Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

Intelligent Systems

Exercise 7- Segmentation

Simon Reichhuber

December 12, 2019

University of Kiel, Winter Term 2019

Table of Content

Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

Top down segmentation

1. A

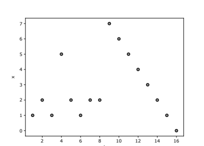
Given two time series in Figure 1, apply a top down segmentation with a maximum approximation error of 1 by using the error function:

$$\sum_{i=1}^{n} |x_{t_i} - \tilde{x}(t_i)|,$$

where $S = \{t_i\}_{i=1}^{n}$ is a segment of length n, x_t are the measurements at time t, and $\tilde{x} : \mathbb{R} \to \mathbb{R}$ is the approximation. For the approximation function $\tilde{x}$ use:

A. A constant function.

B. A polynom of degree 1.



Offline techniques

Top-down approach

- Given: Time series, modelling or approximation method (often approximation with polynomial of degree 1) and (abort) criterion (often threshold value for approximation error)
- Wanted: Segmentation of the time series into as few sections as possible so that the specified criterion is met for all segments (e.g. approximation error smaller than specified threshold value).
- Basic idea: starting from only one segment (entire time series), each segment is further subdivided until the termination criterion for all sub-segments is fulfilled.

Offline techniques (2)

Christian-Albrechts-Universität zu Kiel

Process (using the example of an approximation and consideration of the error as a termination criterion):

- Calculate the approximation error for the currently considered segment (initially: entire time series).
- For each possible position at which the segment can be segmented into two parts, calculate the sum of the approximation errors of the two sub-segments.
- Split the segment at the point where the greatest reduction in error is achieved.
- If the error of one of the two partial segments is greater than the given threshold value, subdivide this further

Approximate n points with the help of a constant function means:

$$\tilde{x}(t) = c,$$

and choose $c \in \mathbb{R}$ that the minimum approximation error is minimised.

$$c^* = \min_{c \in \mathbb{R}} \sum_i |c - x_i|$$

(Remark: Simplified notation: $x_i := x_{t_i}$)

Through differentiation of above equation and setting it to 0:

$$c^* = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x_i}$$

1. B

A. A constant function.

B. A polynom of degree 1.

Approximate n points with the help of a polynom of degree 1 means:

$$\tilde{x}(t) = at + b,$$

and choose $a, b \in \mathbb{R}$ that the minimum approximation error is minimised.

$$(a^*, b^*) = \min_{a,b \in \mathbb{R}} \sum_i |at_i + b - y_i|$$

Through differentiation of above equation* and setting it to 0:

$$a^* = \frac{\sum_{i=1}^n t_i x_i - n\bar{t}\bar{x}}{\sum_{i=1}^n t_i^2 - n\bar{t}^2}$$

$$b^* = \bar{x} - a\bar{t}$$

$$g = \sum_{i=1}^{n} (at_i + b - x_i)^2 = \sum_{i=1}^{n} a^2 t_i^2 + 2abt_i - 2at_i x_i - 2bx_i + b^2 + x_i^2$$

$$\frac{\partial g(a,b)}{\partial b} = \sum_{i=1}^{n} 2at_i - 2x_i + 2b = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} at_i - \sum_{i=1}^{n} x_i + nb = 0$$

$$\Leftrightarrow b = -a\frac{\sum_{i=1}^{n} t_i}{n} + \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\Leftrightarrow b = \overline{x} - a\bar{t} \quad (*)$$

1. B A Top down segmentation with polynom of degree 1 - Derivation II

$$\frac{\partial g(a,b)}{\partial a} = \sum_{i=1}^{n} 2at_i^2 + 2bt_i - 2t_ix_i = 2\sum_{i=1}^{n} t_i(at_i + b - x_i) = 0$$

Important: Here (*) substitute b, because b depents on a (b = b(a)) and has to be taken into account for the partial derivative!

$$\Leftrightarrow 2\sum_{i=1}^{n} t_i(at_i + (\overline{x} - a\overline{t}) - x_i) = 0$$

1. B A Top down segmentation with polynom of degree 1 - Derivation III

CAU
Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

$$\Leftrightarrow \sum_{i=1}^{n} a t_i^2 - a\bar{t} t_i + t_i \bar{x} - t_i x_i = 0$$

$$\Leftrightarrow a \sum_{i=1}^{n} t_i^2 - \bar{t} t_i = \sum_{i=1}^{n} t_i x_i - \sum_{i=1}^{n} t_i \bar{x}$$

$$a = \frac{\sum_{i=1}^{n} t_i x_i - \sum_{i=1}^{n} t_i \bar{x}}{\sum_{i=1}^{n} t_i^2 - t_i \bar{t}} = \frac{\sum_{i=1}^{n} t_i x_i - n\bar{t}\bar{x}}{\sum_{i=1}^{n} t_i^2 - n\bar{t}^2}$$

Bottom up segmentation

2. A

A. a constant function.

B. a polynom of degree 1.

Offline techniques (5)

Christian-Albrechts-Universität zu Kiel

Bottom-up approach:

- Requirements as for top-down procedures: Time series, error function and threshold value
- A segmentation of the time series into as few segments as possible is also required so that the threshold value is not exceeded for all segments.
- Basic idea: Starting from a segmentation as fine as possible, segments are grouped together step by step until it is no longer possible to group them together without violating the targets.

Offline techniques (6)

Process (using the example of an approximation and consideration of the error as a termination criterion):

- Segment the entire time series into as many sub-segments as possible (e.g. for an approximation of each segment with a polynomial of degree K in segments of length K + 1).
- Calculate the increase of the approximation error for the combination of two adjacent segments.
- Combine adjacent segments with the minimum increase of the error as long as the error of the resulting segment does not exceed the threshold.

2. A/B

See *07_Segmentation.ipynb*.

Signature Task

3. A

Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

A. Discuss about the optimal segmentation technique for the Signature Task.
B. Apply the chosen segmentation technique.

Python: Segmentation

4. A

A. Download the jupyter notebook *4_SAX.ipynb* from Open Olat. First, calcluate the Euclidean distance of the two time series. Afterwards, apply the steps of the *SAX* algorithm and compare the distance of the two strings. What attracts your attention? Which paramaters can be adapted two achieve better results?