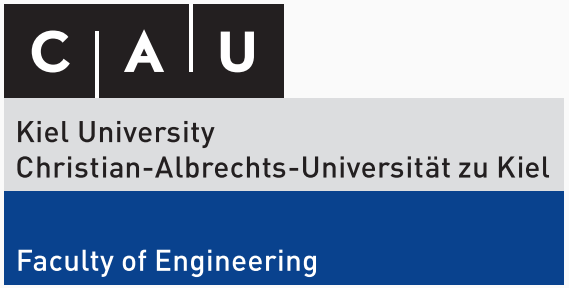


Intelligent Systems

Excersice 4- Representation

Simon Reichhuber

November 18, 2019

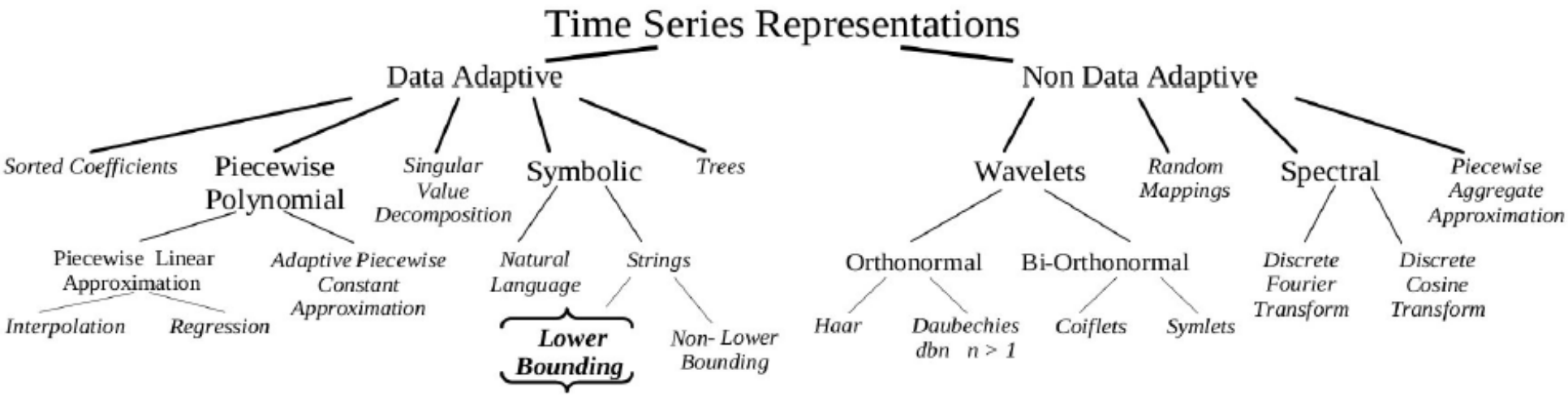
University of Kiel, Winter Term 2019

Table of Content

Data representation

Non data adaptive vs. data adaptive Approximation

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



C|A|U
Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

The FT represents the time series in the frequency domain. The signal is constructed as a sequence of sine and cosine terms.

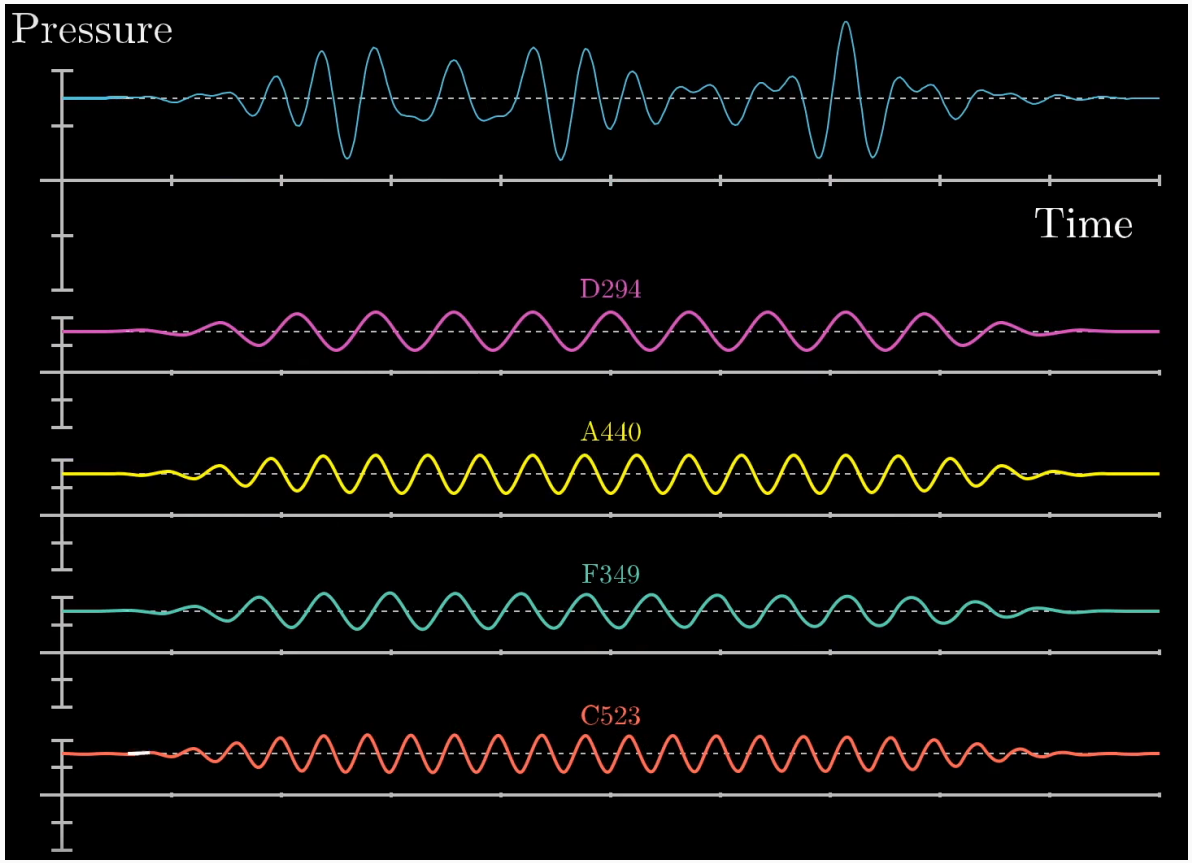


Figure 1: Fourier Transform

C|A|U
Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

The FT represents the time series in the frequency domain. The signal is constructed as a sequence of sine and cosine terms.

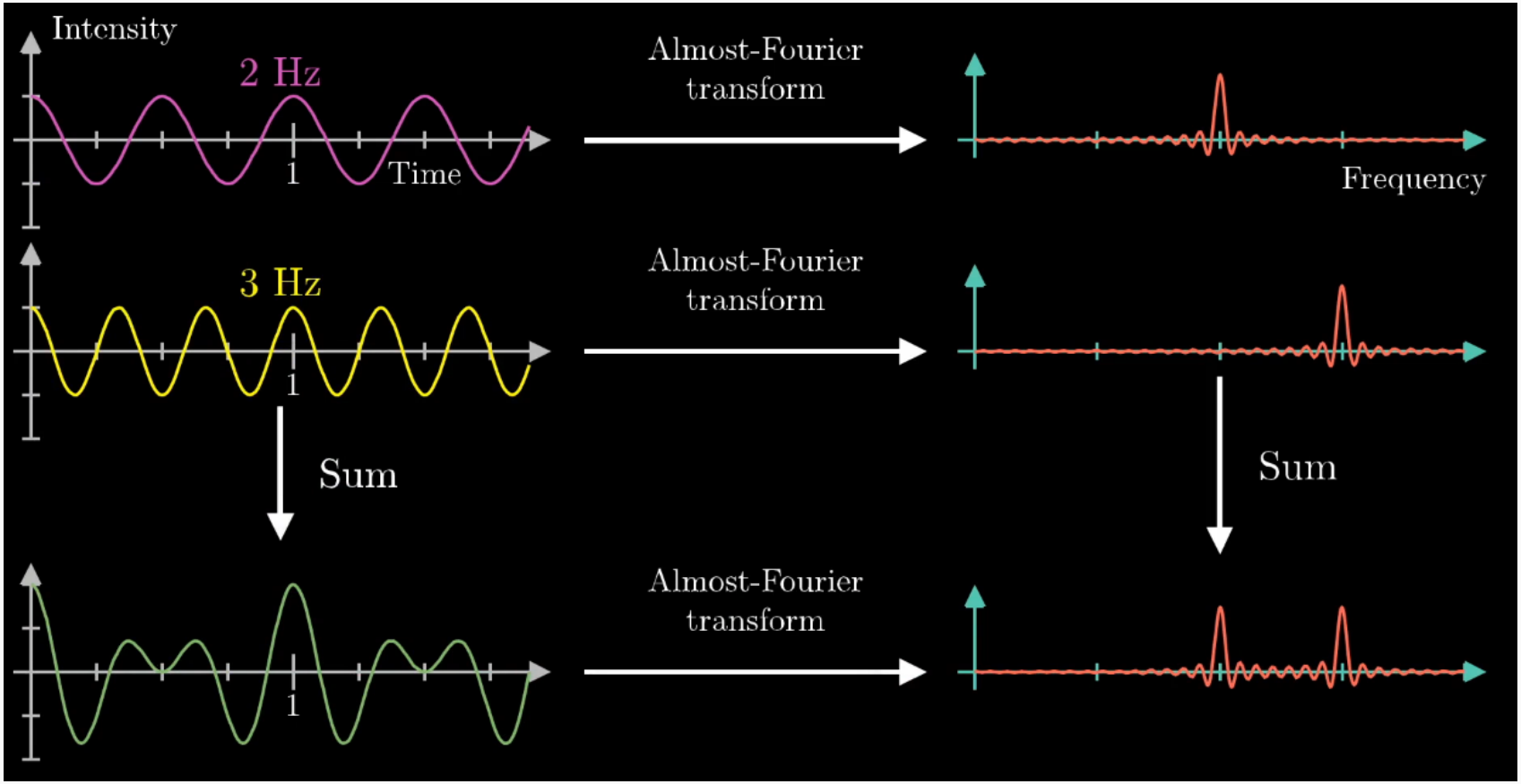


Figure 2: Fourier Transform

C|A|U
Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

A. **Explain the idea of the *Shape Definition Language* and its application?**
B. Approximate the time series with the following approximations:
   - *Piecewise Aggregate Approximation* (*PAA*) with 4 segemtns.
   - *Clipping* to binary values ($\rightarrow$ search the procedure on the internet).
   - *Picewise Linear Approximation* with 4 segments.
   - *Run-Length Encoding* (*RLE*).
C. Aggregate the timeseries to the following statistical measures:
   - *Mean*
   - *Standard deviation*
   - *Mode*
D. What are the advantages and disatvantages of the *clipping* procedure?
E. What is the main difference between the *Adaptive Picewise Aggregate Approximation* (*APAA*) and the *PAA*?

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

Given the case that the most important information of a time series can be extracted from the rough shape, limited terms of the *Shape Definition Language* (*SDL*) are enough to model it, i.e. *Up, up, stable, zero, down, Down*. These represent the different slopes of the time series. Advantageously, the representation can easily be processed by algorithms for symbolic sequences, like *Longest Common Subsequence* (*LCSS*).

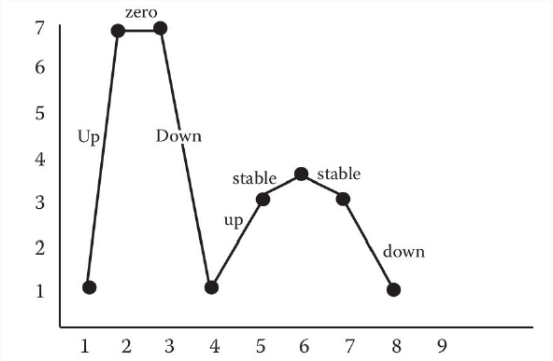


Figure 3: Shape Definition Language[1]

[1] Mitsa, Theophano. Temporal data mining. CRC Press, 2010

C|A|U
Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

A. Explain the idea of the *Shape Definition Language* and its application?

B. Approximate the time series with the following approximations:

- *Piecewise Aggregate Approximation* (*PAA*) with 4 segemtns.
- *Clipping* to binary values ($\rightarrow$ search the procedure on the internet).
- *Picewise Linear Approximation* with 4 segments.
- *Run-Length Encoding* (*RLE*).

C. Aggregate the timeseries to the following statistical measures:

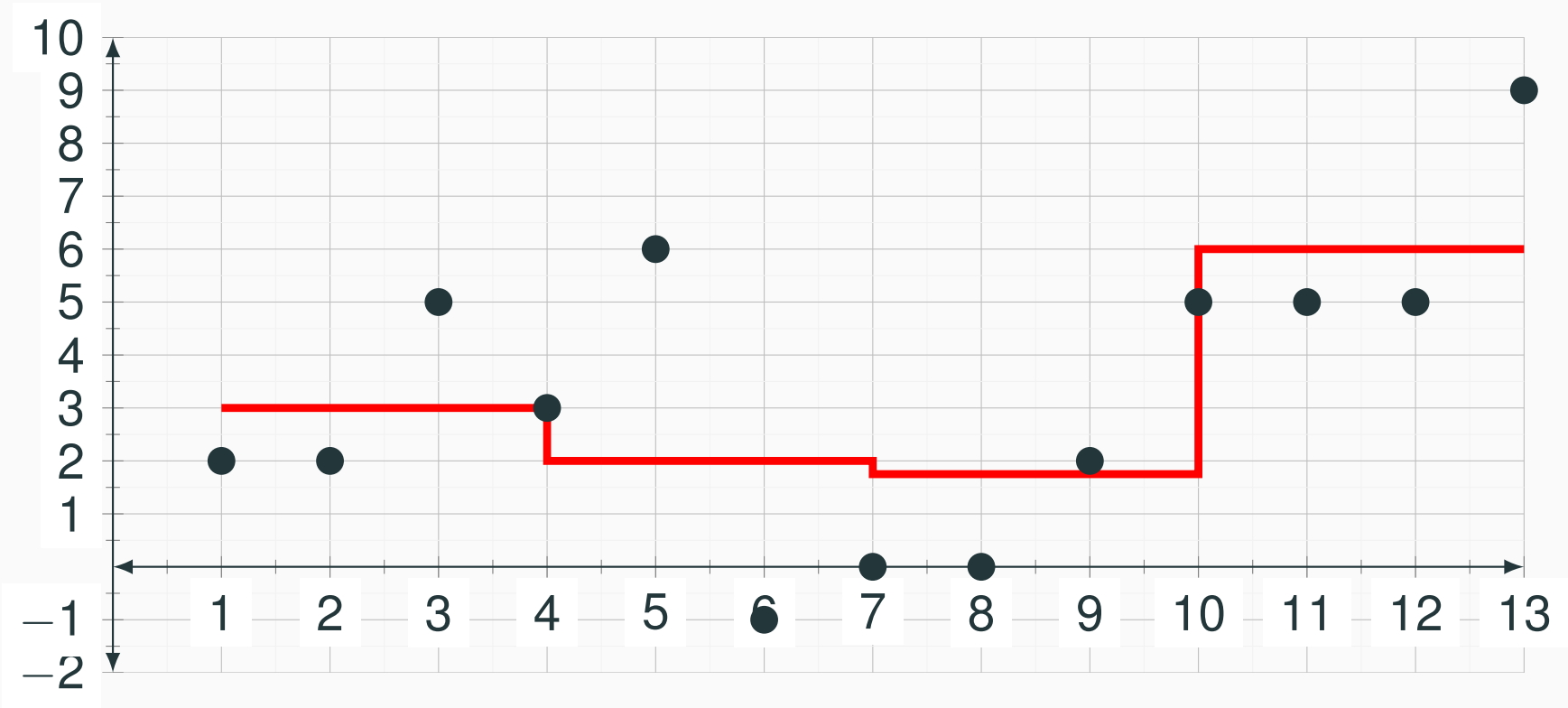
- *Mean*
- *Standard deviation*
- *Mode*

D. What are the advantages and disatvantages of the *clipping* procedure?

E. What is the main difference between the *Adaptive Picewise Aggregate Approximation* (*APAA*) and the *PAA*?

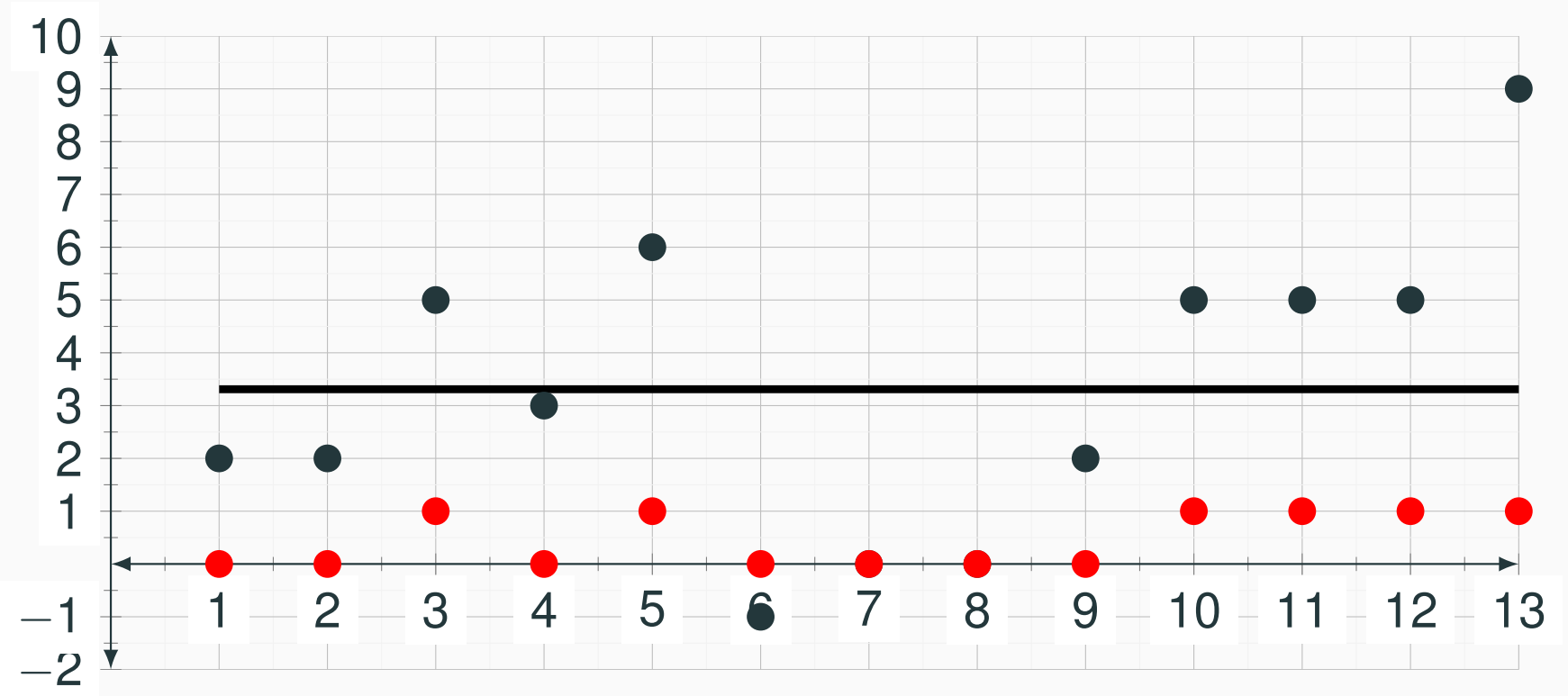
Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

- PAA Segment Length $= \frac{t_{end} - t_{start}}{\#Segments}$
- 4 segments $\rightarrow$ PAA Segment Length $= \frac{13-1}{4} = 3$
- Segment 1: $\frac{2+2+5+3}{4} = 3$
- Segment 2: $\frac{3+6+(-1)+0}{4} = 2$
- ...

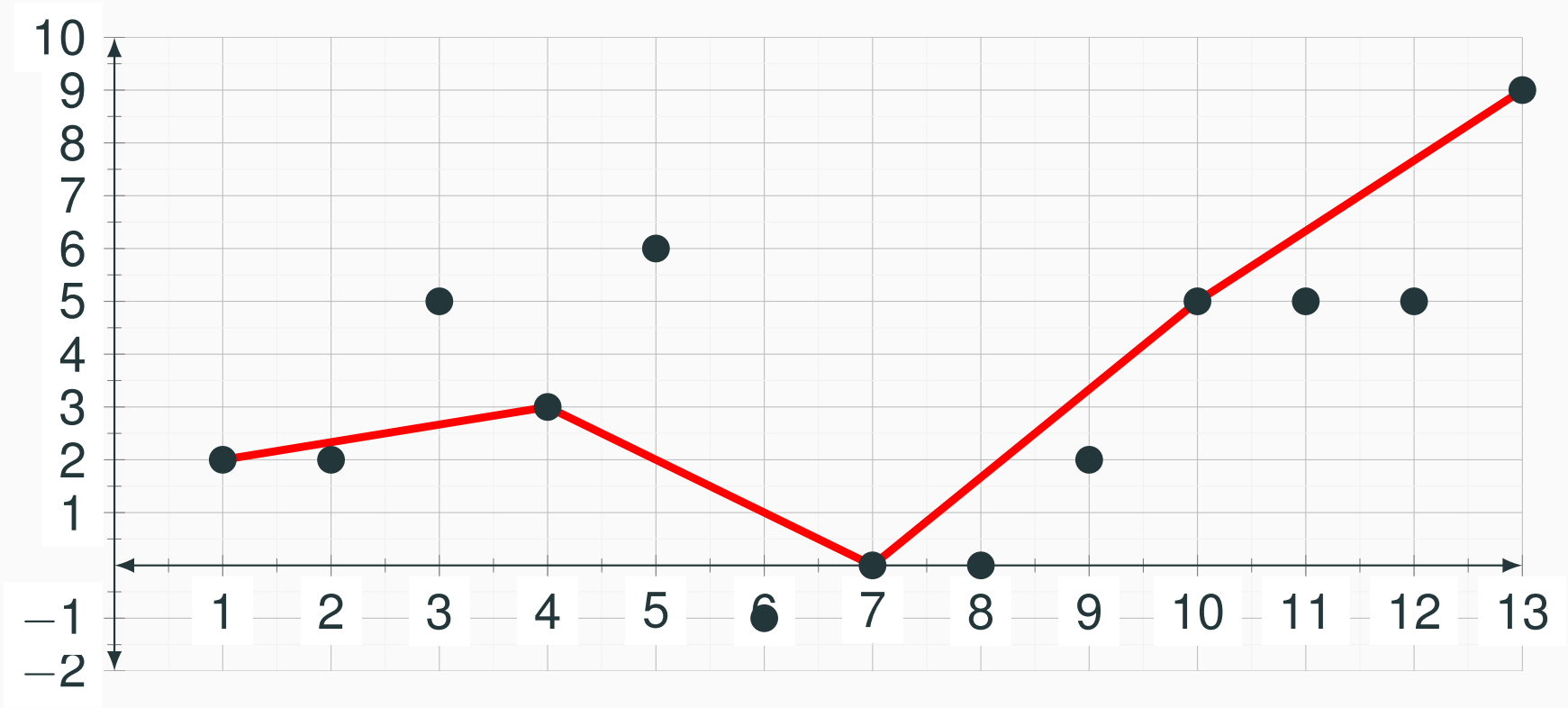


Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

- Calculate time series' mean: $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$
- The clipped values are given by: $y_i^* = \begin{cases} 1 \text{ , for } x_i \geq \mu \\ 0 \text{ , otherwise} \end{cases}$
- $\mu = \frac{2+2+5+3+6+(-1)+0+0+2+5+5+5+9}{13} \approx 3.31$
- Clipping: $0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1$



1. B Piecewise Linear Approximation



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

- Counter $\{n\}$ represents the number of repititions of the following symbol
- $2, 2, 5, 3, 6, -1, 0, 0, 2, 5, 5, 5, 9 \rightarrow$
  $\{2\}2, 5, 3, 6, -1, \{2\}0, 2, \{3\}5, 9$

A. Explain the idea of the *Shape Definition Language* and its application?
B. Approximate the time series with the following approximations:
   - *Piecewise Aggregate Approximation* (*PAA*) with 4 segemtns.
   - *Clipping* to binary values ($\rightarrow$ search the procedure on the internet).
   - *Picewise Linear Approximation* with 4 segments.
   - *Run-Length Encoding* (*RLE*).

C. Aggregate the timeseries to the following statistical measures:
   - *Mean*
   - *Standard deviation*
   - *Mode*

D. What are the advantages and disatvantages of the *clipping* procedure?
E. What is the main difference between the *Adaptive Picewise Aggregate Approximation* (*APAA*) and the *PAA*?

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

Given N D-dimensional points $x_i \in \mathbb{R}^d$, $i = 1, \ldots, N$, the following statistical measures can be defined:

- *Mean:* $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \quad \rightarrow 3.31$
- *Standard deviation*: $\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu)^2} \rightarrow 2.84$
- *Mode*: $m = argmax_x |\{y | y \in \{x_1, \ldots, x_N\}, x = y\}| \rightarrow 5$

A. Explain the idea of the *Shape Definition Language* and its application?

B. Approximate the time series with the following approximations:
   - *Piecewise Aggregate Approximation* (*PAA*) with 4 segemtns.
   - *Clipping* to binary values ($\rightarrow$ search the procedure on the internet).
   - *Picewise Linear Approximation* with 4 segments.
   - *Run-Length Encoding* (*RLE*).

C. Aggregate the timeseries to the following statistical measures:
   - *Mean*
   - *Standard deviation*
   - *Mode*

D. What are the advantages and disatvantages of the *clipping* procedure?

E. What is the main difference between the *Adaptive Picewise Aggregate Approximation* (*APAA*) and the *PAA*?

What are the advantages and disatvantages of the *clipping* procedure?

Advantages:

- Extreme compression rate (*float* $\rightarrow$ *bool*)
- Simple representation (only binary values)
- Rough patterns can be found easily

Disatvantage:

- Very unprecise representation of the signal

A. Explain the idea of the *Shape Definition Language* and its application?

B. Approximate the time series with the following approximations:
   - *Piecewise Aggregate Approximation* (*PAA*) with 4 segemtns.
   - *Clipping* to binary values ($\rightarrow$ search the procedure on the internet).
   - *Picewise Linear Approximation* with 4 segments.
   - *Run-Length Encoding* (*RLE*).

C. Aggregate the timeseries to the following statistical measures:
   - *Mean*
   - *Standard deviation*
   - *Mode*

D. What are the advantages and disatvantages of the *clipping* procedure?

E. What is the main difference between the *Adaptive Picewise Aggregate Approximation* (*APAA*) and the *PAA*?

What is the main difference between the *Adaptive Picewise Aggregate Approximation* (*APAA*) and the *PAA*?

Main difference:

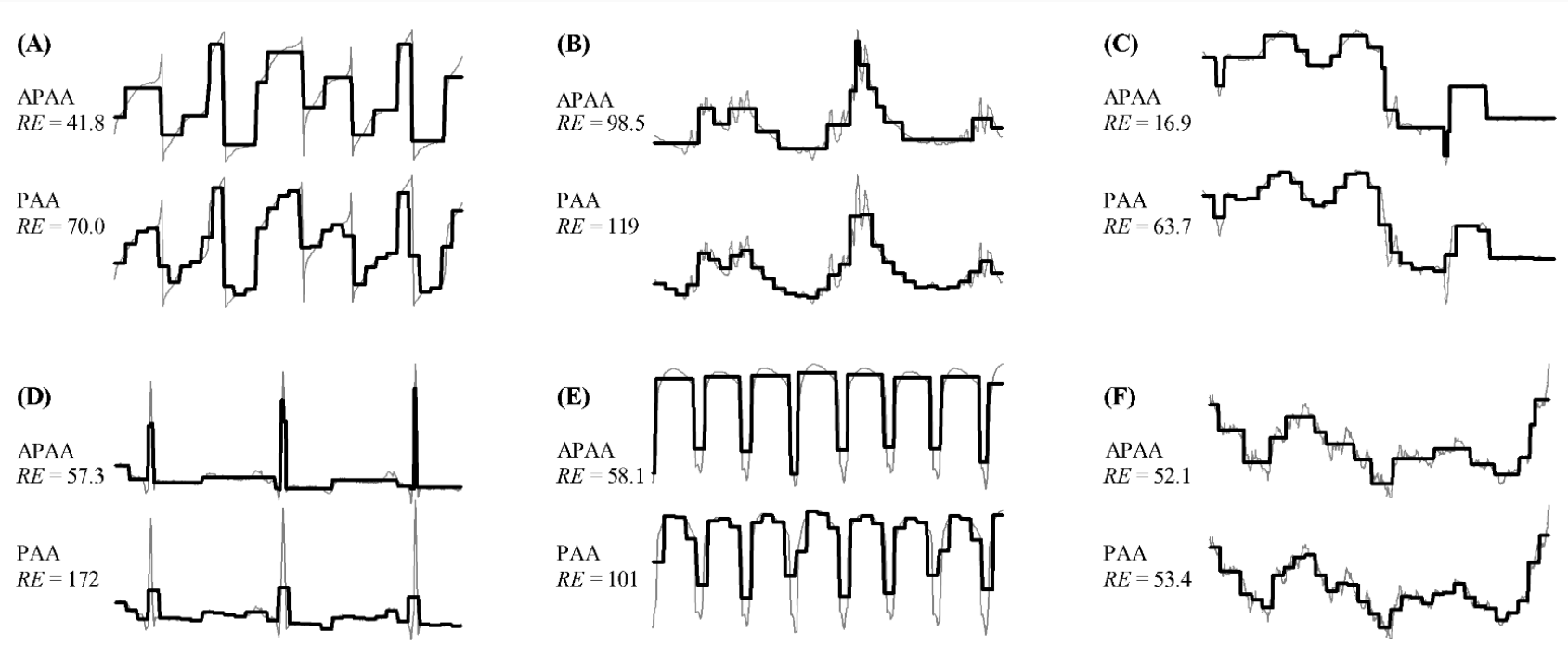
- Variable length of the (temporal) sections
- Adaptive according to local details of a time series
  - Sections of frequent movement will be parted in smaller intervals
  - A Section without significant information will be represented as a large interval

1. E AAPA vs. PAA II

What is the main difference between the *Adaptive Picewise Aggregate Approximation* (*APAA*) and the *PAA*?

Advantage:

- The error (e.g. Least-Squares) between the raw data and the representation is reduced.[2]



[2]Koegh et. al. 2001

Principal component analysis

A. **What is the goal of the *Principal Component Analysis* (*PCA*) and what is its basic assumption.**

B. **What is the benefit of the *PCA*?**

C. Describe the following items:
   - Zero-mean feature
   - Variance
   - Standard deviation
   - Covariance matrix
   - Arithmetic mean
   - Eigenvector
   - Eigenvalue
   - Projection onto new feature space

D. How can we get a dimensionality reduction with the means of Eigenvalues?

What is the goal of the *Principal Component Analysis* (*PCA*) and what is its basic assumption. **Basic assumption:**

- The larger the variance, the higher is the level of information.

Goal:

- Dimensionality reduction:
  - Remove dimensions with poor information gain
  - Generate new dimensions that better fit to the structure of the data, i.e. at which the data has the largest variance.

Benefits:

- **Reduce computing time**: Through the usage of data mining algorithms applied on dimensionality-reduced data sets.
- **Feature selection**: Very easy; just select the PCA meta features.
- **Comprehensability**: Easy detection of structures, e.g. by projecting the data onto the two or three most important meta features.

A. What is the goal of the *Principal Component Analysis* (*PCA*) and what is its basic assumption.

B. What is the benefit of the *PCA*?

C. **Describe the following items:**
   - Zero-mean feature
   - Variance
   - Standard deviation
   - Covariance matrix
   - Arithmetic mean
   - Eigenvector
   - Eigenvalue
   - Projection onto new feature space

D. How can we get a dimensionality reduction with the means of Eigenvalues?

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

Given: Data set with N samples and D dimensions (features).

1. Standardisation:

- Calculate the arithmetic mean for each feature j:
  $\mu_j = \frac{1}{N} \sum_{i=1}^{N} x_{j,i}$
- Zero-mean sample:
  $x'_{j,i} = x_{j,i} - \mu_j$
- Calculate the standard deviation for each feature j:
  $\sigma_j = \sigma_j = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} x'^2_{j,i}}$
- Standardisation for every feature j:
  $x_{j,i} = \frac{x'_{j,i}}{\sigma_j}$

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

2. Calculate covariance matrix:

- $s_{j,j'} = \frac{1}{N-1} \sum_{i=1}^{N} x'_{j,i} \cdot x'_{j',i}$
- $S = \begin{pmatrix} s_{1,1} & \cdots & s_{1,D} \\ \vdots & \ddots & \vdots \\ s_{D,1} & \cdots & s_{D,D} \end{pmatrix}$

3. Calculate eigenvectors and eigenvalues:

- Eigenvalues represent the ratio of the variance along their corresponding eigenvectors
- The eigenvector with the largest eigenvalue represents the direction, in which the data has the largest variance
- The eigenvector with the second largest eigenvalue represents a orthogonal direction w.r.t. the first eigenvector, in which the data has the second largest variance, etc.
- Eigenvectors are the *principal components* (*PC*), which are more suitable to model the structure of the data than the original features (assuming that high information gain corresonds to high variance).

4. Dimensionality reduction: Choose the most important eigenvalues according to their eigenvalues:

- Method 1: The sum of the remaining eigenvalues shall be larger than a predefined ratio (e.g. 0.75) of the sum of all eigenvalues.
- Method 2: Dimensions shall be removed if the eigenvalue of the corresponding eigenvector is lower than the mean of all eigenvalues.

4. Projection onto the new feature space:

- Spanned by the new selected eigenvectors (*PC*s).

Python PCA

A. Download the file *04_Representation.ipyn* from *OpenOlat*.
B. In order to solve the tasks, you can use the library *numpy*.
C. Compare your results afterwards with the help of *sklearn*.