

Knowledge Discovery in Databases II  
 SS 2019

Exercise 10: Time Series Data

Exercise 10-1 Time Series Preprocessing

Given a time series  $X = (3, 5, 10, 4, 1, 7, 7, 9, 1, 3)$

- (a) Compute the normalization (shifting and scaling) series  $\tilde{X}$  of  $X$ .
- (b) Compute the smoothing series  $\bar{X}$  of  $X$  with  $k = 1$ .

Exercise 10-2 Uniform and Dynamic Time Warping

Given the following two time series:  $X = (3, 5, 9, 2, 3, 6, 3)$  and  $Y = (3, 4, 6, 10, 1, 3, 2, 7, 4)$ , compute the following distances:

- (a) Uniform Time Warping Distance  $D_{UTW}^2$
- (b) Dynamic Time Warping  $DTW^2$
- (c) k-Dynamic Time Warping Distance  $D_{k-DTW}^2$  where  $k = 3$  (Optional)

Visualize the optimal alignment between the time series.

Exercise 10-3 Longest Common Subsequence for Time Series

Given the longest common subsequence distance for time series is defined as following:

Given two time series  $X = (x_1, \dots, x_n)$  and  $Y = (y_1, \dots, y_m)$  and two threshold  $\epsilon$  and  $\delta$ ,

$$LSC(X, Y) = \begin{cases} 0 & \text{if } n = 0 \vee m = 0 \\ LCS(start(X), start(Y)) + 1 & \text{if } match(last(X), last(Y)) \\ \max(LCS(start(X), Y), LCS(X, start(Y))) & \text{else} \end{cases}$$

where the matching function is defined as:

$$match(x_i, y_j) = \begin{cases} true & \text{if } |x_i - y_j| \leq \epsilon \wedge |i - j| \leq \delta \\ false & \text{else} \end{cases}$$

Then the distance is:

$$D_{LCS}(X, Y) = 1 - \frac{LCS(X, Y)}{\min(n, m)}$$

- (a) Given  $X = (3, 5, 9, 2, 3, 6, 3)$  and  $Y = (3, 4, 6, 10, 1, 3, 2, 7, 4)$ ,  $\epsilon = 2$ ,  $\delta = 2$ , what is the  $D_{LCS}(X, Y)$ ?

(b) Prove or disprove if the  $D_{LCS}$  is a metric distance.

**Exercise 10-4 PAA / DWT(Optional)**

Given two time series:  $X = (6, -2, -7, -1, 1, -3, 6, 8)$ ,  $Y = (1, 3, -8, -4, 5, -1, 2, 10)$ ,

(a) compute the  $L_1$ - and  $L_\infty$ -distance.

(b) Compute the dimension reduction representations of  $X$  using: DWT with Haar-Wavelet and PAA with  $M = 4$  boxes.