Mining and Processing Biomedical Data

Dr. rer. nat. Krisztian Buza
adiunkt naukowy
Faculty of Mathematics, Informatics and Mechanics
University of Warsaw, Poland
chrisbuza@yahoo.com

Preprocessing Time Series

Time Series

- Sequence of numbers:

  10, 12, 13 …

- Example: observation of the temperature

  6:00 am – 10 ºC
  7:00 am – 12 ºC
  8:00 am – 13 ºC
  9:00 am – 15 ºC
  ...

Multivariate Time Series

- Electrocardiograph (ECG) and Electroencephalograph (EEG) signals
- Several values (observations) at each point of time

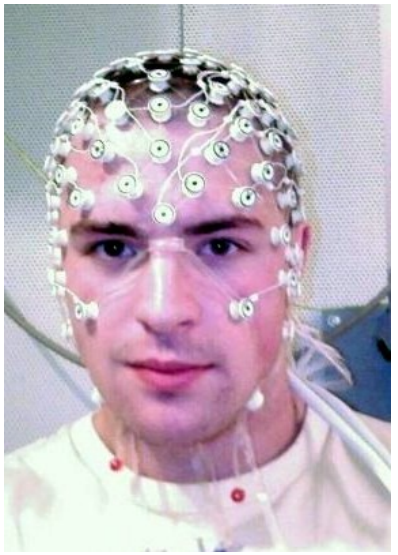


Image: http://en.wikipedia.org/wiki/File:EEG_cap.jpg

Excerpt from the publicly available EEG database at https://archive.ics.uci.edu/ml/datasets/EEG+Database

Time point	Senor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	...
1	-8.921	0.834	-19.847	8.148	-2.146	...
2	-8.433	3.276	-12.522	1.801	-2.146	...
3	-2.574	5.717	1.149	-2.594	-1.658	...
4	5.239	7.67	14.821	-4.547	-0.682	...
5	11.587	9.623	20.681	-5.035	2.248	...
...	...	...	...	...	...	...

Nearest Neighbor Classification of Time Series

- How to compare two time series?

Nearest Neighbor Classification of Time Series

- How to compare two time series?
- Compare the values one by one:

first time series:	-1.34	-2	3.5	1.7	...
second time series:	0.32	-1.5	2.8	0.9	...

Nearest Neighbor Classification of Time Series

- How to compare two time series?
- Compare the values one by one:

first time series:	-1.34	-2	3.5	1.7	...
second time series:	0.32	-1.5	2.8	0.9	...

Nearest Neighbor Classification of Time Series

- How to compare two time series?
- Compare the values one by one:

first time series:	-1.34	-2	3.5	1.7	...
second time series:	0.32	-1.5	2.8	0.9	...

Nearest Neighbor Classification of Time Series

- How to compare two time series?
- Compare the values one by one:

first time series:	-1.34	-2	3.5	1.7 ...
second time series:	0.32	-1.5	2.8	0.9 ...

- Problems with this simple approach
  - *What if time series are of different length?*
  - *Shiftings and elongations*, i.e., patterns reflecting real-world phenomena may have various length and may begin at (slightly) different position of the time-series
  - *Noise*

Preprocessing of Time Series

- Fourier transformation
- Aggregation of consequitve values
- Transformation to a symbolic representation
- Moving average
- Normalisation
- ...

Aggregation of consequtive values

- Time series:

- Aggregated time-series:

1.63 , 2.4 , ...

(Average of the first 10 values)

(Average of the next 10 values)

Aggregation of consequtive values

- Time series:

1.0, 1.2, 1.3, 1.7, 1.7, 1.8, 1.8, 1.9, 1.9, 2.0, 2.1 2.2 …

- Aggregated time-series:

1.63 , 2.4 , ...

(Average of the first 10 values)

(Average of the next 10 values)

- Symbolic representation

B , C , ...

0..1	1..2	2..3	...
A	B	C	...

Moving Average

- Time series:

1.0, 1.2, 1.3, 1.7, 1.7, 1.8, 1.8, 1.9, 1.9, 2.0, 2.1 2.2 …

- Transformed time-series:

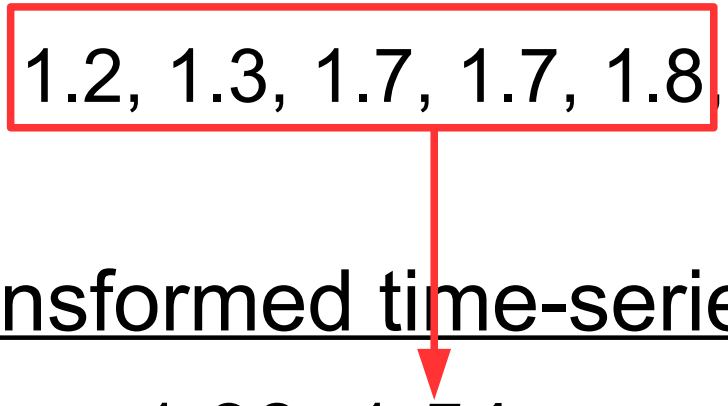
1.38

(Average of the first 5 values)

Moving Average

- Time series:

1.0, 1.2, 1.3, 1.7, 1.7, 1.8, 1.8, 1.9, 1.9, 2.0, 2.1 2.2 …



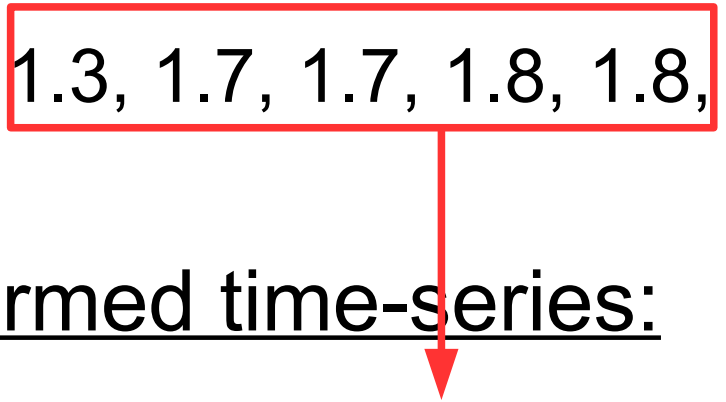
- Transformed time-series:

1.38, 1.54,

Moving Average

- Time series:

1.0, 1.2, 1.3, 1.7, 1.7, 1.8, 1.8, 1.9, 1.9, 2.0, 2.1 2.2 …



- Transformed time-series:

1.38, 1.54, ...

Normalisation

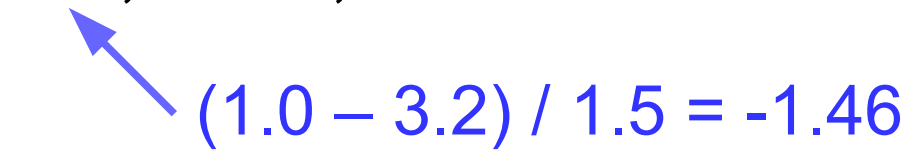
- Time series:

  1.0, 1.2, 1.3, 1.7, 1.7, 1.8, 1.8, 1.9, 1.9, 2.0, 2.1 2.2 …

- Average of all the values: 3.2
  Standard deviation of all the values: 1.5

- Normalised time-series:

  -1.46, -1.33, -1.26 ...



Dynamic Time Warping (DTW)

Proximity measures for time series

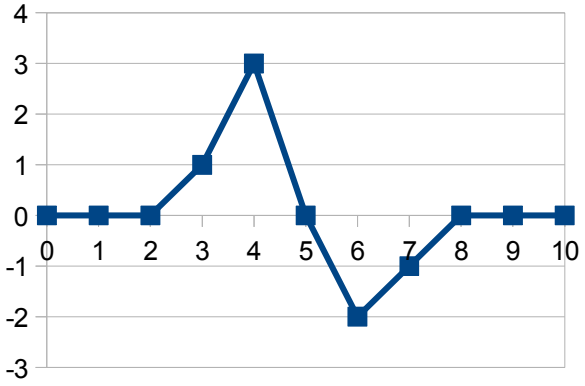
- Similarity measures and distance measure
- Similarity measure
  - High value → two time series are similar
  - Low value → two time series are different
- Distance measure
  - High value → two time series are different (dissimilar)
  - Low value → two time series are similar
- Dynamic Time Warping (DTW) is a distance measure

Example

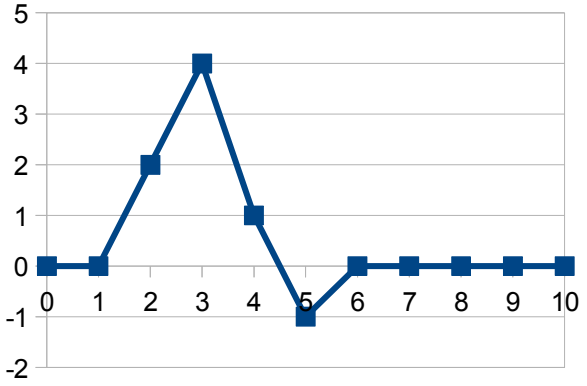
x*: 0, 0, 0, 1, 3, 0, -2, -1, 0, 0, 0

x1: 0, 0, 2, 4, 1, -1, 0, 0, 0, 0, 0

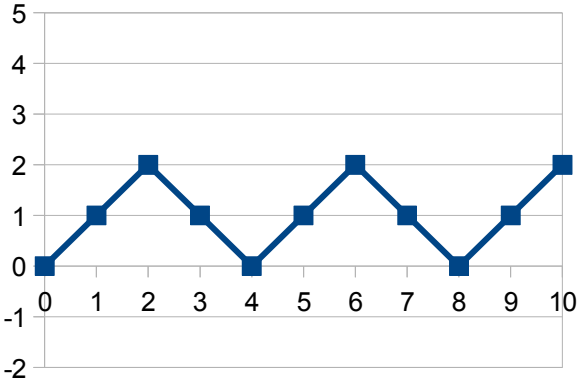
x2: 0, 1, 2, 1, 0, 1, 2, 1, 0, 1, 2



x*

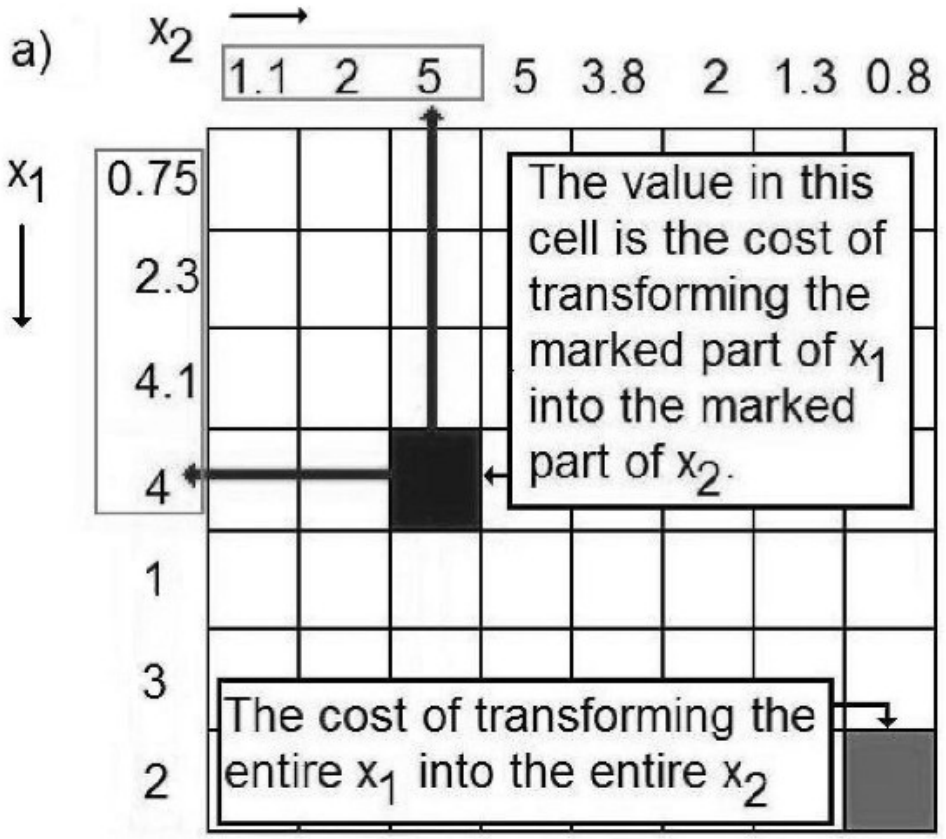


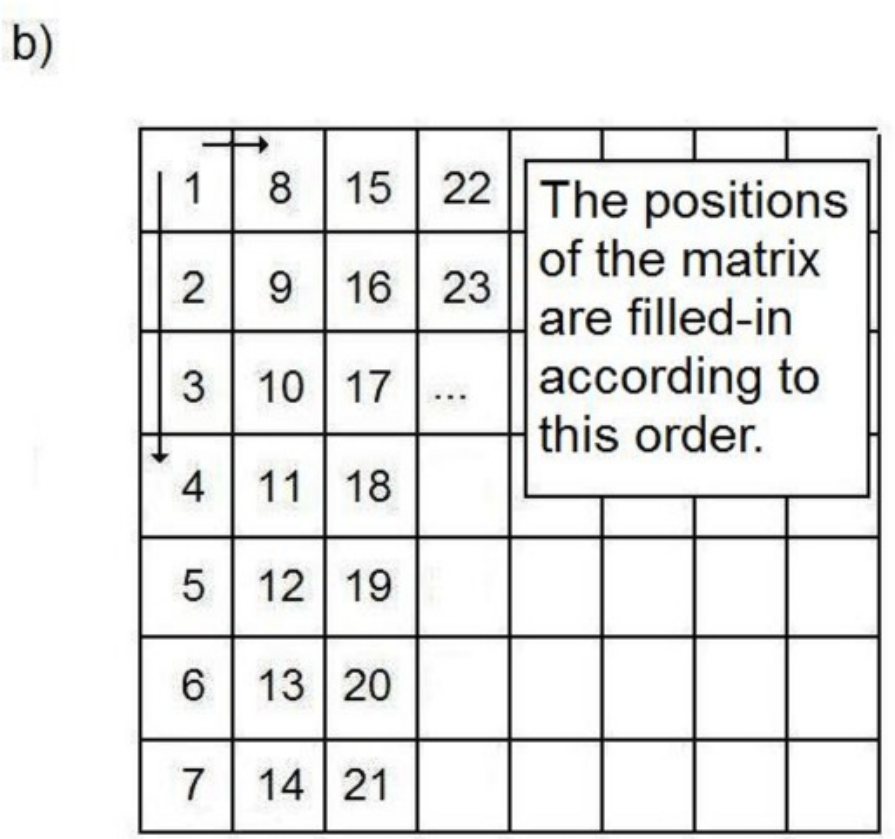
x1



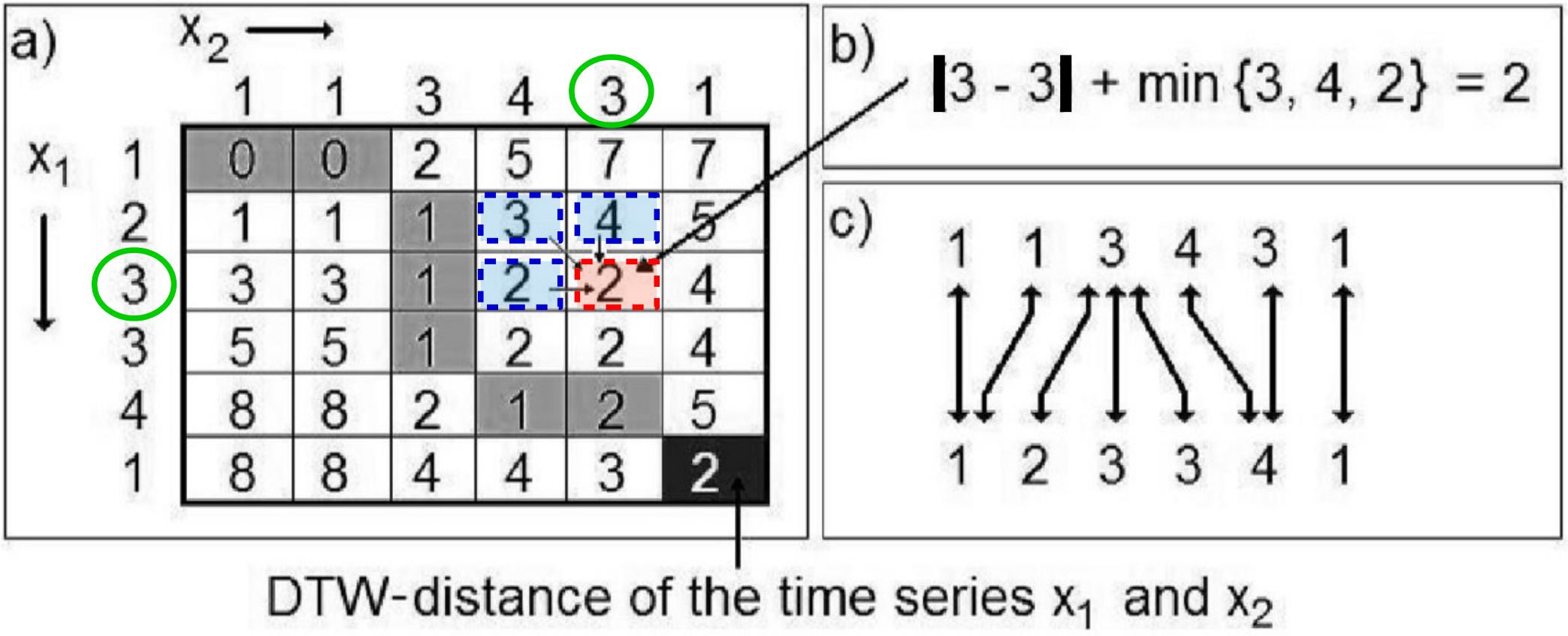
x2

Dynamic Time Warping

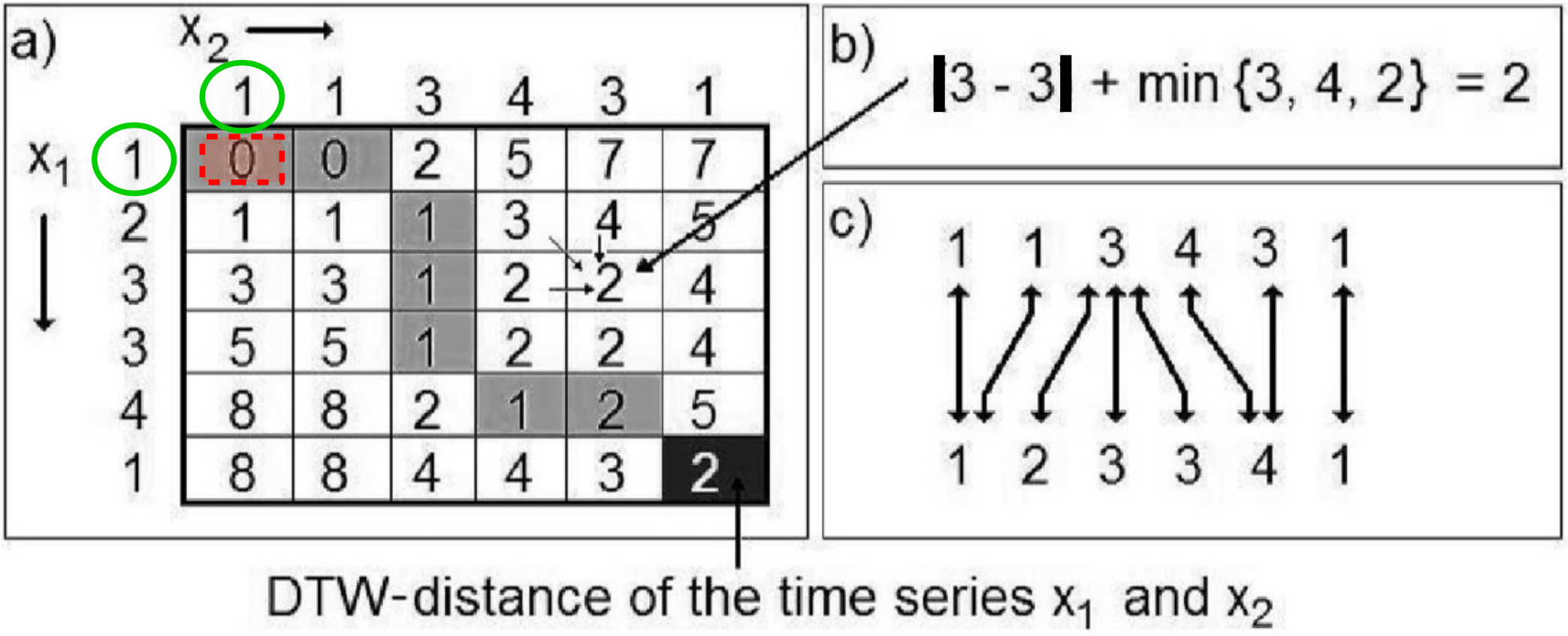




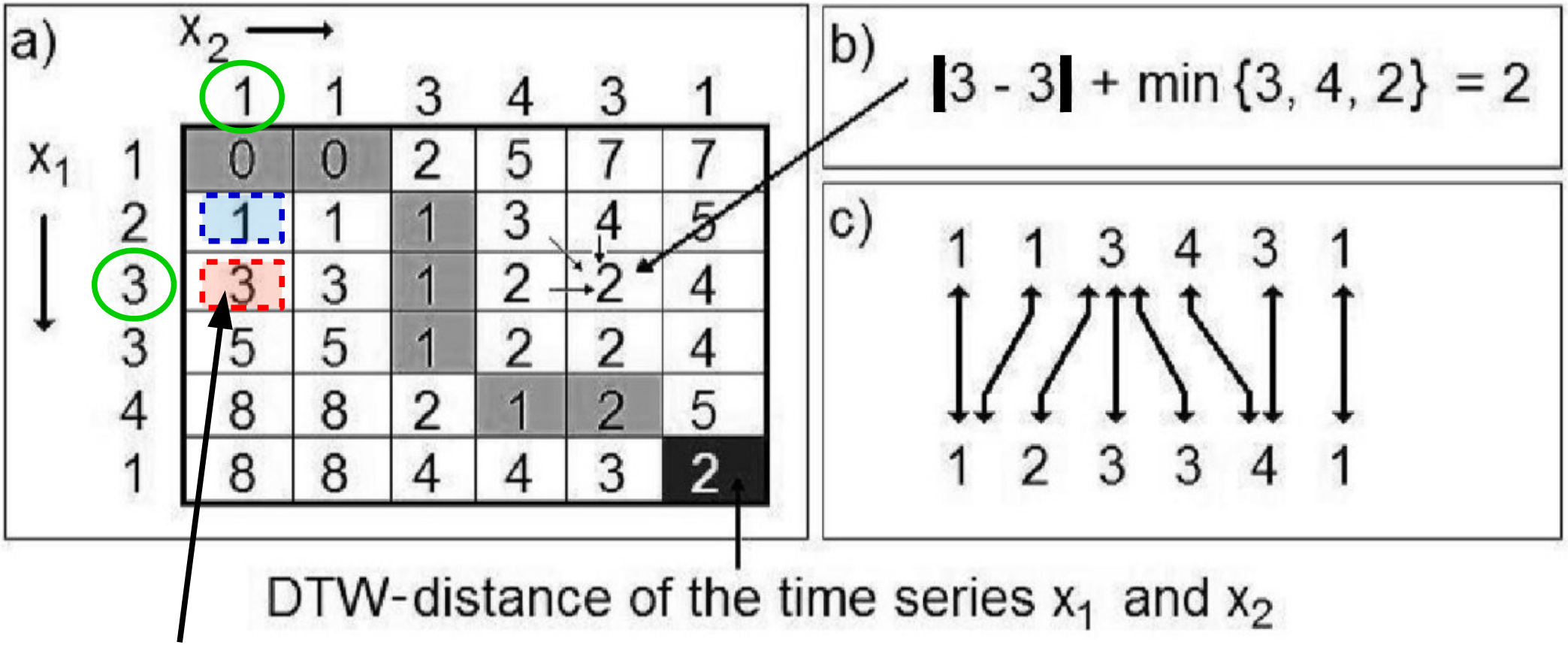
Dynamic Time Warping



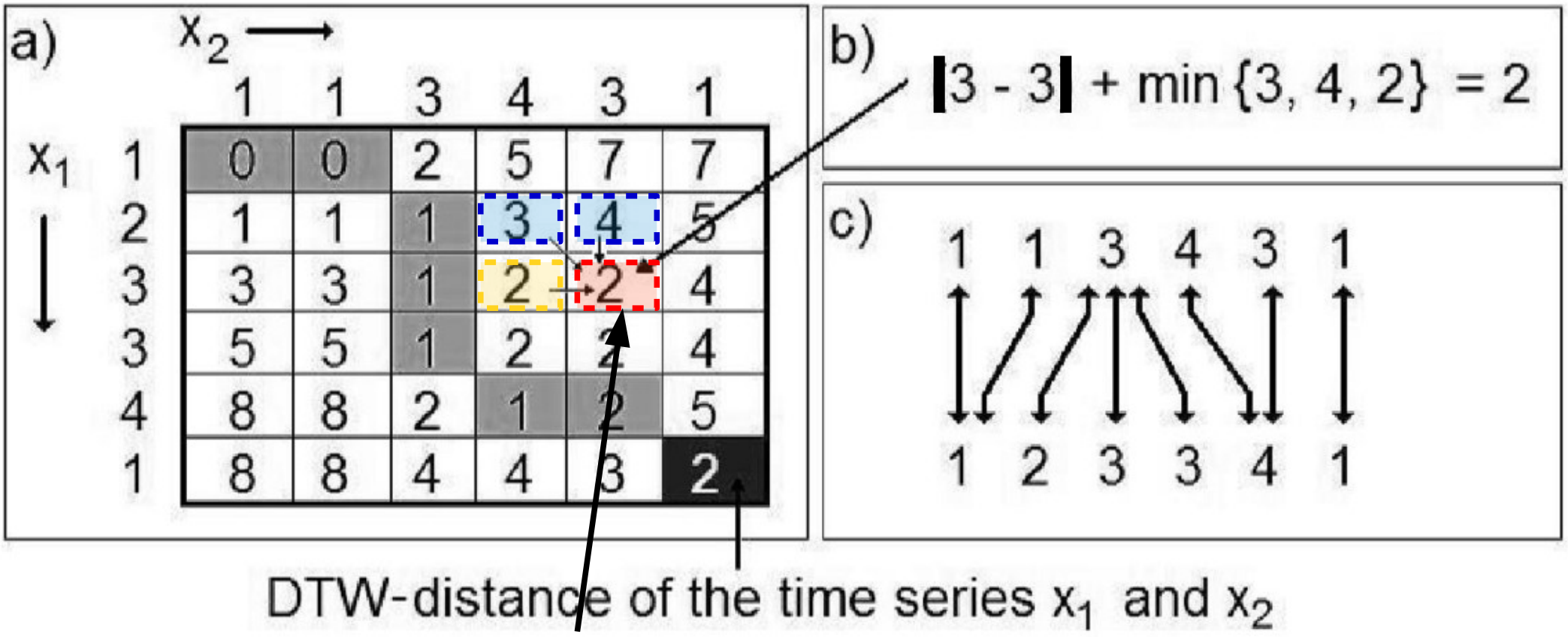
Dynamic Time Warping



Dynamic Time Warping



Dynamic Time Warping



For example, for this entry, we record that the yellow entry had the minimum out of the three entries that were considered while filling-in this entry.