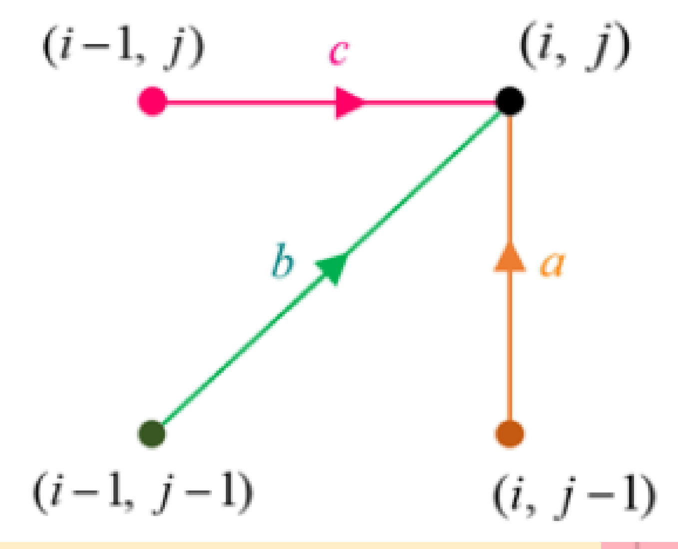
MOCK UP TIME SERIES

(a, b, c)	DTW Alignment	Cost Matrix *g* with Warping Path	*D(A,B)*
(1, 1, 1) symmetric			0.2567
(1, 2, 1) symmetric			0.2454



MOCK UP TIME SERIES

<table>
<tr><th>(a, b, c)</th><th>DTW Alignment</th><th>Cost Matrix g with Warping Path</th><th>D(A,B)</th></tr>
<tr><td>(1, 1, 0)
asymmetric</td><td></td><td>
</td><td>0.2073</td></tr>
<tr><td>(0, 1, 1)
asymmetric</td><td></td><td>
</td><td>0.2209</td></tr>
</table>



TIME SERIES DATASETS

Dataset Name	Number of classes	Size of training set	Size of testing set	Time series Length	Total File Size
Gun-Point	2	50	150	150	484 KB
Lightning-7	7	70	73	319	733 KB
Synthetic Control	6	300	300	60	586 KB
Coffee	2	28	28	286	156 KB
FaceFour	4	24	88	350	629 KB

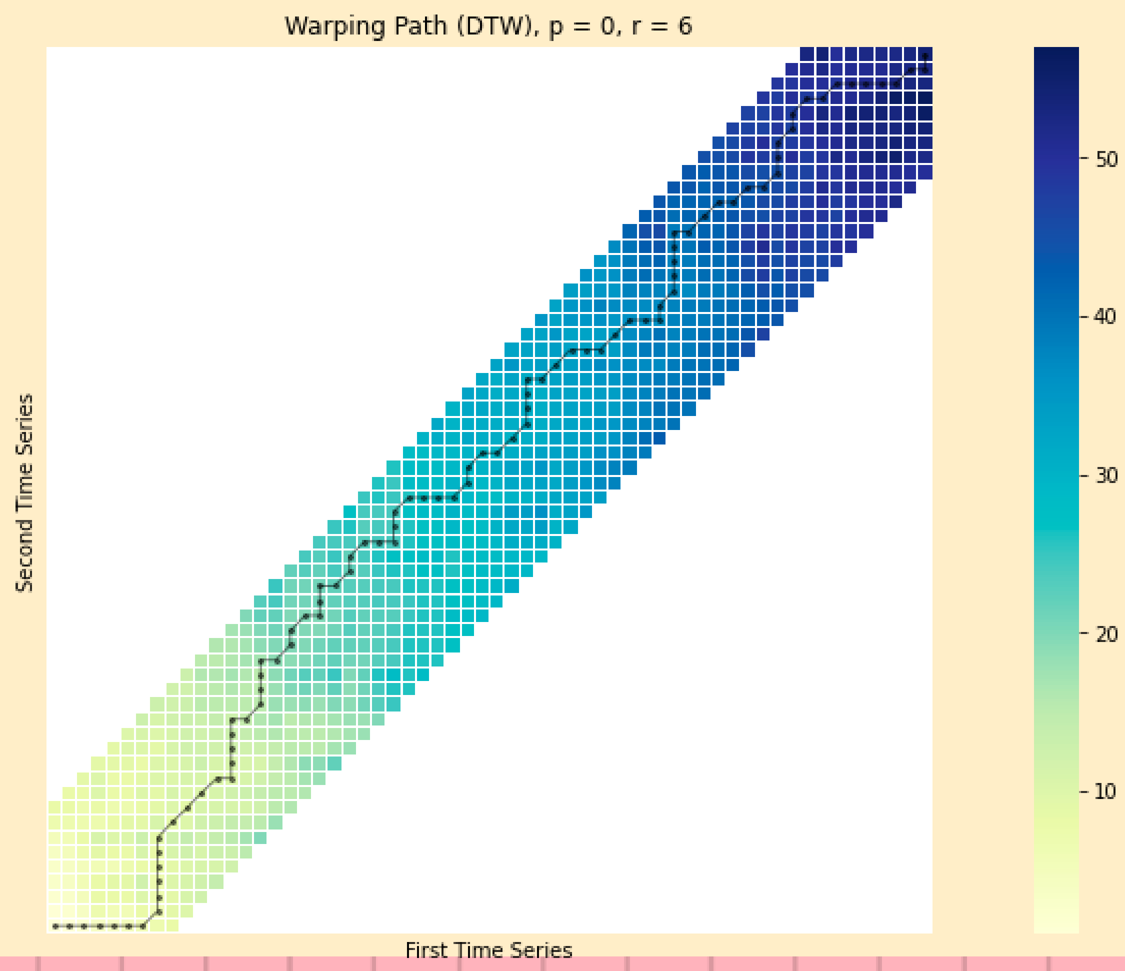
Total 2.58 MB

WINDOW LENGTH (SAKOE-CHIBA BAND)

To reduce computation time and prevent unreasonable warping in DTW.

We also implement Sakoe-Chiba Band by using window length (r) approximately 7-10% of the time-series length in each dataset.

Dataset	Time series Length	Window Length (r)	width (%)
Gun-Point	150	12	8.00%
Lightning-7	319	24	7.52%
Synthetic Control	60	6	10.00%
Coffee	286	28	9.79%
FaceFour	350	28	8.00%



r = 6 (Synthetic Control Dataset)

1-NN CLASSIFICATION ACCURACY

Dataset	Weight *(a, b, c)*						
	Symmetric			Asymmetric			
	(1, 1, 1)	(1, 2, 1)	(1, 0.5, 1)	(1, 1, 0)	(0, 1, 1)	(2, 1, 0.5)	(0.5, 1, 2)
Gun-Point	0.9000	**0.9200**	**0.8467**	0.9067	0.9267	0.9133	**0.9200**
Lightning-7	0.7808	0.7808	0.8219	0.7534	**0.6986**	**0.8493**	0.7260
Synthetic Control	0.9600	0.9400	0.9567	**0.7134**	0.9800	0.8133	**0.9900**
Coffee	0.8214	0.7857	0.8214	**0.6429**	**0.8571**	0.7857	0.7857
FaceFour	0.8181	0.8409	0.7613	**0.4773**	0.8409	0.7159	**0.8523**

CONCLUSION (WEIGHT)

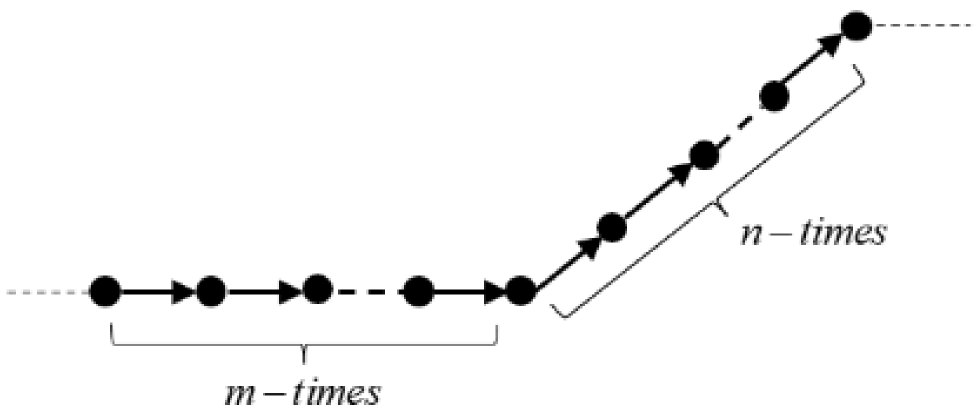
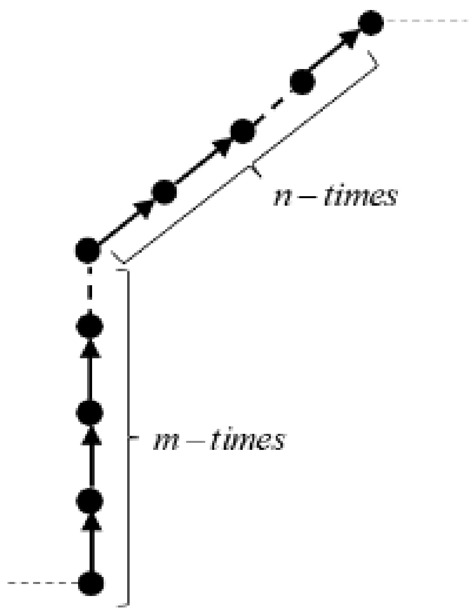
- Using "0" in a weight is not recommended as this will exclude some feature out from the calculation which result in bad performance.
- Using symmetric weight usually give acceptable performance, but it may not the best.
- It's likely that using asymmetric weight might obtain the optimal accuracy, but we cannot tell which weight works best for each dataset. Grid Search is needed but it may took long time to compute.

1.2) SLOPE CONSTRAINT

- Neither too steep nor too gentle warping path's gradient shouldn't be allowed as it may cause undesirable time-axis warping.

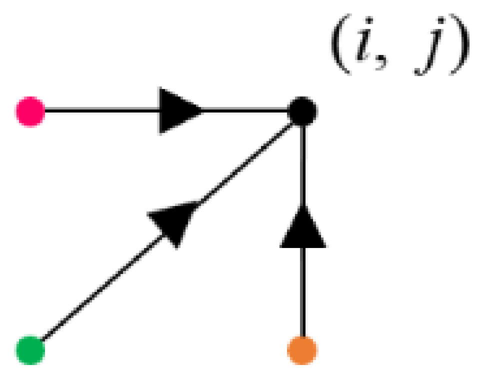
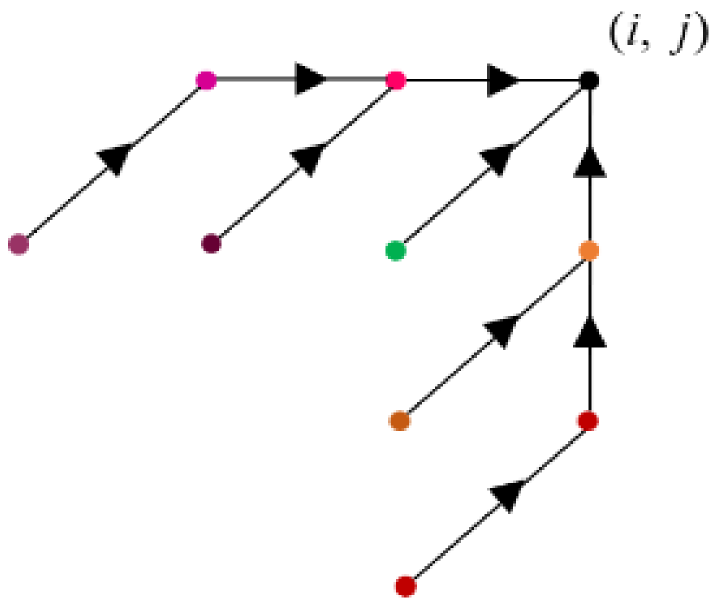
Slope Constraint Condition (p)

$$p = \frac{n}{m}$$

p	Minimum Slope	Maximum Slope
n/m		

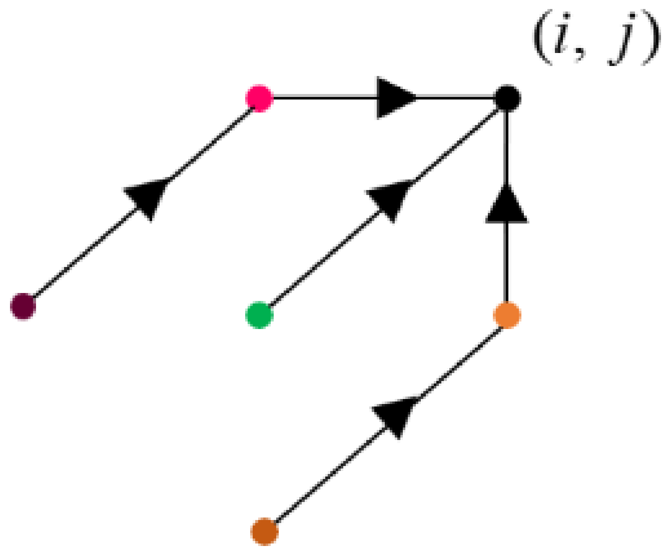
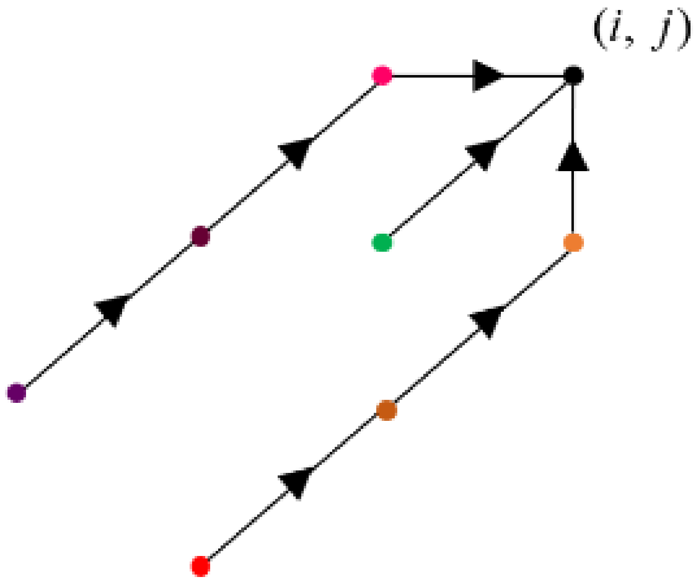
DTW VARIATIONS

Symmetric DP-Algorithms with Slope Constraint Condition (p = 0, 1/2, 1, 2)

p	Schematic Explanation	DP-equation $g(i, j)$ =
0		$\min\begin{cases} g(i, j-1) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i-1, j) + d(i, j) \end{cases}$
½*		$\min\begin{cases} g(i-1, j-3) + 2d(i, j-2) + d(i, j-1) + d(i, j) \\ g(i-1, j-2) + 2d(i, j-1) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i-2, j-1) + 2d(i-1, j) + d(i, j) \\ g(i-3, j-1) + 2d(i-2, j) + d(i-1, j) + d(i, j) \end{cases}$

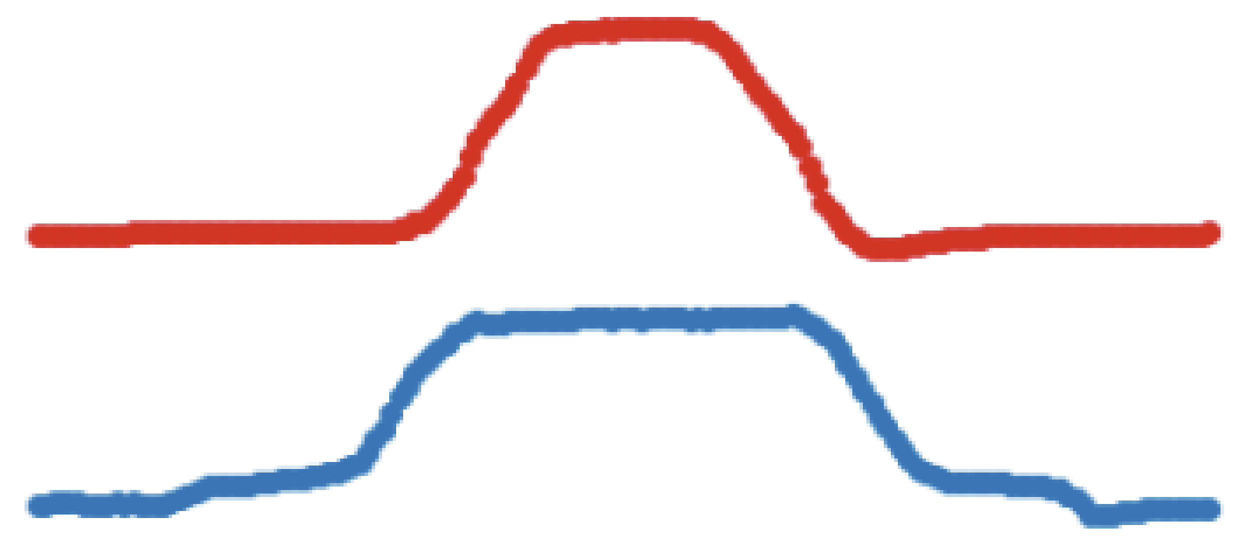
DTW VARIATIONS

Symmetric DP-Algorithms with Slope Constraint Condition (p = 0, 1/2, 1, 2)

1		$\min \begin{cases} g(i-1, j-2) + 2d(i, j-1) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i-2, j-1) + 2d(i-1, j) + d(i, j) \end{cases}$
2		$\min \begin{cases} g(i-2, j-3) + 2d(i-1, j-2) + 2d(i, j-1) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i-3, j-2) + 2d(i-2, j-1) + 2d(i-1, j) + d(i, j) \end{cases}$

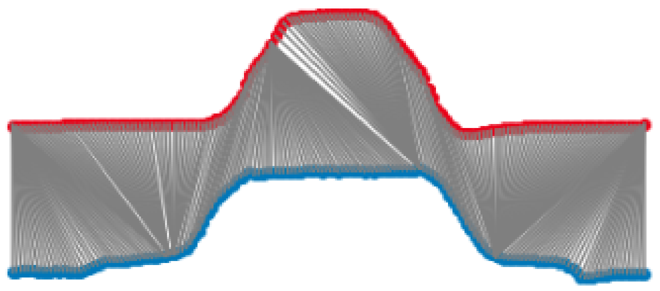
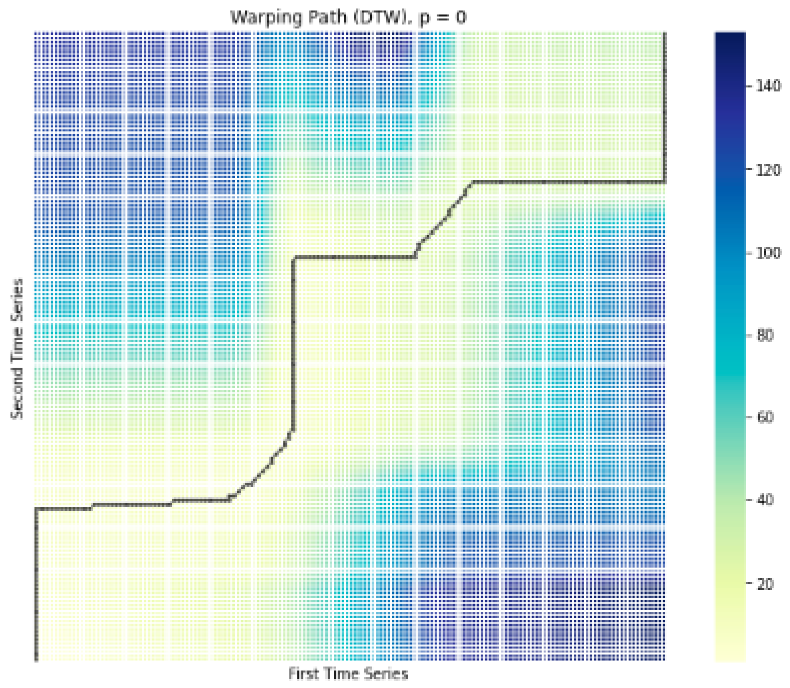
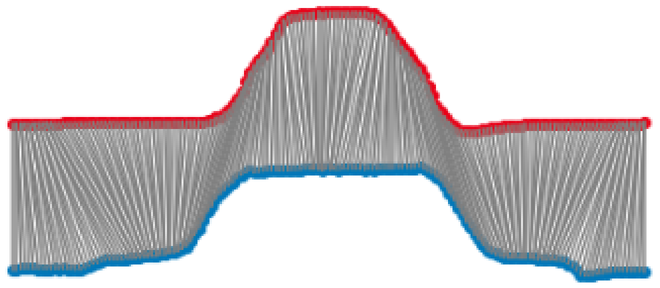
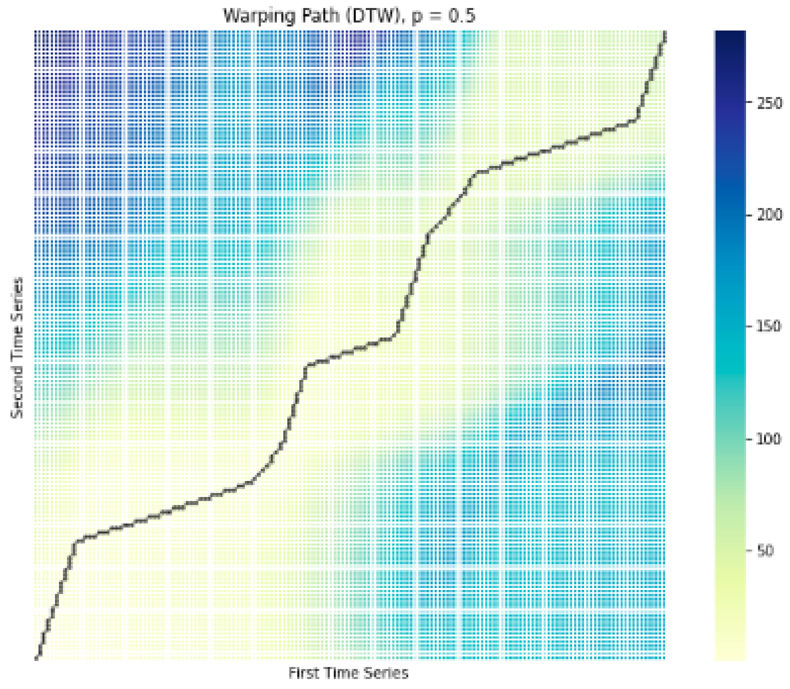
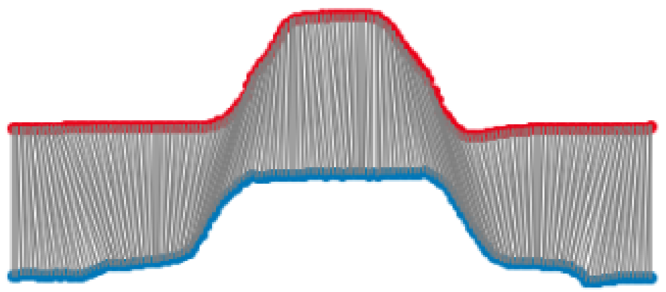
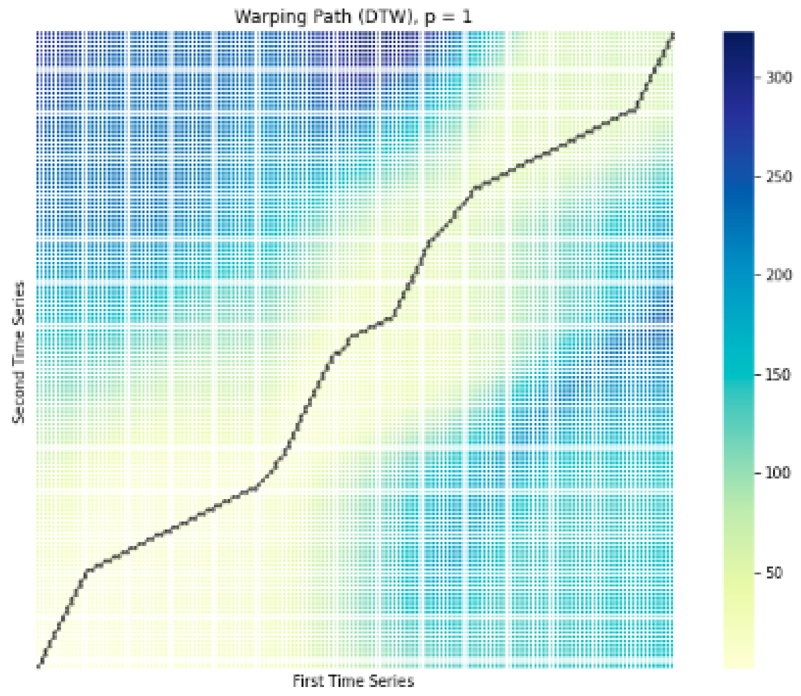
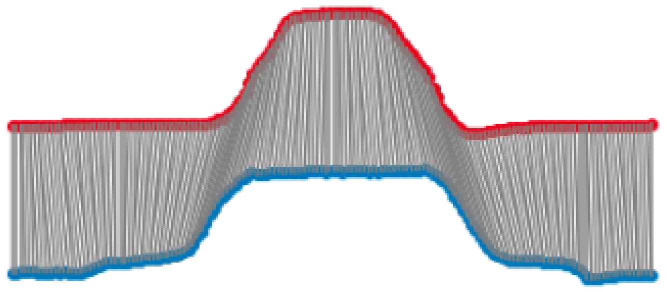
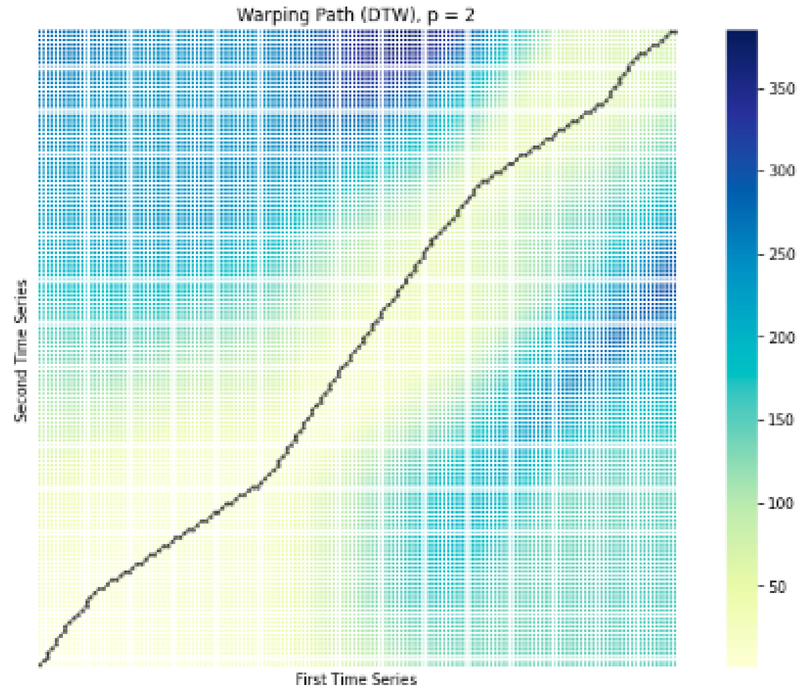
MOCK UP TIME SERIES

Two mock-up time series are sampled from Gun-Point dataset to visualize the difference of the warping path obtained from using diffrent p values.

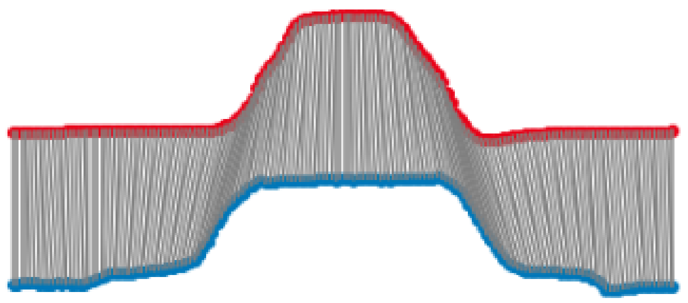
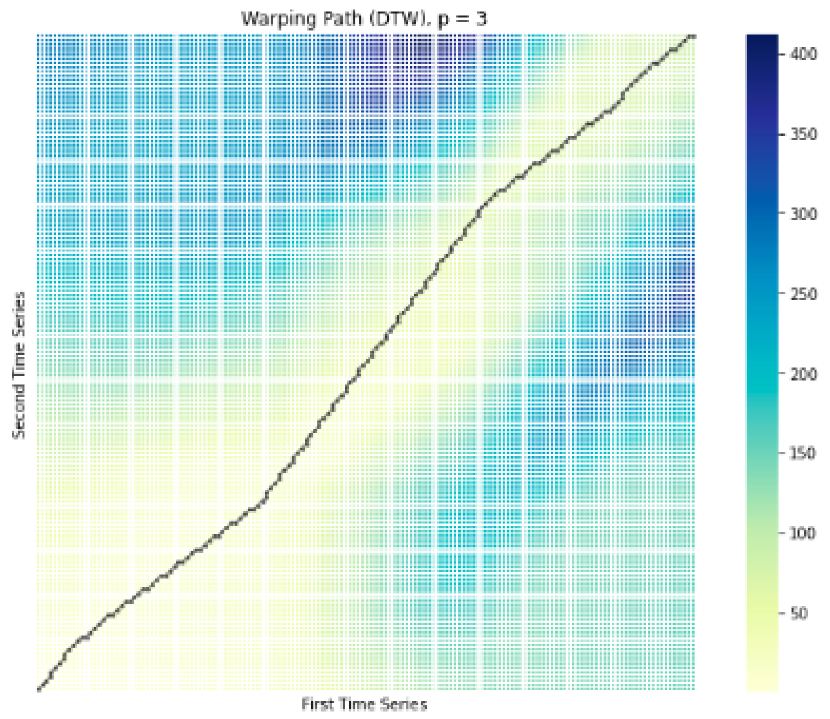
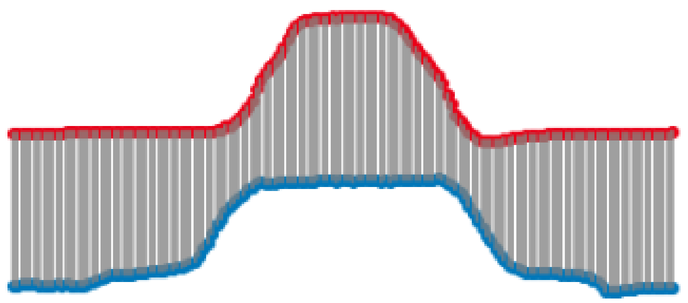
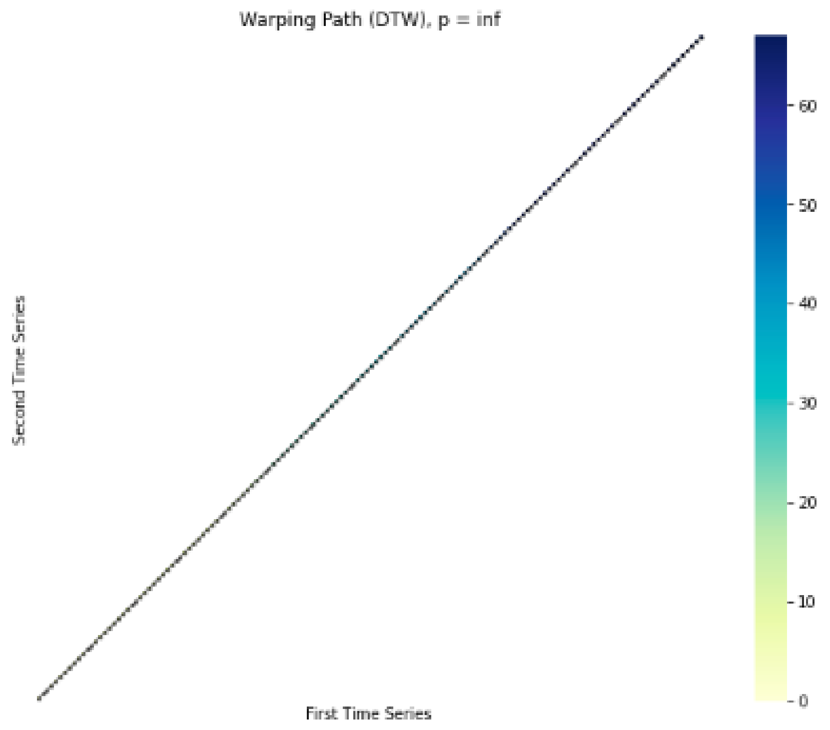


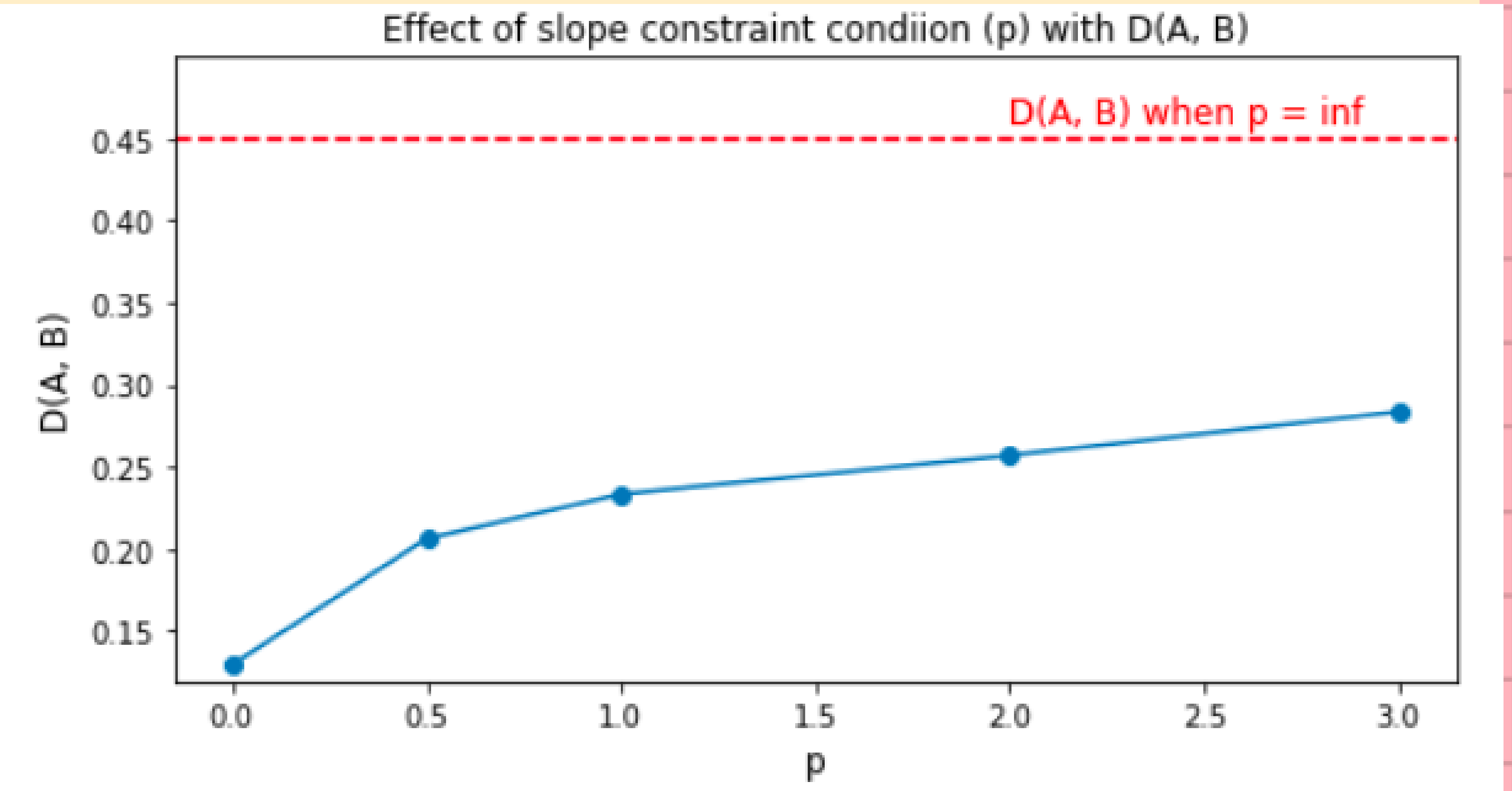
Two Mock-up Time Series from Gun-Point Dataset (A, B)

MOCK UP TIME SERIES

p	DTW Alignment	Cost Matrix *g* with Warping Path	*D(A,B)*
0			0.1294
1/2			0.2062
1			0.2332
2			0.2568

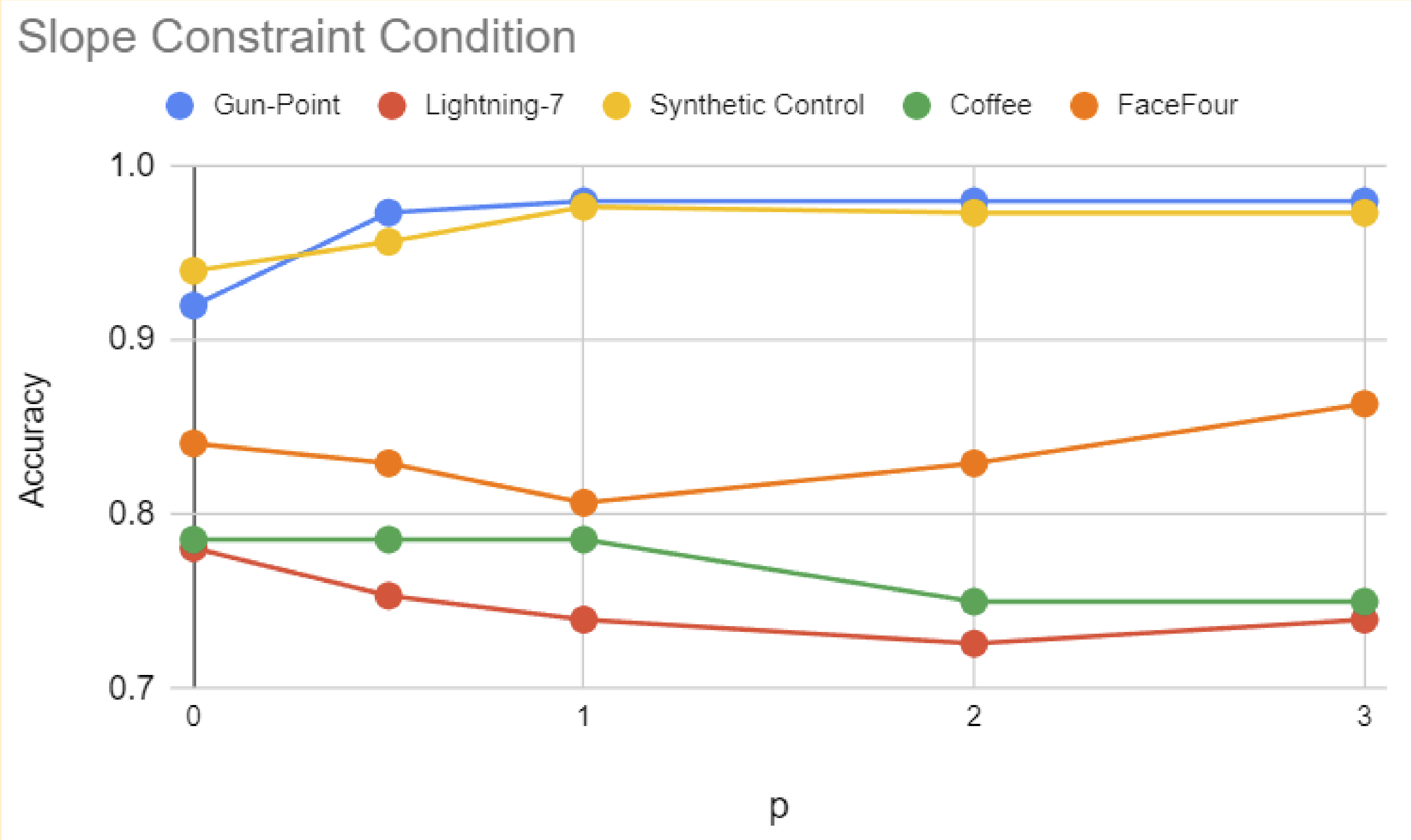
MOCK UP TIME SERIES

3			0.2834
inf			0.4496



1-NN CLASSIFICATION ACCURACY

Dataset	Slope Constraint Value (p)				
	p = 0	p = 1/2	p = 1	p = 2	p = 3
Gun-Point	0.9200	0.9734	0.9800	0.9800	0.9800
Lightning-7	0.7808	0.7534	0.7397	0.7260	0.7397
Synthetic Control	0.9400	0.9567	0.9767	0.9733	0.9733
Coffee	0.7857	0.7857	0.7857	0.7500	0.7500
FaceFour	0.8409	0.8295	0.8068	0.8295	0.8636

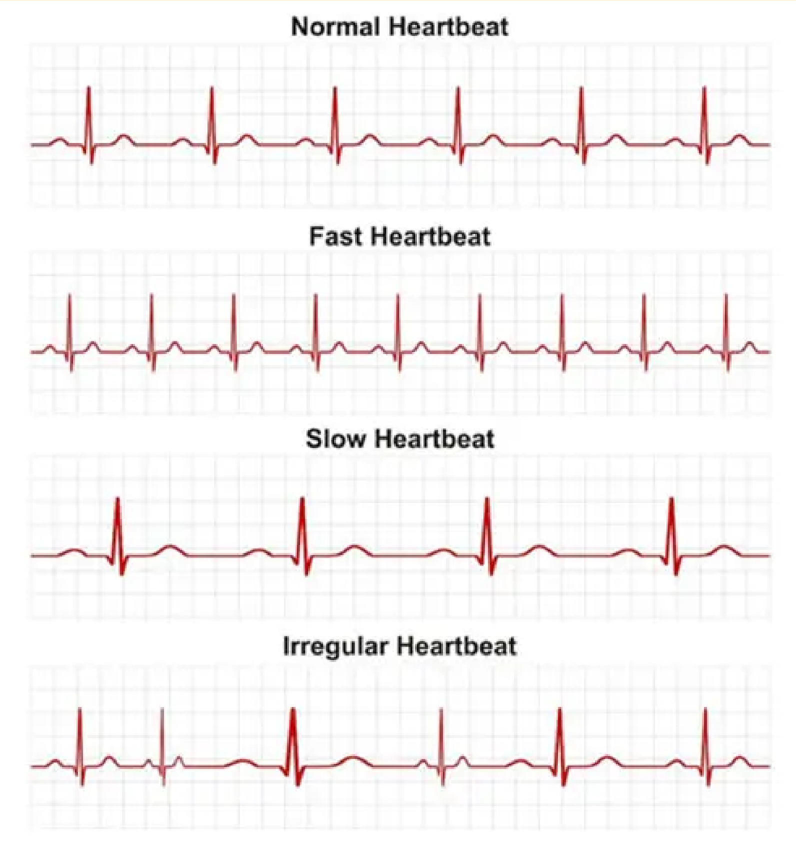


CONCLUSION (SLOPE CONSTRAINT)

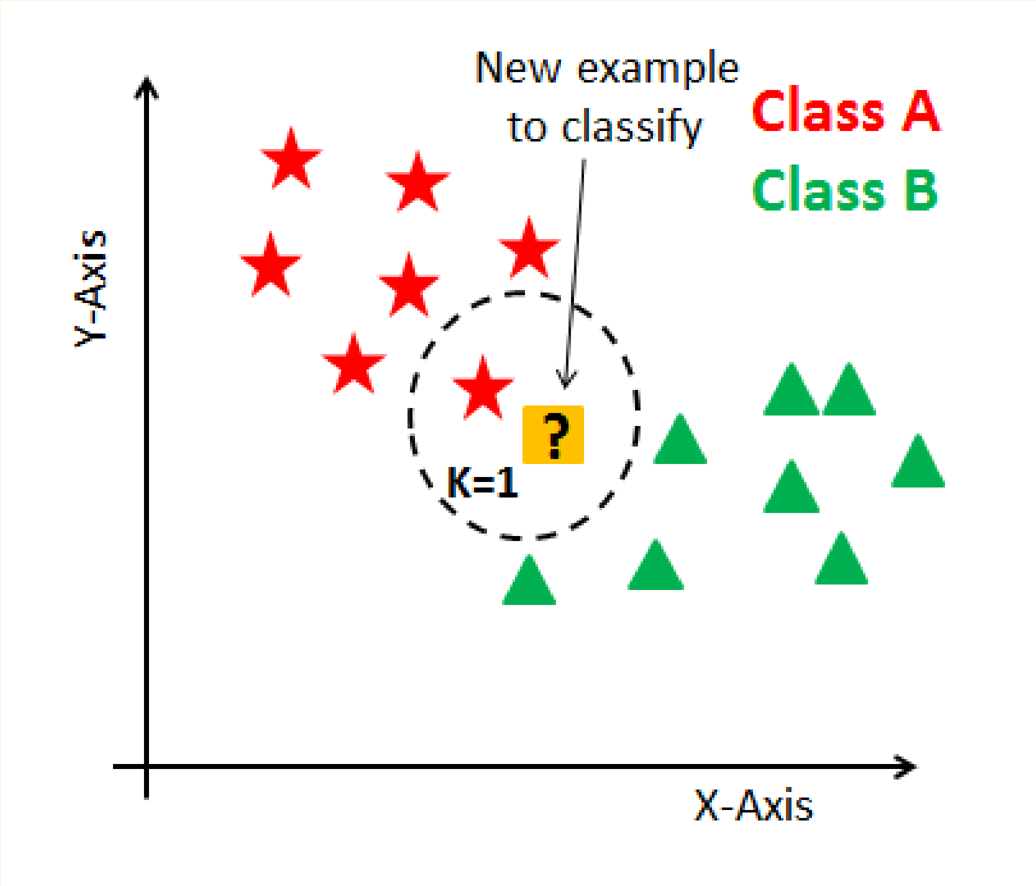
- When the value of p = 0 (no slope constraint condition is employed), there might be a case that the warping path is too steep or too gentle which may cause an undesirable time-axis warping which sometimes can result in poor performance in classification accuracy
- The preferable value of p should not also be too high. For the reason that if the slope condition is too strict, then the dynamic time warping algorithm might not handle the time fluctuation in time series data
- Try using p around 1/2, 1, 2 is likely to obtain the optimal accuracy in most datasets. This is because the warping path obtained is more sensible and the restriction is not too lax or too strict :).

2.) SHAPE AVERAGING METHOD FOR MULTIPLE TIME SERIES SEQUENCES

MOTIVATION

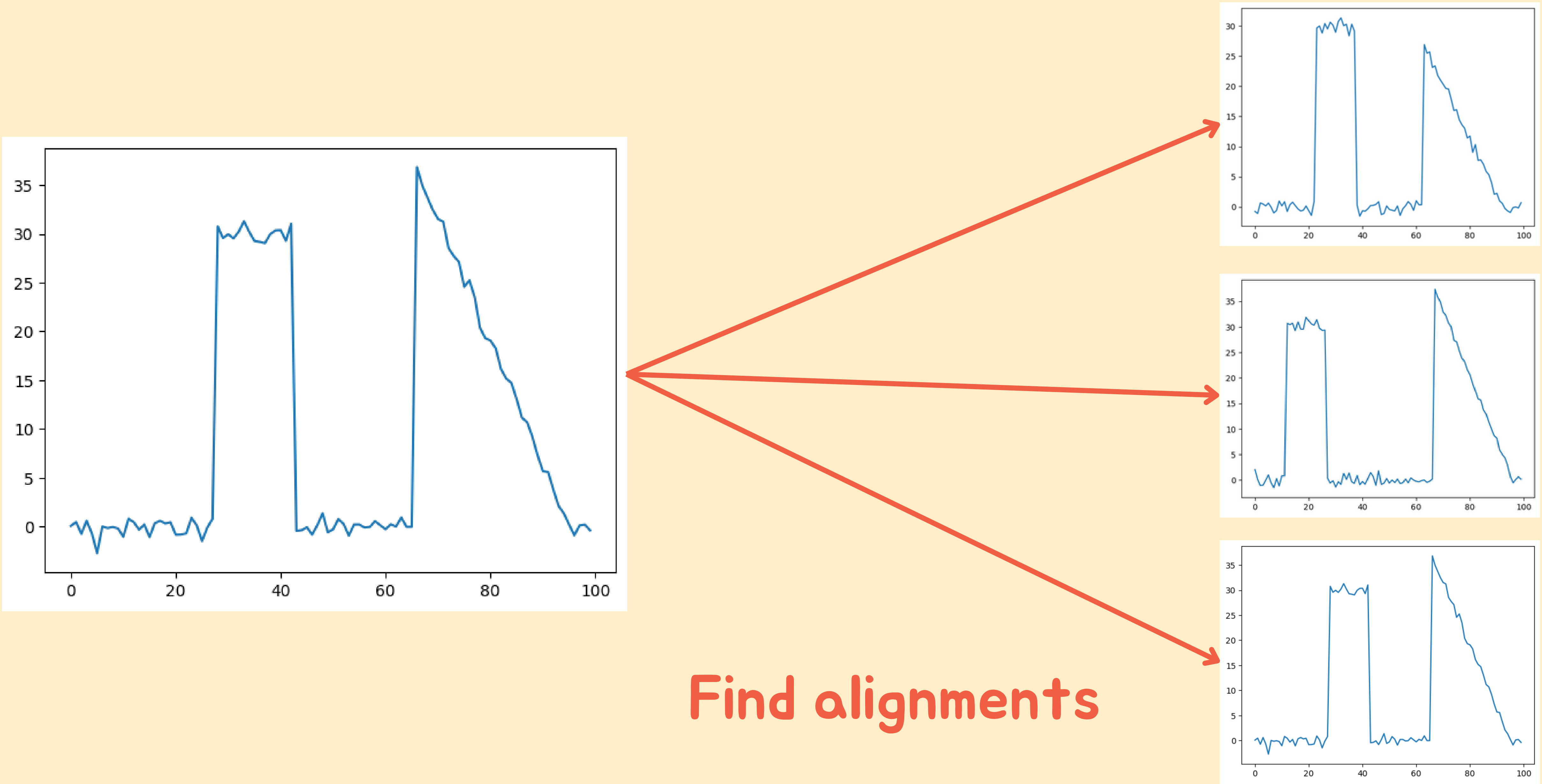


Time Series Classification



One Nearest Neighbor

DTW BARYCENTER AVERAGING



Pivot Time series

Input multiple time series

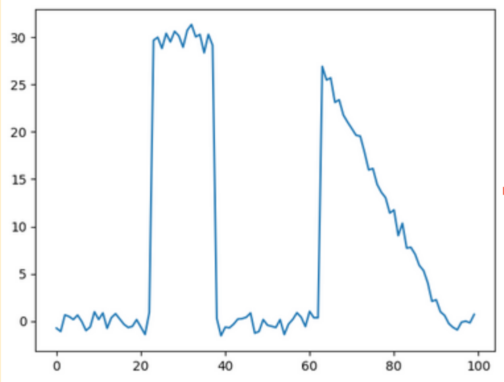
INITIALIZE PIVOT TIME SERIES

$$T_n = \{T_{n_1}, T_{n_2}, \cdots T_{n_l}\}$$

$$S_n = \sum_{i=1}^{l} T_{n_i}$$

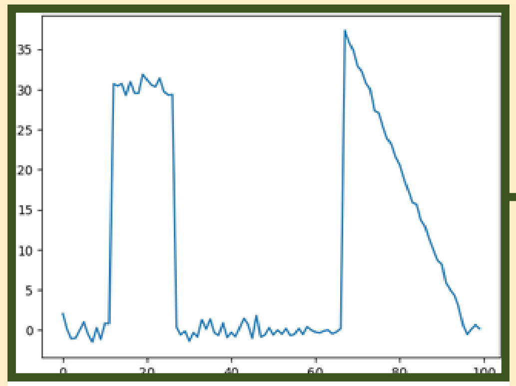
$$\text{mean} = \frac{(\sum_{i=1}^{k} S_i)}{k}$$

T1



S1

T2



Pivot

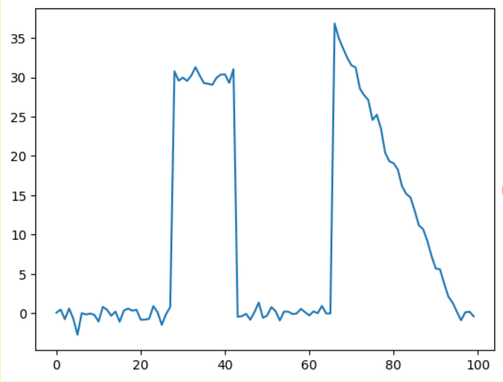
Time series

S2

closest

Mean

T3



S3

$$\text{Pivot} = \arg \min_{n \in \{1 \cdots k\}} (|\text{mean} - S_n|)$$

Input multiple time series

sum of time series

mean of sum

DTW SHAPE AVERAGING

Other	Cost						
5	5.25	9.25	1.25	1.25	0.25	0.25	
4	5.25	5.25	1.25	0.25	1.25	2.25	
3	4.25	4.25	0.25	1.25	1.25	1.25	
2	4.25	0.25	4.25	3.5	6.75	7	
1	0.25	2.25	2.5	2.75	3	3.25	
0	0	4	4	5	5	5	
	0	1	2	3	4	5	Pivot

Cost Matrix

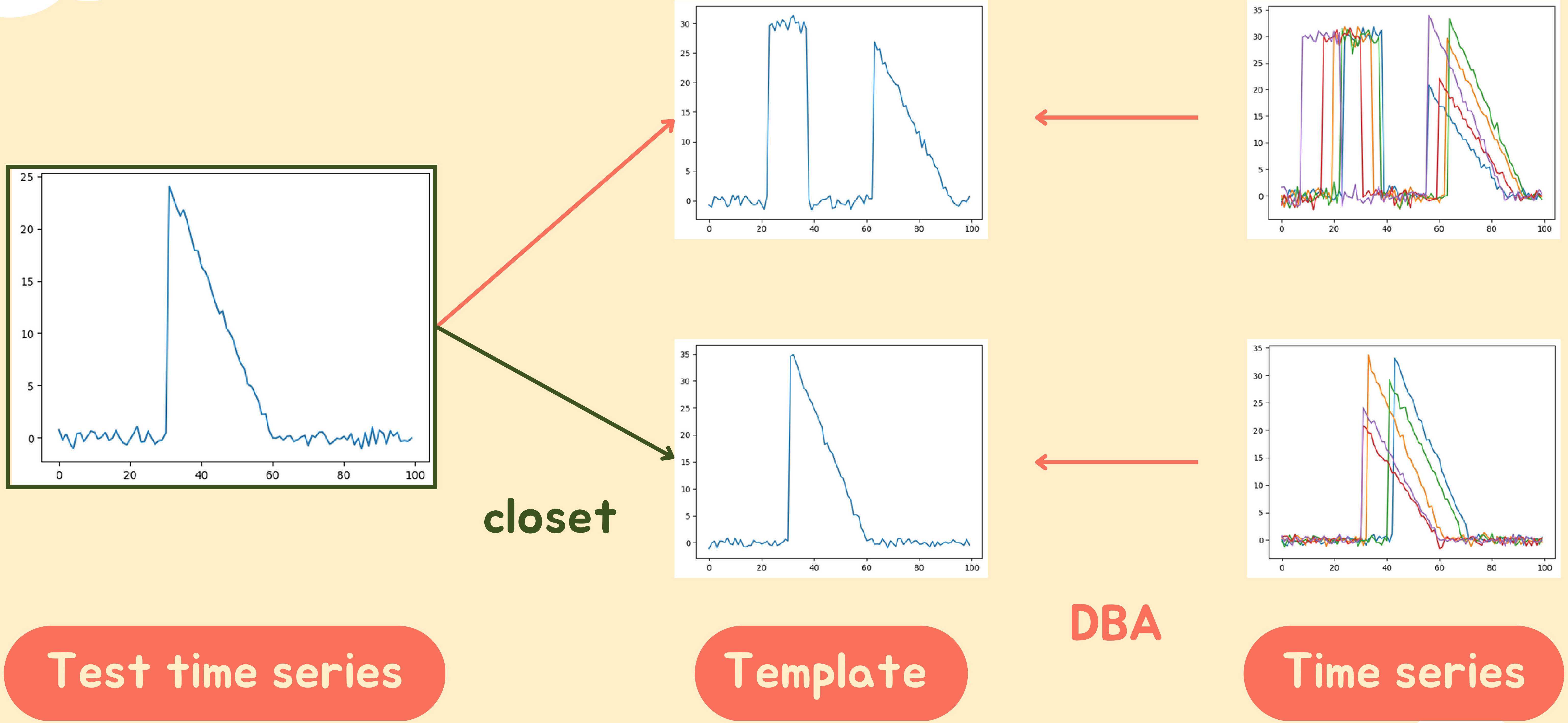
Pivot	0	2	0	1	0	0
Other	0	0.5	2	0	1	0

Time series

Alignment	[0.5, 0]	[2]	[0]	[1]	[0]	[0]
Result	0.25	2	0	1	0	0

Result Time series

EVALUATION



ONE NEAREST NEIGHBOR ACCURACY

Dataset	DBA	Amplitude Avg
Gun Point	66%	52%
Lighting7	75.71%	41.43%
synthetic control	100%	78%
Coffee	75%	75%

THANK YOU FOR LISTENING