Algorithmic Data Analysis (ADA)
Spring 2024
Exercise sheet #5: Mining Temporal Data
Due 12.02.24 at 10:00 (Exercise session at 12:00, self-marking by the end of the day)

UNIVERSITY OF EASTERN FINLAND

Please carefully read and follow the general instructions regarding exercises. Failing to meet the requirements might lead to penalties. https://elearn.uef.fi/mod/page/view.php?id=248672

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 (Binning, smoothing, etc.). Consider the following time-series

$$\mathcal{S}_X = \langle 2, 6, 6, 3, 3, 3, 5, 5, 3, 4, 6, 3, 2, 3, 6, 4 \rangle .$$

a) Determine the binned time-series for bins of length 4.

b) Construct the moving average series for a window of width 4.

c) Construct the exponentially smoothed series with smoothing parameter $\alpha = 0.6$, setting the initial value of the smoothed series to the value of the first point of the original series.

Problem 2 (Discrete Fourier transform, *Aggarwal C14 Q7*). For a real-valued time series $x_0, \ldots, x_{n-1}$ with Fourier coefficients $f_0, \ldots, f_{n-1}$, show that $f_k + f_{n-k}$ is real-valued for each $k = 1 \ldots n-1$.

Problem 3 (Discrete wavelet transform).

a) Compute the weights of the discrete wavelet transform for time-series $\mathcal{S}_X$ above. Give the vector of weights and the corresponding wavelet matrix, i.e. the matrix containing unnormalized basis vectors as its rows.

b) Compute the reconstructed time-series after dropping $3/4$ of the coefficients while minimizing the reconstruction error. What is the ratio of energy retained in the approximated time-series?

Problem 4 (A chicken and egg problem). Consider a chicken K. Every day, the chicken might lay one egg (`Y`) or none (`N`). Whether or not the chicken lays an egg on a given day depends on its mood on that day. The chicken is rather inexpressive, so that its mood is not outwardly discernible, but it is known not to have a volatile character, so that its mood does not change during the course of the day. Furthermore, it is known to be either angry (`A`), broody (`B`) or cheerful (`C`). If the chicken is broody one day, four times out of six it will continue to brood the next day, else it will be angry or cheerful instead, with equal chances. Similarly, if the chicken is angry one day, three times out of five it will remain angry the next day, else it will instead brood or cheer, with equal chances. On the other hand, if it was cheerful the previous day, the chicken will wake up in any state of mind with equal probability. The chicken lays its daily egg with probability 0.8 when angry, 0.3 when broody, and 0.9 when cheerful. At the beginning of an observation period, it is safe to assume the chicken to be angry with probability 0.6 and broody with probability 0.2.

a) Model this scenario using a Markov model. Draw a schematic representation of the model and indicate its parameters.

Last week, the chicken laid eggs on Monday, Tuesday and Thursday but not on Wednesday nor on Friday.

b) Under the above model, what is the probability for this egg laying sequence to happen and the chicken being cheerful on Monday and broody on other days? How about with the chicken being cheerful on Monday, Tuesday and Thursday but broody on Wednesday and Friday?

c) What is the probability for this egg laying sequence to be observed under the above model?

d) What is the most likely underlying sequence of daily moods of the chicken?