Exercise sheet #4: Mining Temporal Data
Due 17.11.22 at 08:00 (Exercise session on 17.11.22 at 10:00, self-marking by 17.11.22 at 23:59)

Please carefully read and follow the general instructions regarding exercises. Failing to meet the requirements might lead to penalties. https://moodle.uef.fi/mod/page/view.php?id=1935632
If you suspect that something is wrong with some exercise question, please contact the lecturer.
If you face persistent issues while working on an exercise, do ask for help, e.g. during a course meeting or by contacting the lecturer via email.

Problem 1 (DTW distance for time-series). Compute the dynamic time warping distance $D_{DTW}(S_A, S_B)$ between the time-series below, a) first without any window constraint, then b) with a window of width at most one.

Problem 2 (Aligning sequences). Consider the pair of discrete sequences $S_A = abacc$ and $S_B = bdabcd$.

a) Compute the dynamic time warping distance without window constraint.
b) Compute the edit distance with costs $c_{del} = 0$, $c_{ins} = 2$, $c_{sub} = 3$.
c) Compute the longest common subsequence length.

Problem 3 (Frequent sequence mining). Consider the following sequence


a) Compute the support of the four short subsequences below, allowing gaps of at most 1 (i.e. consecutive itemsets from the subsequence are allowed to map to itemsets in the data sequence separated by at most one unmapped itemset).


b) Enumerate the subsequences of $S$ having size at most 4 and 5 or more occurrences, not allowing gaps (i.e. consecutive itemsets from the subsequence must map to consecutive itemsets in the data sequence).