Data Mining 2

Module 3 - 2020/2021

Name _____ Surname _____ ID: _____ Test id. AUTO

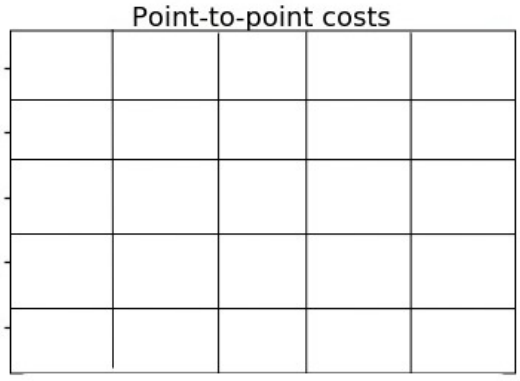
Q1. Given $w = 4$ and the time series $X =< 20, 15, 7, 7, 22, 29, 19 >$, apply the Moving Average Smoothing. What type of distortion the smoothing reduce?

A1. _____

Q2. Given the time series $X =< 2, 1, 7, 1, 2, 1, 2 >$, build the Matrix Profile with $m = 3$ using the Manhattan distance. Which is/are correct value/s for m that would have retrieved more motifs with distance equals to 0?

A2. _____

Q3. Given the time series $X_1 =< 4, 3, 2, 4, 1 >$ and $X_2 =< 2, 3, 4, 6, 3 >$, compute their distance using the DTW with distance between points computed as $d(x, y) = |x - y|$



A3. _____

Q4. It approximates a time series with the mean value of the points in a segment as well as the length of the segment itself:

1) PAA 2) APCA 3) DFT 4) DWT 5) PLA

A4. _____

Q5. What is change point detection?

1) The identification of seasonality in the underlying model of a time series

2) The identification of breakpoints in the underlying model of a time series

3) A specific method for time series clustering

4) The identification of motifs in the underlying model of a time series

5) A specific method for time series classification

A5. _____

Q6. What is a shapelet dataset?

1) A set of N time series represented as k-vectors of distances w/r/t representative approximations

2) A k-sample of N time series maximally representative of a class

3) A set of N time series represented as k-vectors of the most representative global features

4) A set of N times series represented as k-vectors of distances w/r/t representative elements of a class

5) A k-sample of N time series maximally representative of a set of patterns

A6. _____

Q7. A criterion for shapelet representativeness should be...

1) Using association rules algorithms

2) Extracting the most frequent motifs

3) Computing DTW

4) Computing Information Gain

5) Using clustering for finding representative centroids

A7. _____