UNI KASSEL VERSITÄT

Chair for Communication Technology

Prof. Dr.-Ing. Klaus David

ComTec

Segmentation and Data Features

M. Sc. Judith S. Heinisch

ComTec, Kassel, 04.05.2019

© ComTec 2019

Topics for this lecture

ComTec

Sensor Data

Segmentation → Feature Extraction → Model creation → Model evaluation

Segmentation & Features (JH)

Bayesian Classification (MM)
Decision Tree Classifier (CA)
Clustering/ KNN (MB)
Hidden Markov Models (IK)

© ComTec 2019

Topics for this lecture

ComTec

Sensor Data

Segmentation → Feature Extraction → Model creation → Model evaluation

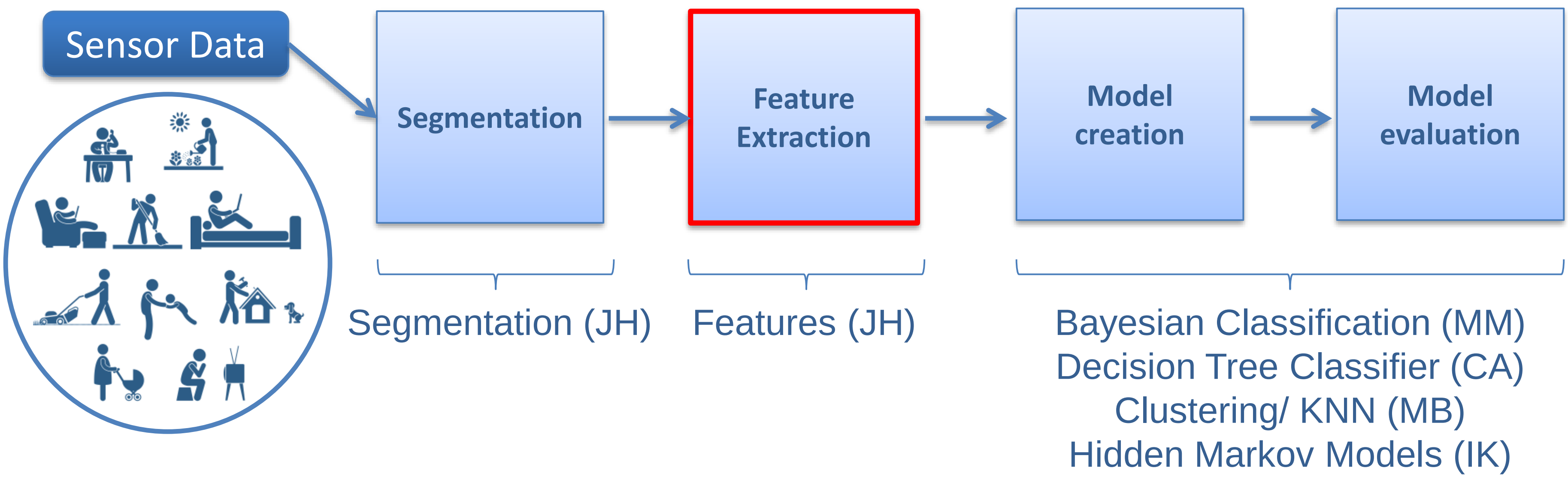
Segmentation (JH)

Features (JH)

Bayesian Classification (MM)
Decision Tree Classifier (CA)
Clustering/ KNN (MB)
Hidden Markov Models (IK)

© ComTec 2019

Topics for this lecture



Segmentation Algorithms

Time Series

Time series segmentation

Sliding Window Algorithm

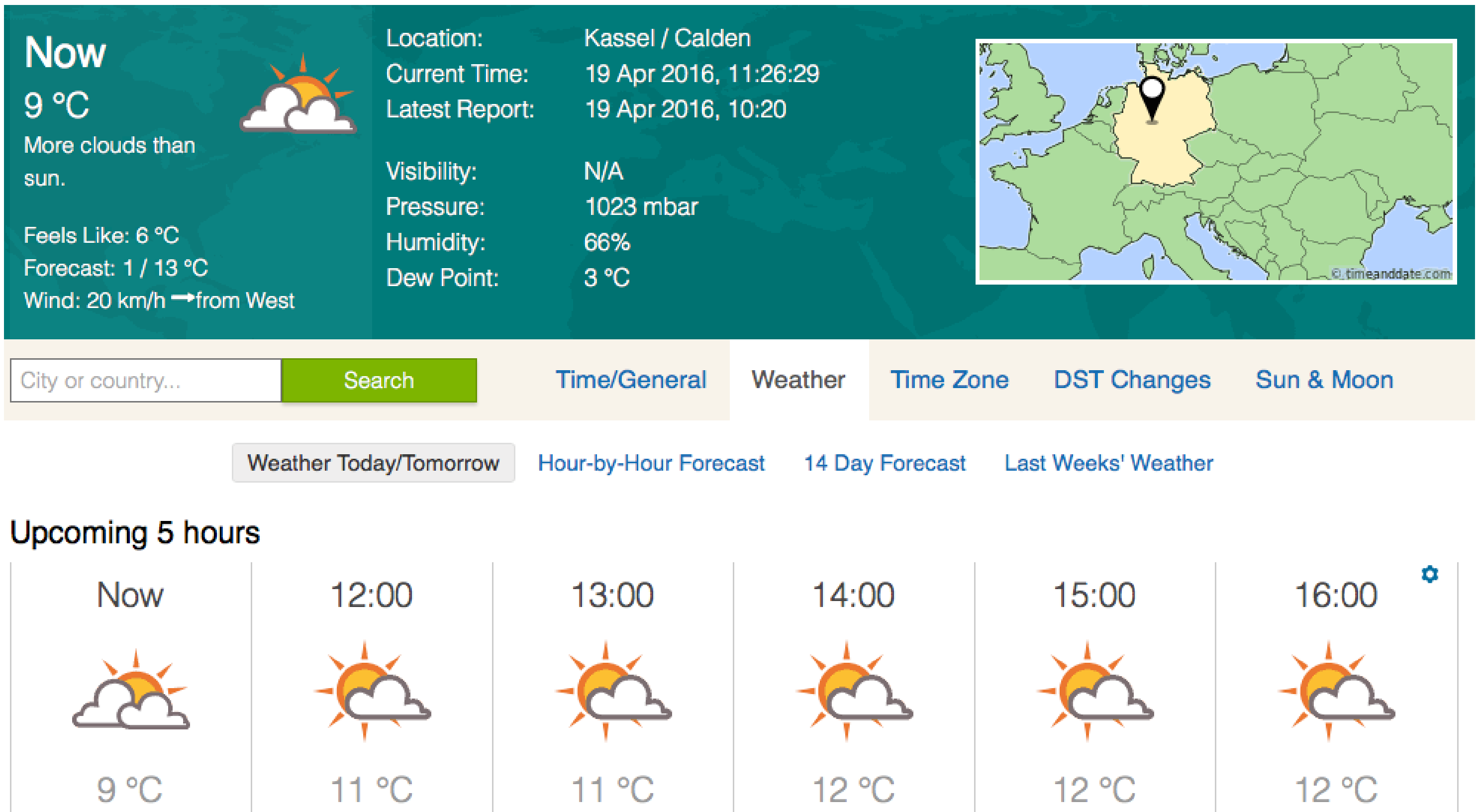
© ComTec 2019

What are time series?

- Some examples for time series(?)

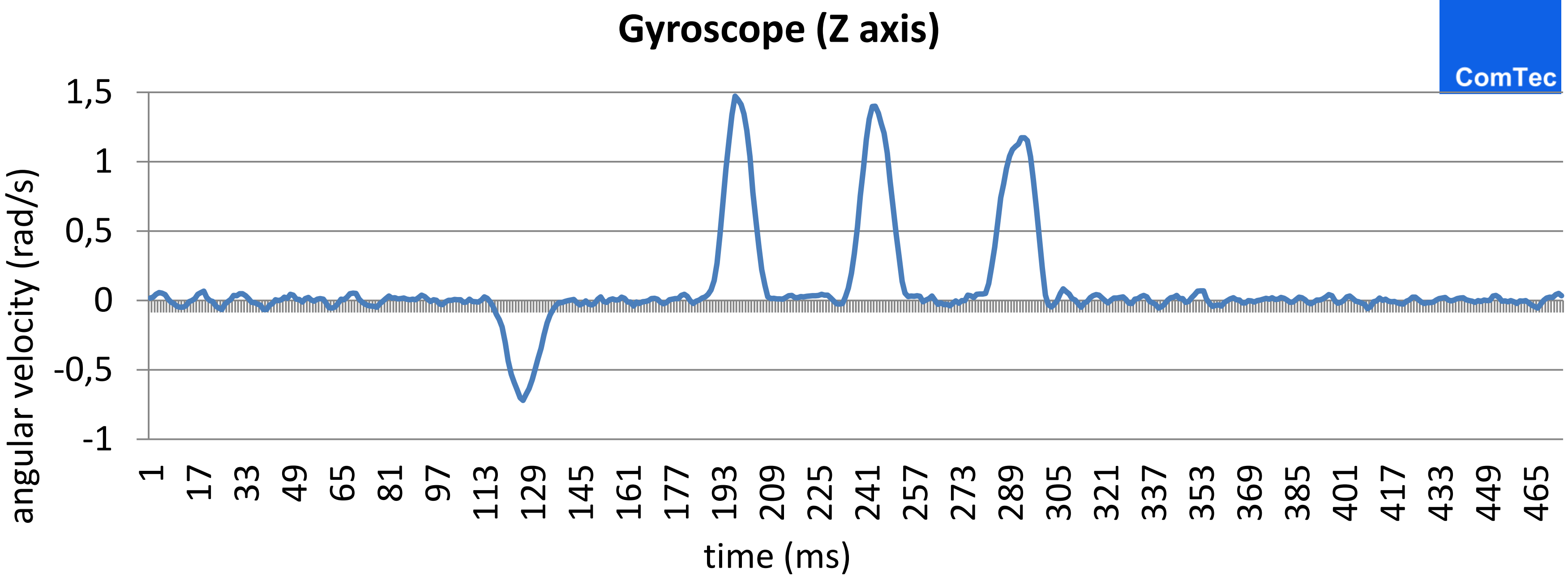
What are time series?

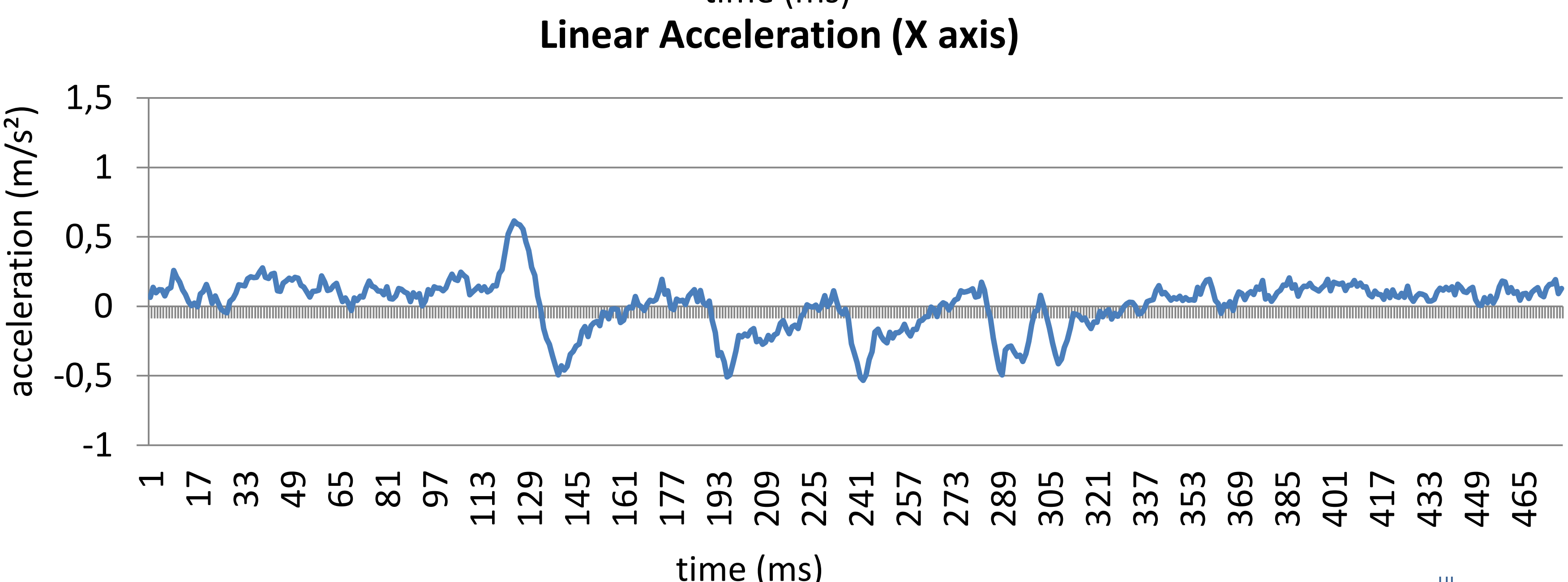
- Some examples for time series



Weather in Kassel [1]

© ComTec 2019


Gyroscope (Z axis)
angular velocity (rad/s)
1,5
1
0,5
0
-0,5
-1
1 17 33 49 65 81 97 113 129 145 161 177 193 209 225 241 257 273 289 305 321 337 353 369 385 401 417 433 449 465
time (ms)


Linear Acceleration (X axis)
acceleration (m/s²)
1,5
1
0,5
0
-0,5
-1
1 17 33 49 65 81 97 113 129 145 161 177 193 209 225 241 257 273 289 305 321 337 353 369 385 401 417 433 449 465
time (ms)

© ComTec 2019

What are time series?

“When a variable is *measured sequentially in time over* or *at a fixed interval*, known as the sampling interval, the resulting data form a *time series*.” [2]

Time intervals between observations

- Equally spaced time series
  - One tax return per year, one payslip per month, popular holiday destinations per season, ..
- Unevenly spaced time series
  - Earthquakes, floods, astronomical observations (e.g. supernovas), ..
- Typical sampling rate unit for sensor values is Hertz (Hz), the cycles per second

Time series segmentation

ComTec

- Types of segmentation algorithms which need complete time series as input
  - Top-Down

    "The time series is recursively partitioned until some stopping criteria is met [3]"

  - Bottom-Up

    "Starting from the finest possible approximation, segments are merged until some stopping criteria is met [3]"

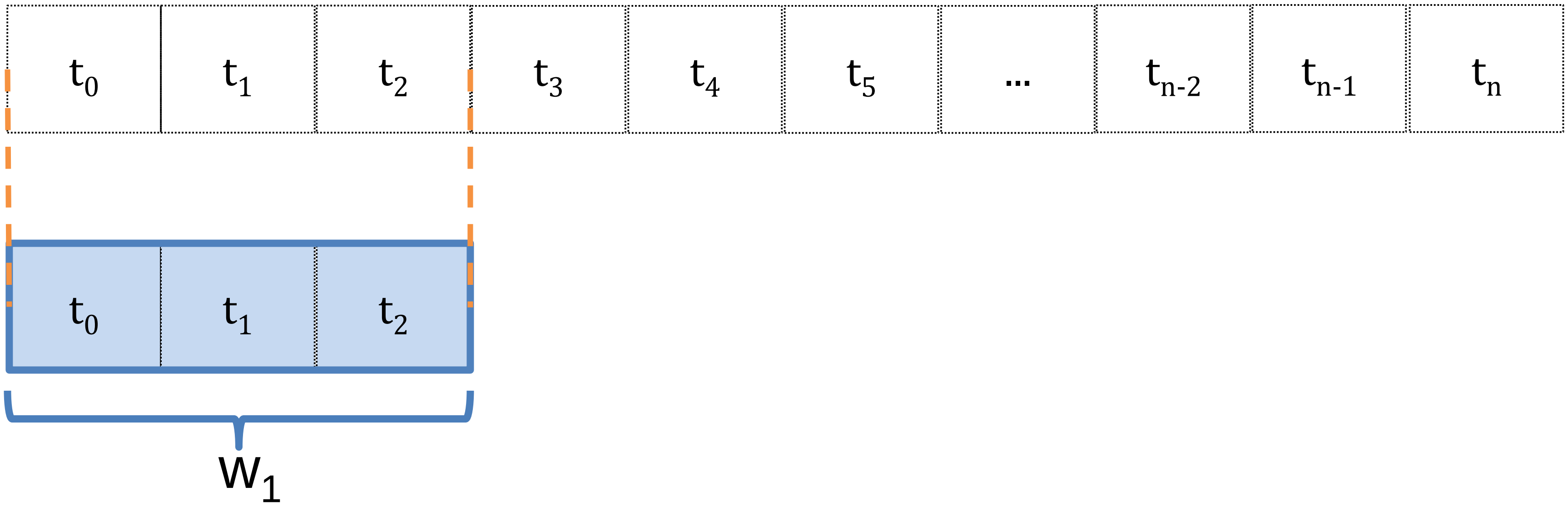
  - Sliding Window Algorithm

Sliding Window Algorithm

- The data is segmented in so-called windows.
- The window slides over the time series

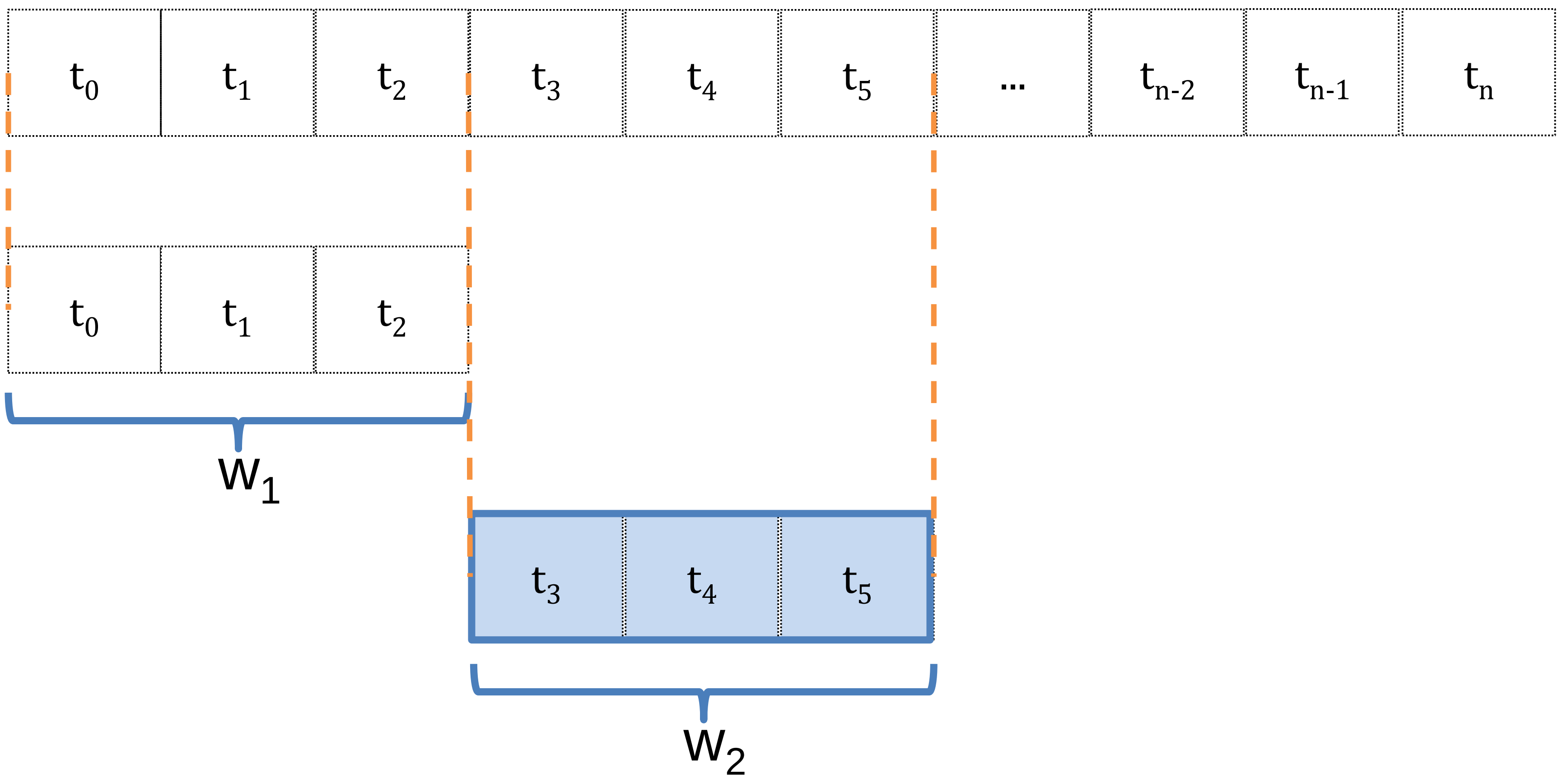
- Two parameters:
  - Window length w, numbers of instances per window
  - Overlap [in percentage], how many instances are used of the previous window for the current window

Sliding Window Algorithm
no overlapping, window length = 3

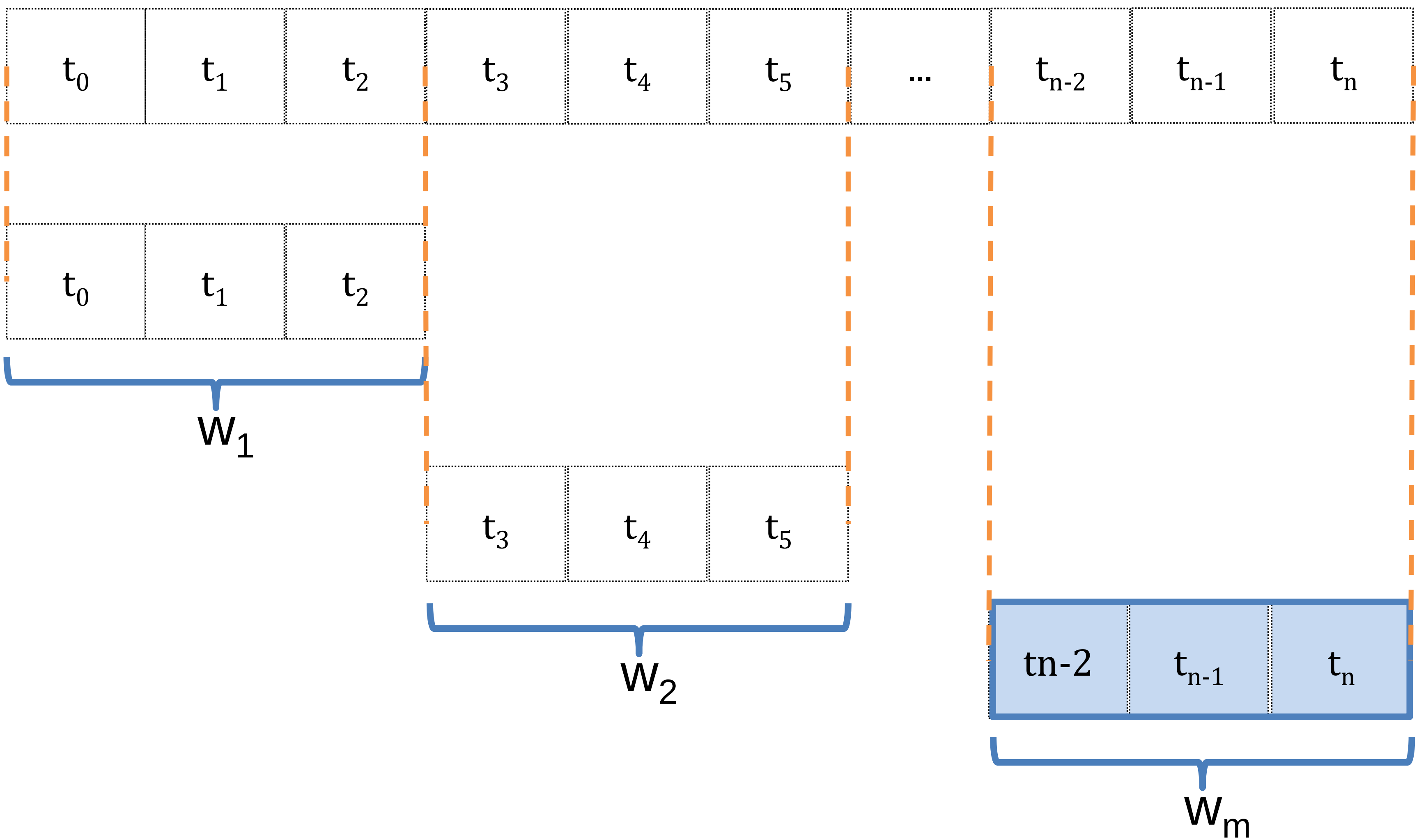


© ComTec 2019

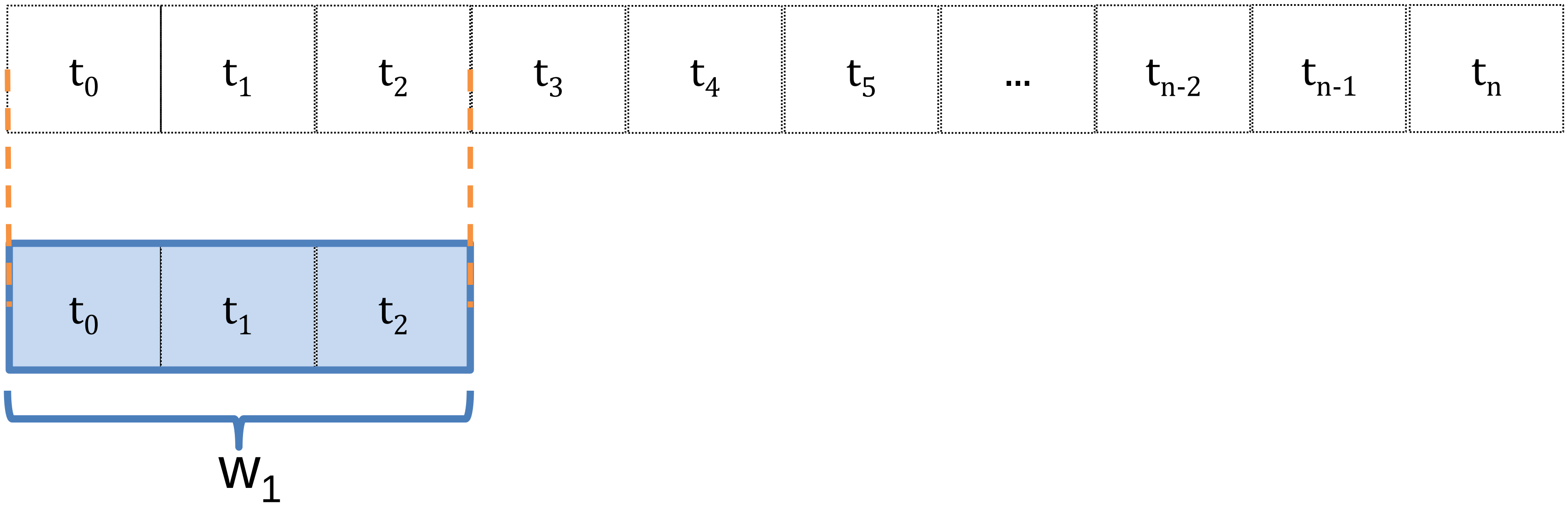
Sliding Window Algorithm
no overlapping, window length = 3



Sliding Window Algorithm
no overlapping, window length = 3

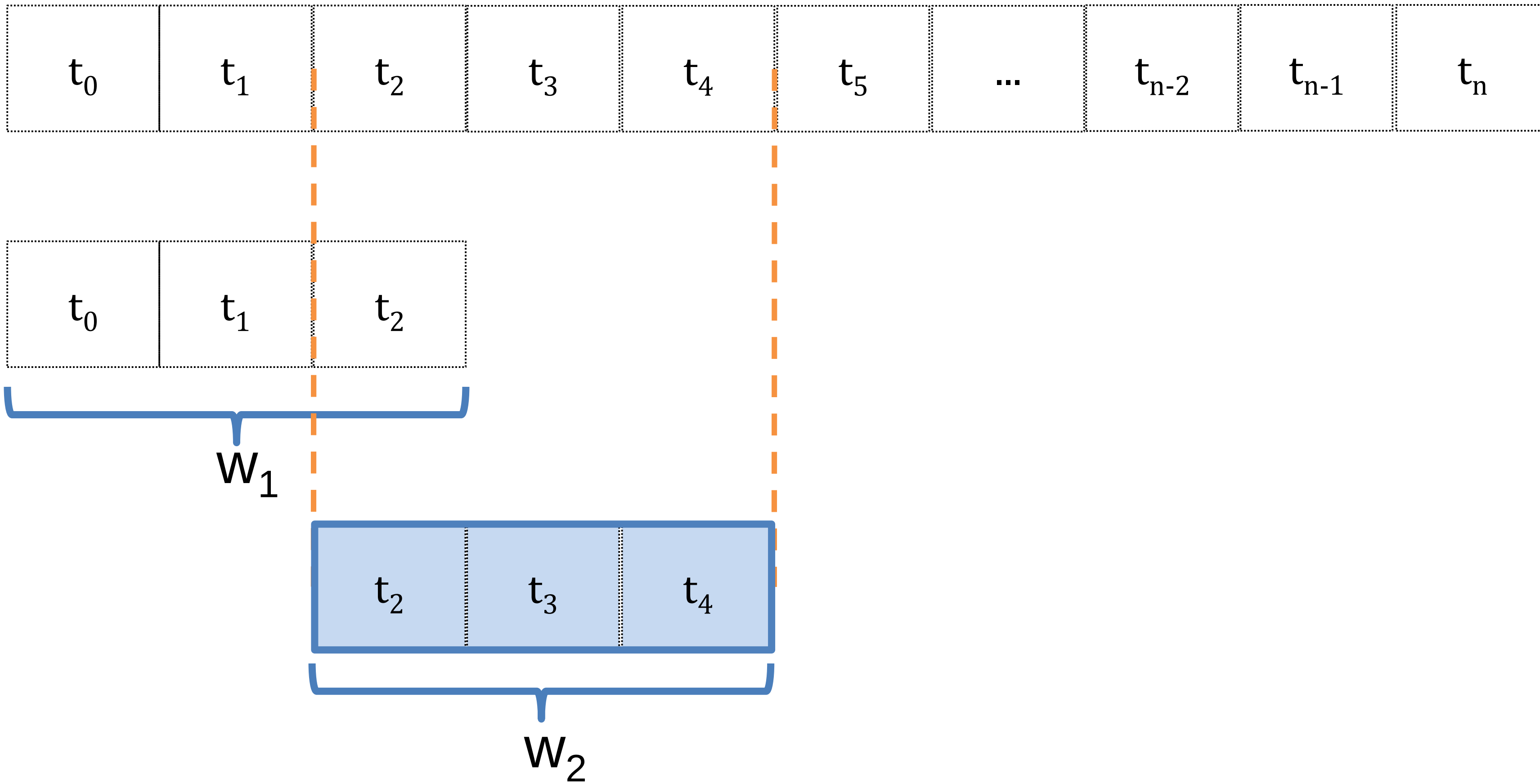


Sliding Window Algorithm
overlap = 33%, window length = 3

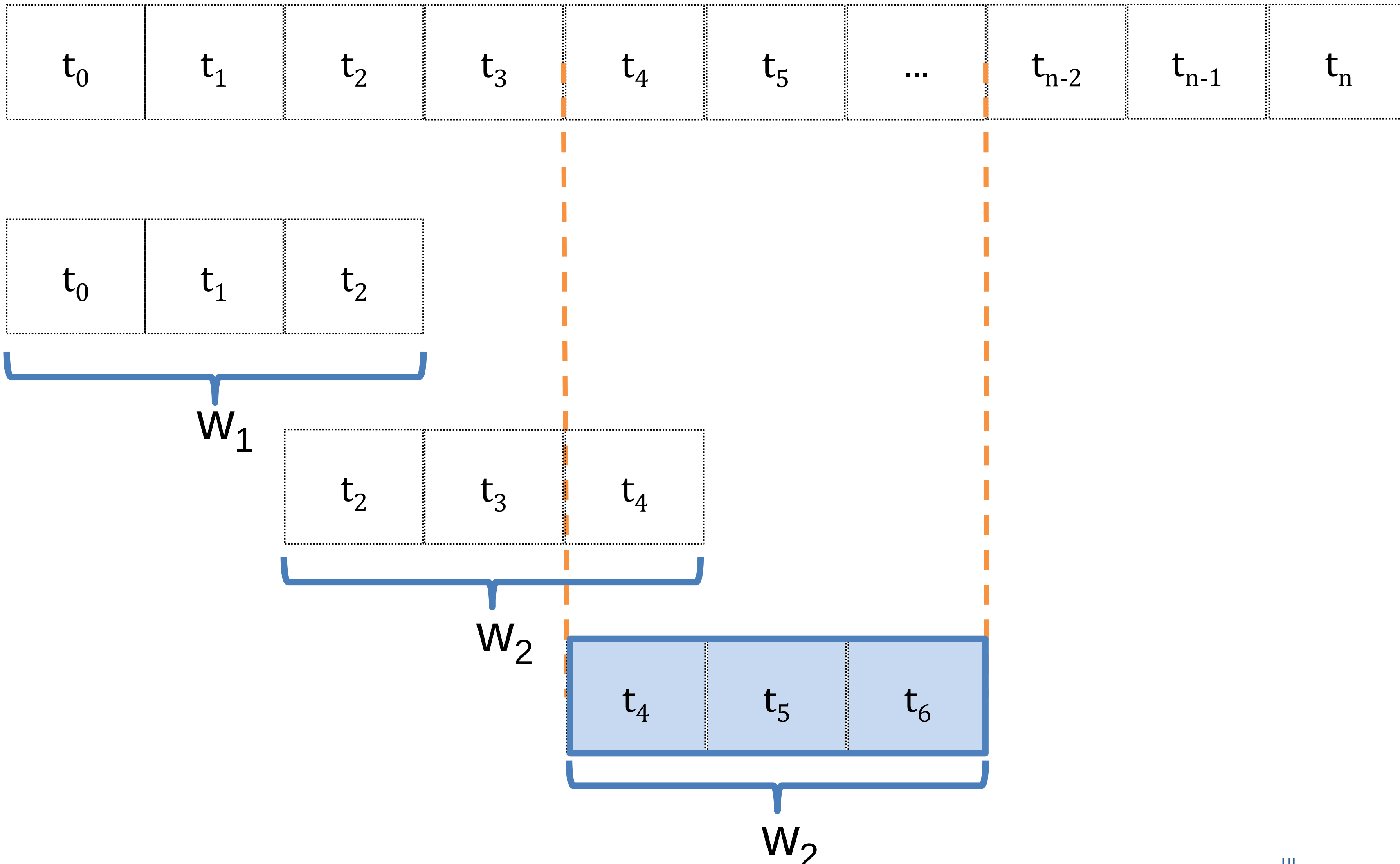


© ComTec 2019

Sliding Window Algorithm
overlap = 33%, window length = 3

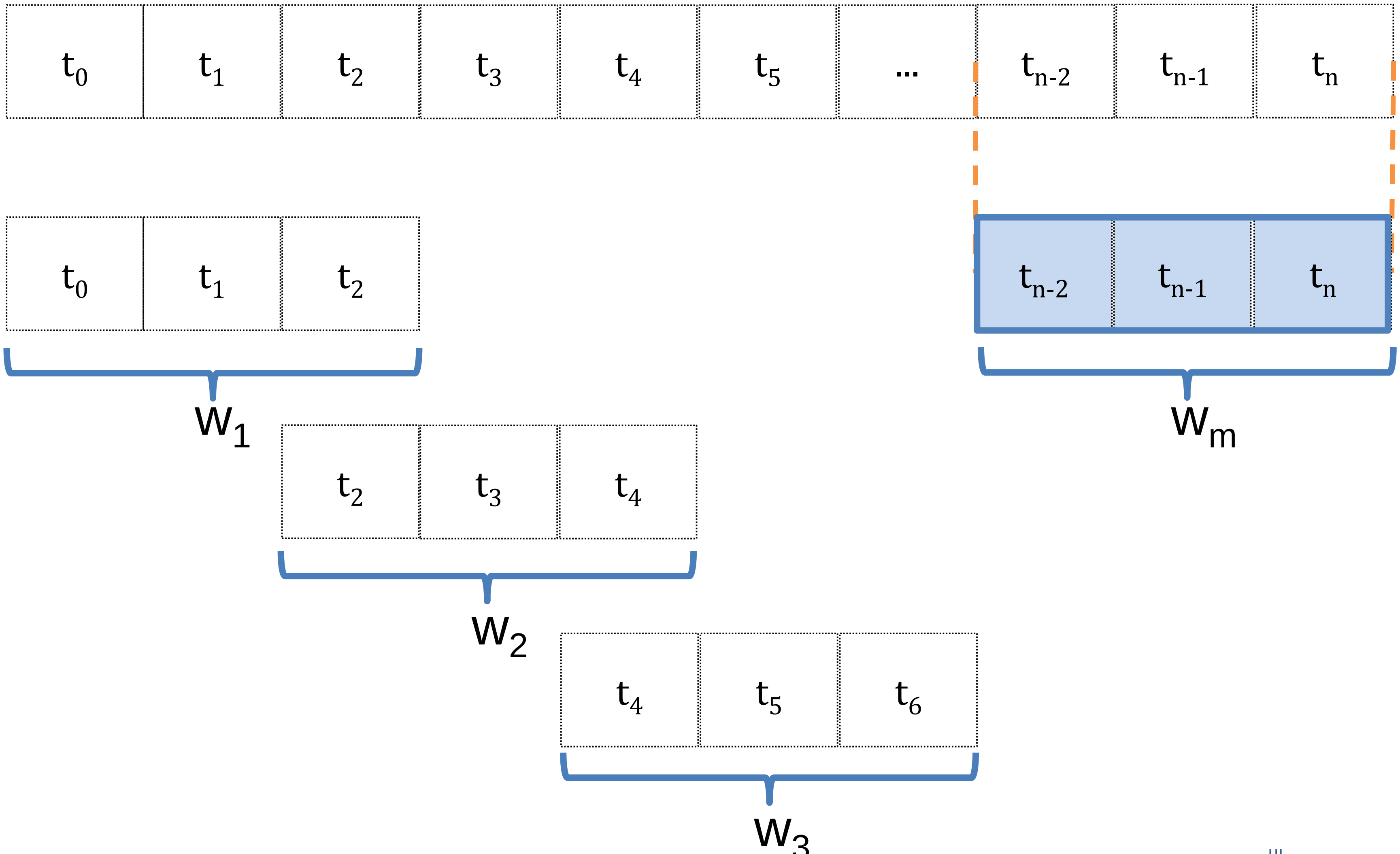


Sliding Window Algorithm
overlap = 33%, window length = 3

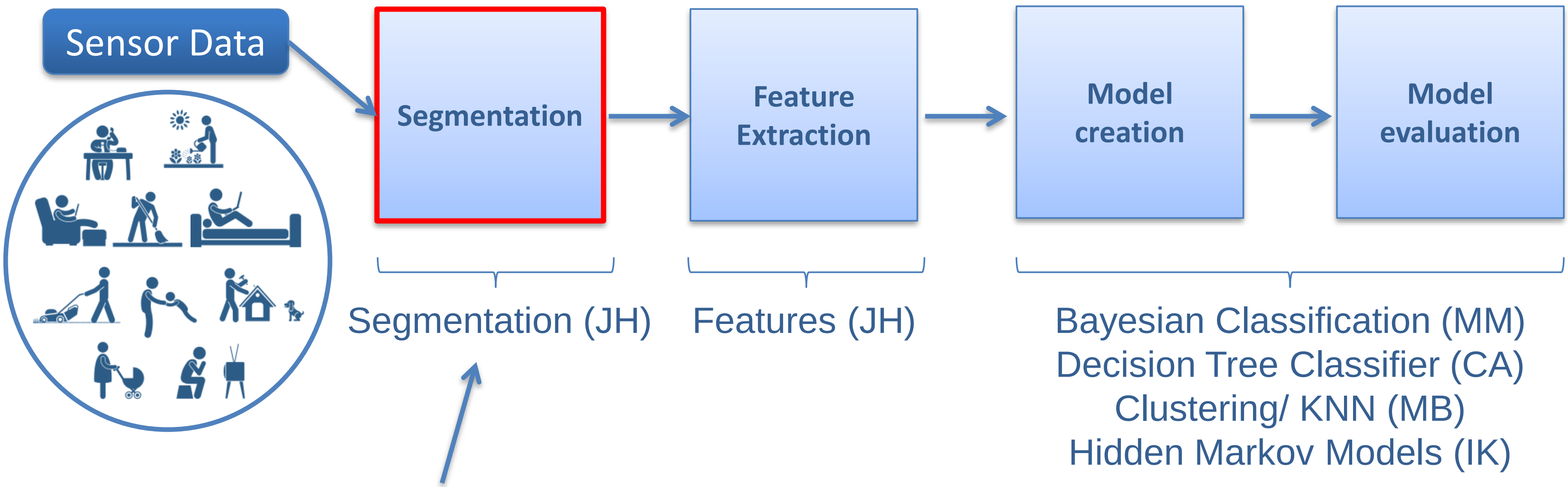


© ComTec 2019

Sliding Window Algorithm
overlap = 33%, window length = 3

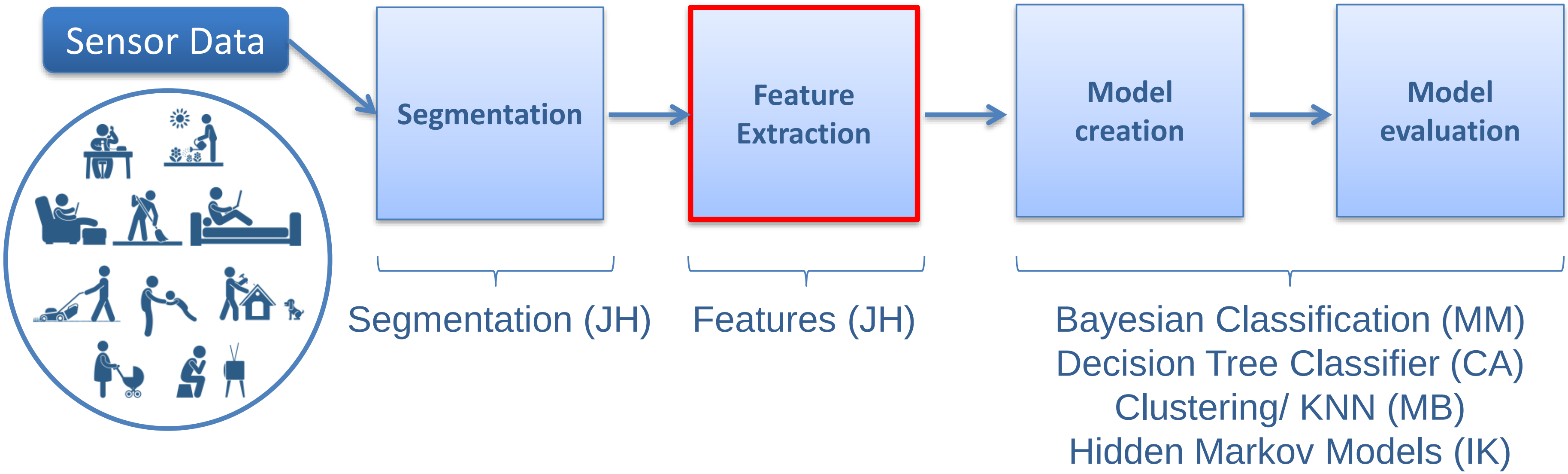


Topics for this lecture



- Time Series
- Time intervals
- Segmentation Algorithms
  - Sliding Window Algorithm

Topics for this lecture



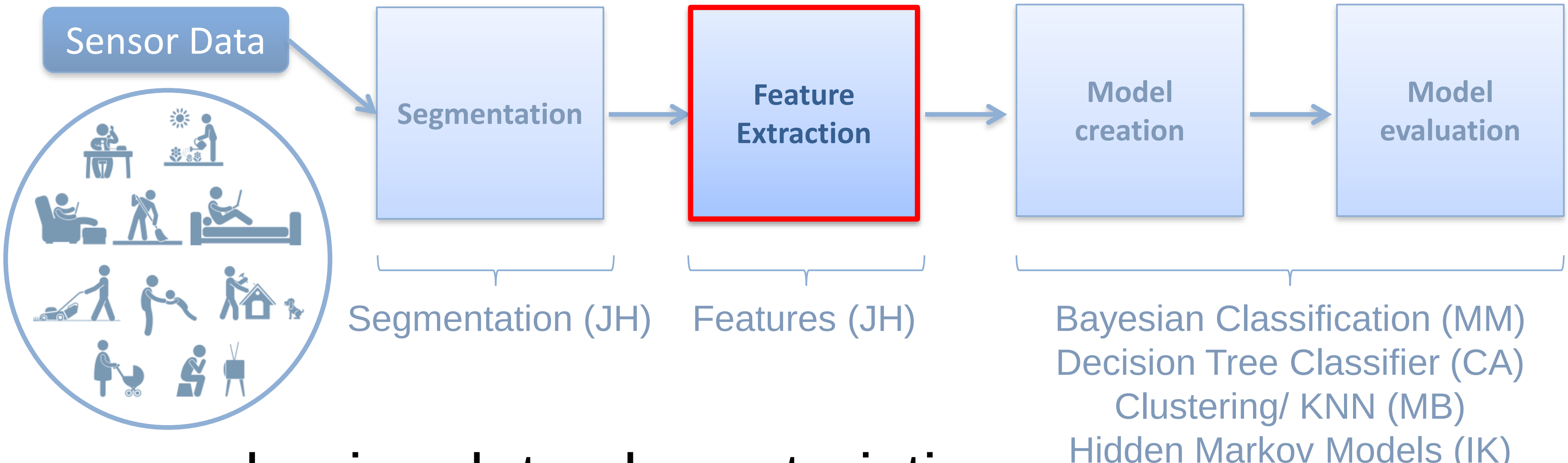
© ComTec 2019

Feature Extraction

- Data Features
  - Central tendency
  - Measures of Spread
  - Skewness
- Feature Selection
  - Filter
  - Wrapper
  - Embedded methods
  - Hybrid methods

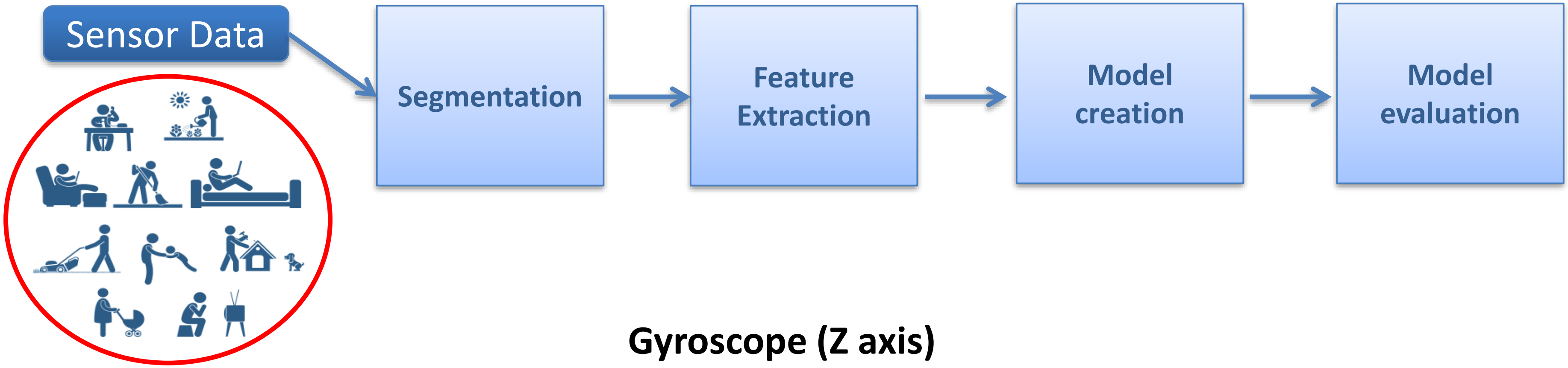
Feature Extraction

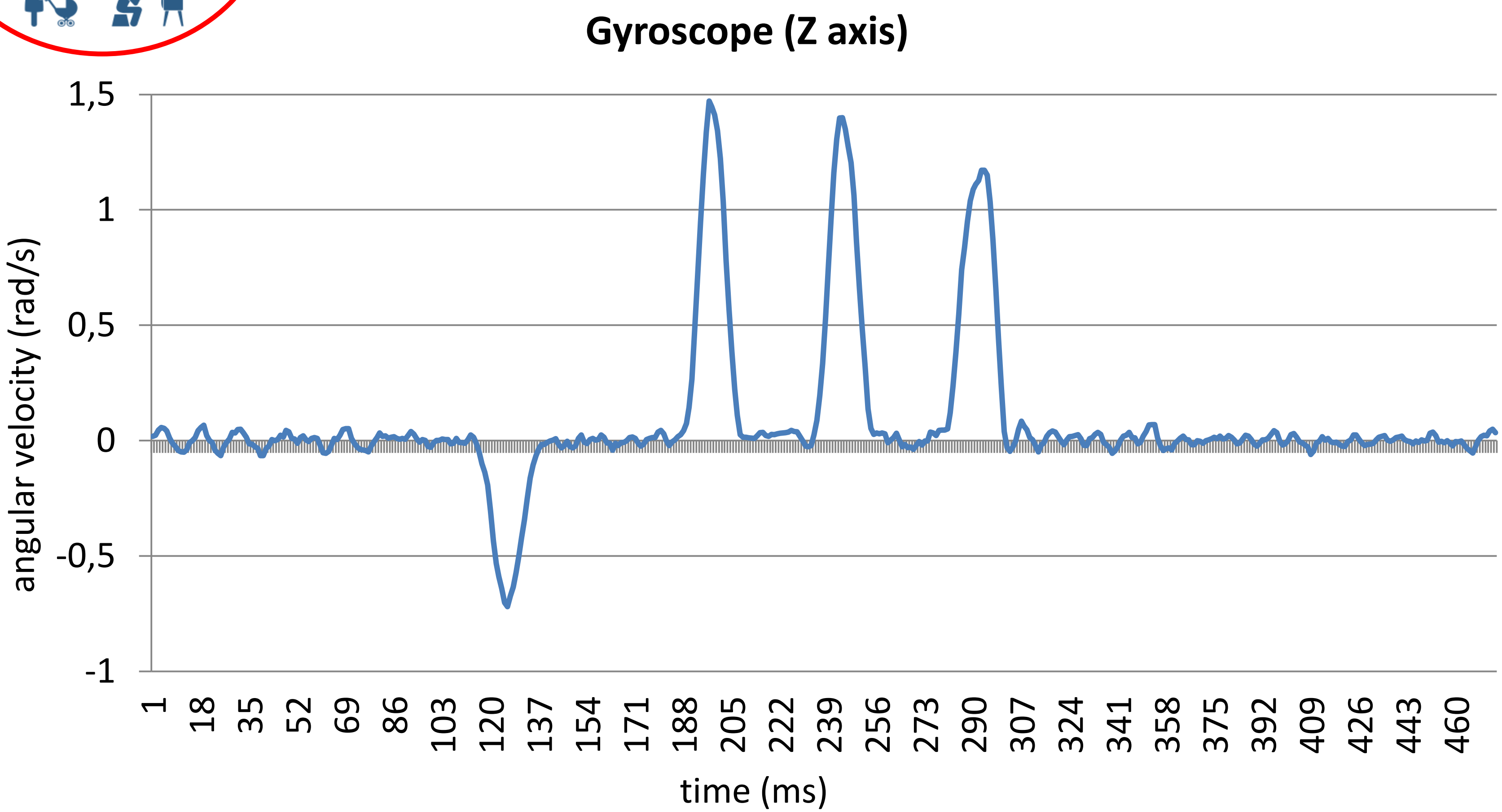
ComTec

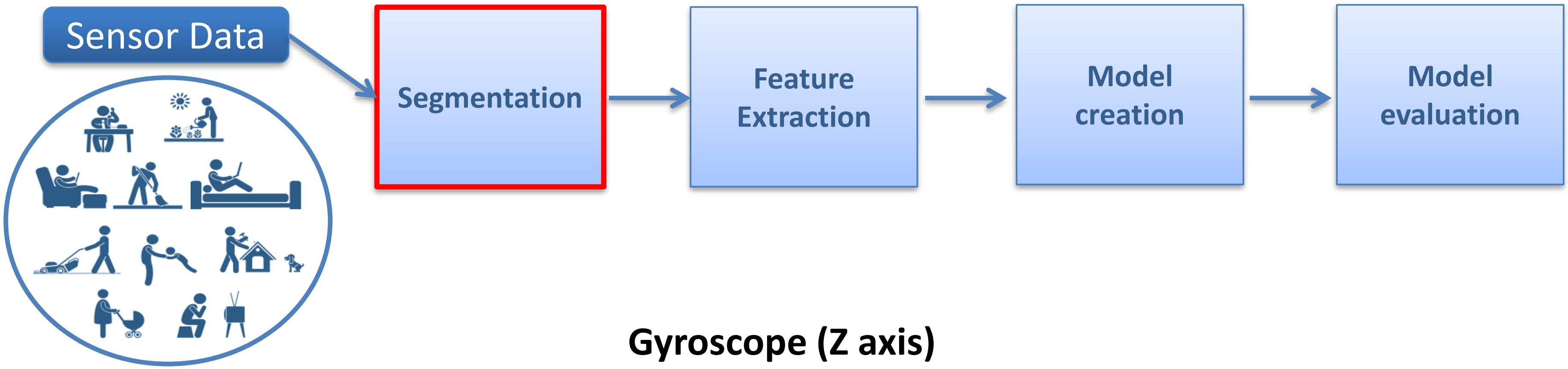


- emphasize data characteristics
- reduce amount of data
- save energy

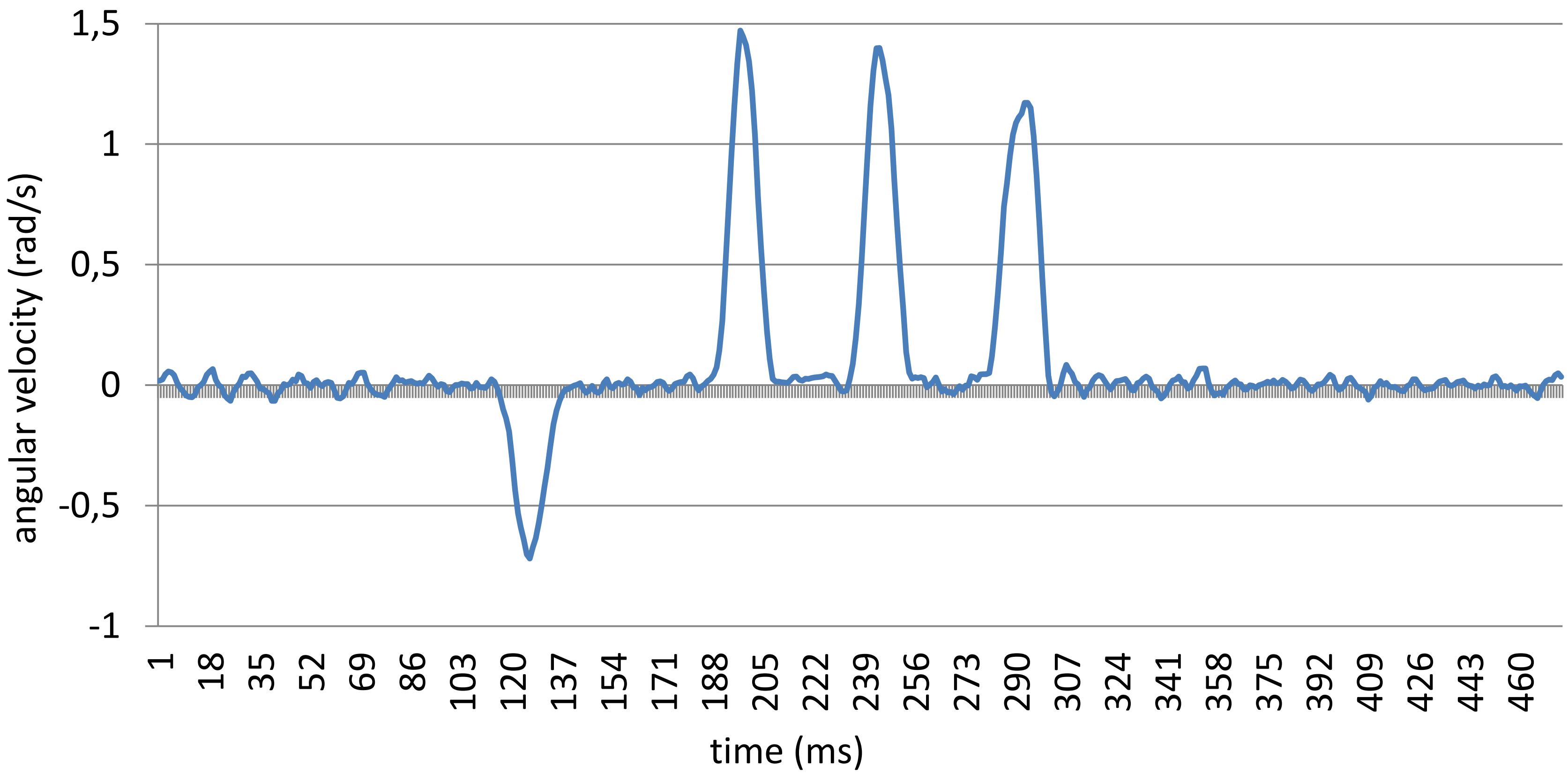
© ComTec 2019


Sensor Data
Segmentation
Feature Extraction
Model creation
Model evaluation

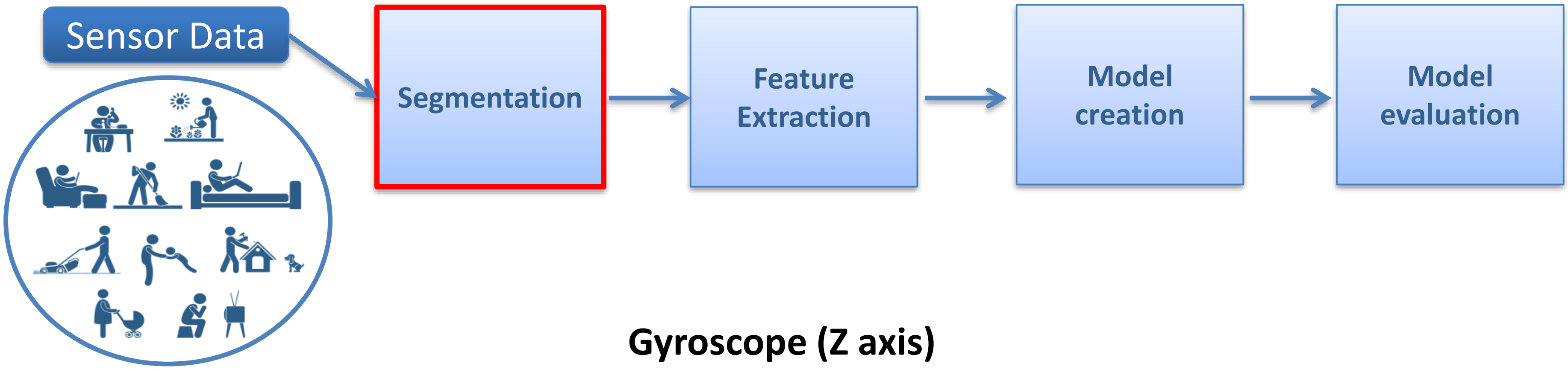

Gyroscope (Z axis)
angular velocity (rad/s)
1,5
1
0,5
0
-0,5
-1
1
18
35
52
69
86
103
120
137
154
171
188
205
222
239
256
273
290
307
324
341
358
375
392
409
426
443
460
time (ms)

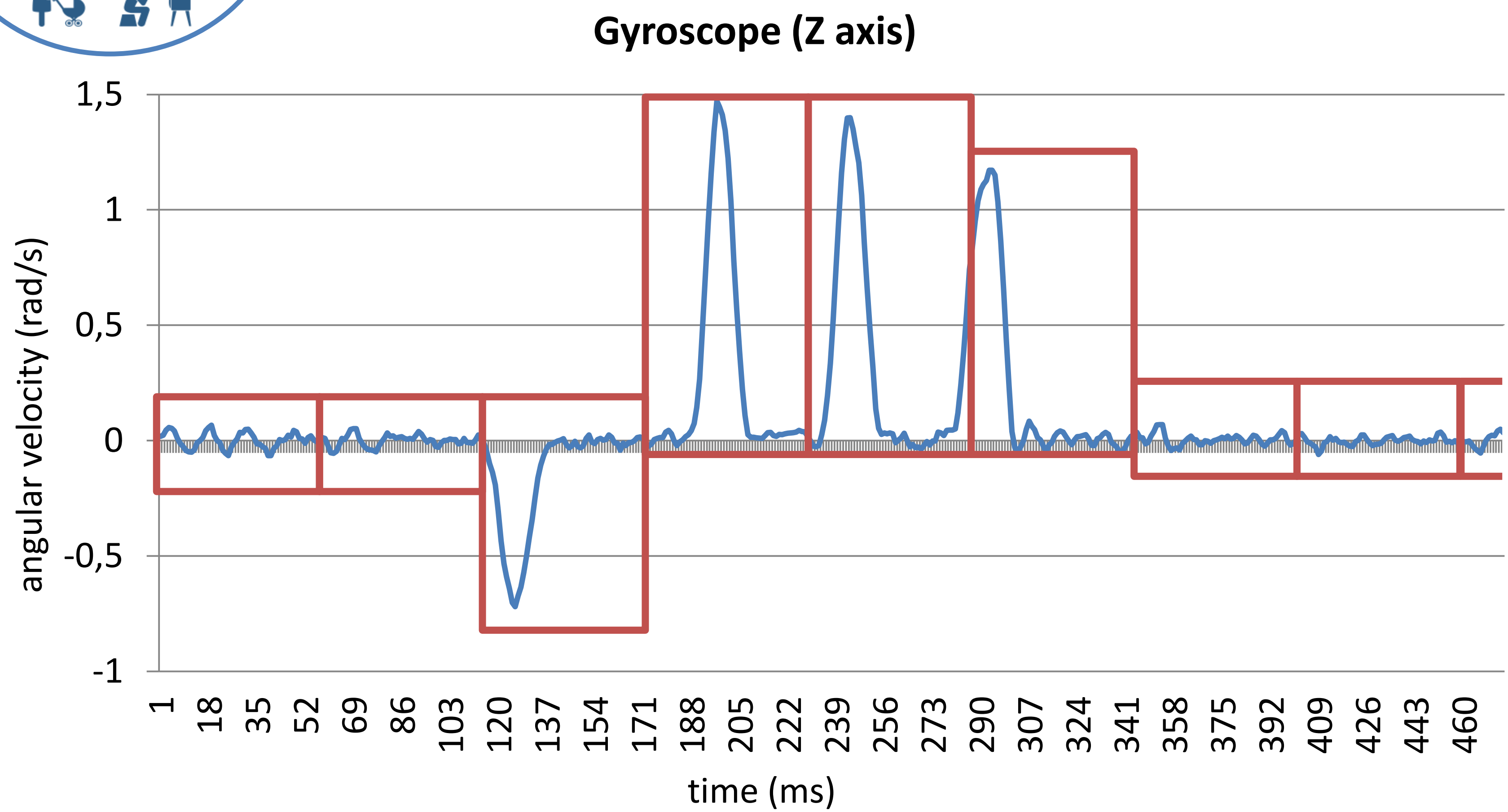


Gyroscope (Z axis)

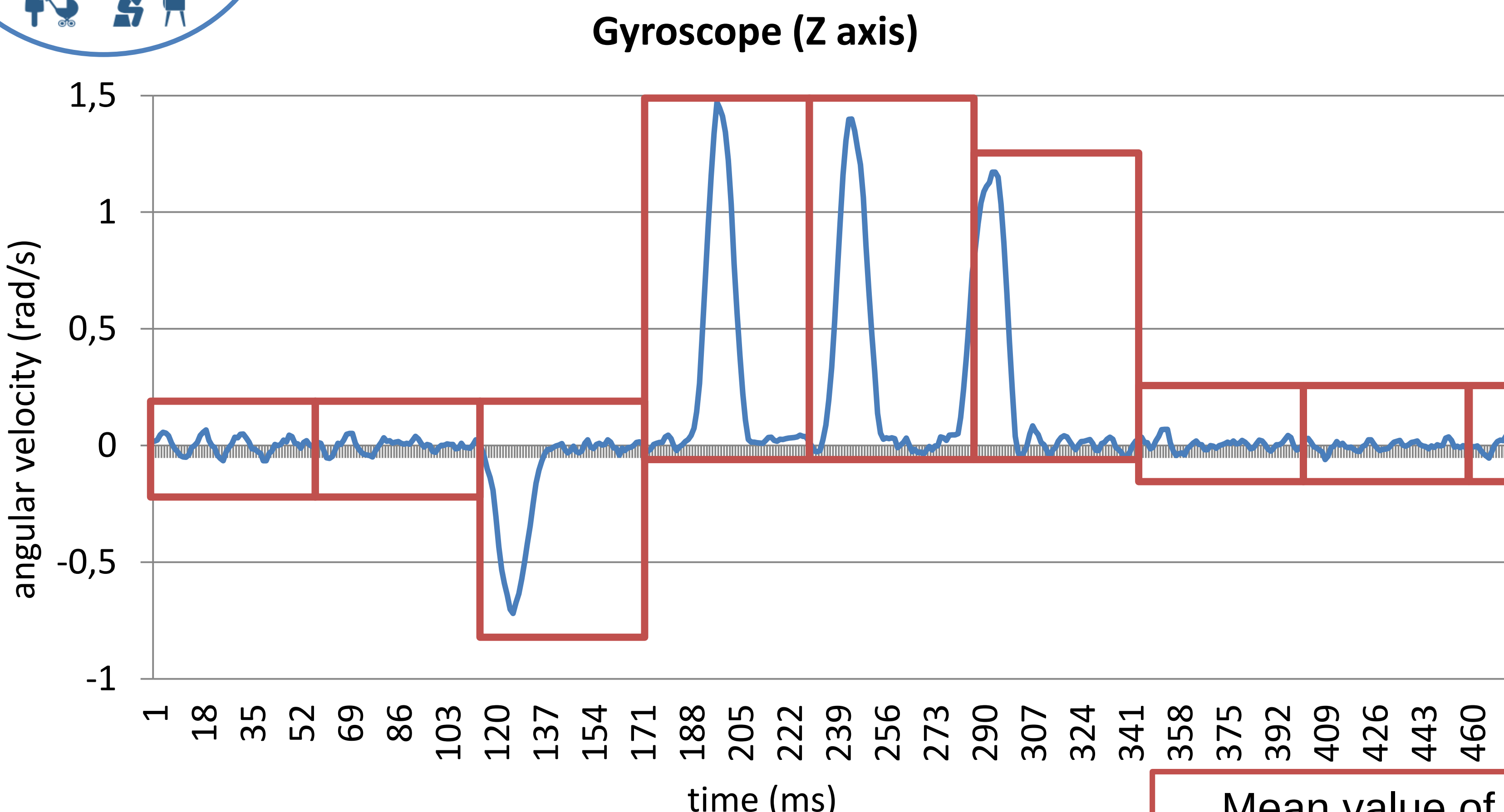


© ComTec 2019


Sensor Data
Segmentation
Feature Extraction
Model creation
Model evaluation


Gyroscope (Z axis)
angular velocity (rad/s)
1,5
1
0,5
0
-0,5
-1
1
18
35
52
69
86
103
120
137
154
171
188
205
222
239
256
273
290
307
324
341
358
375
392
409
426
443
460
time (ms)

Sensor Data → Segmentation → **Feature Extraction** → Model creation → Model evaluation

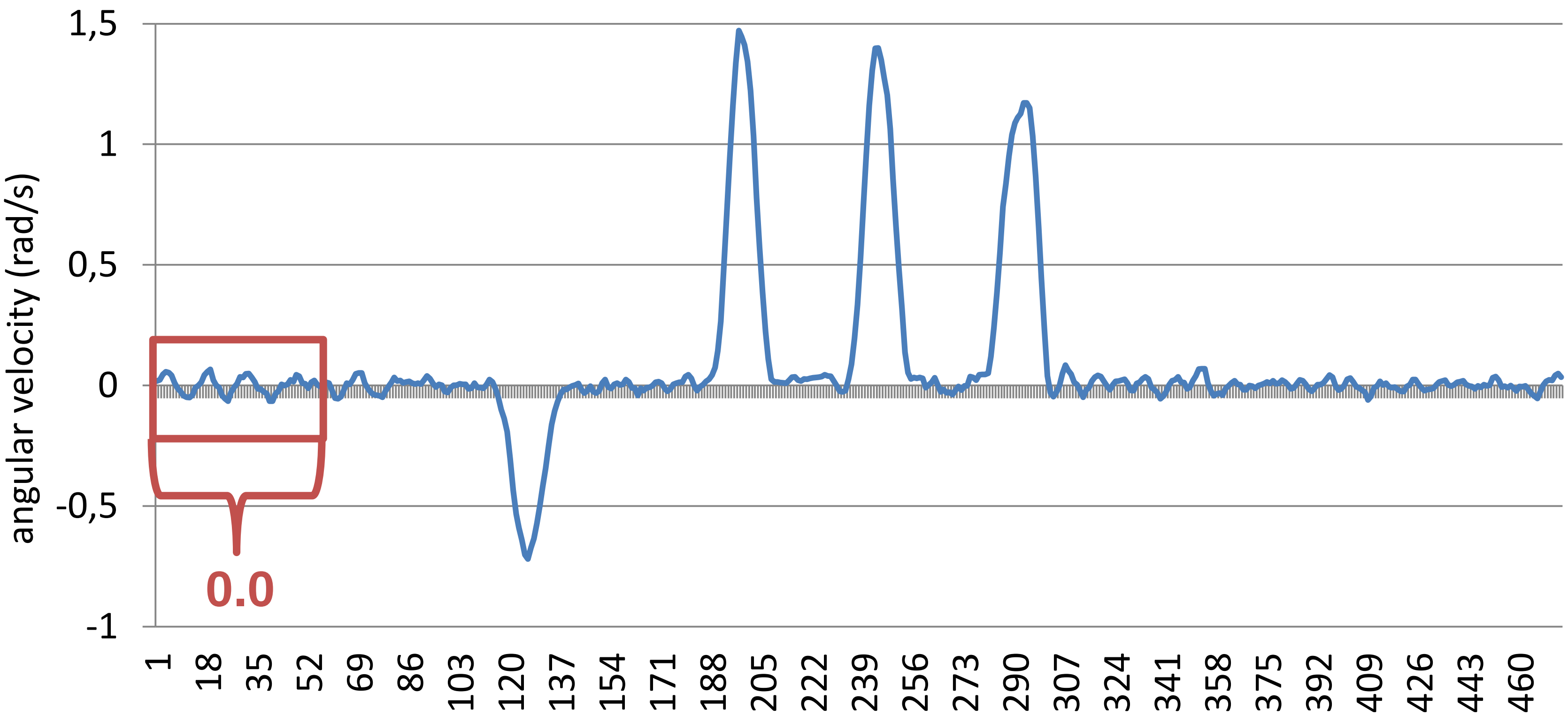


Mean value of each window

© ComTec 2019

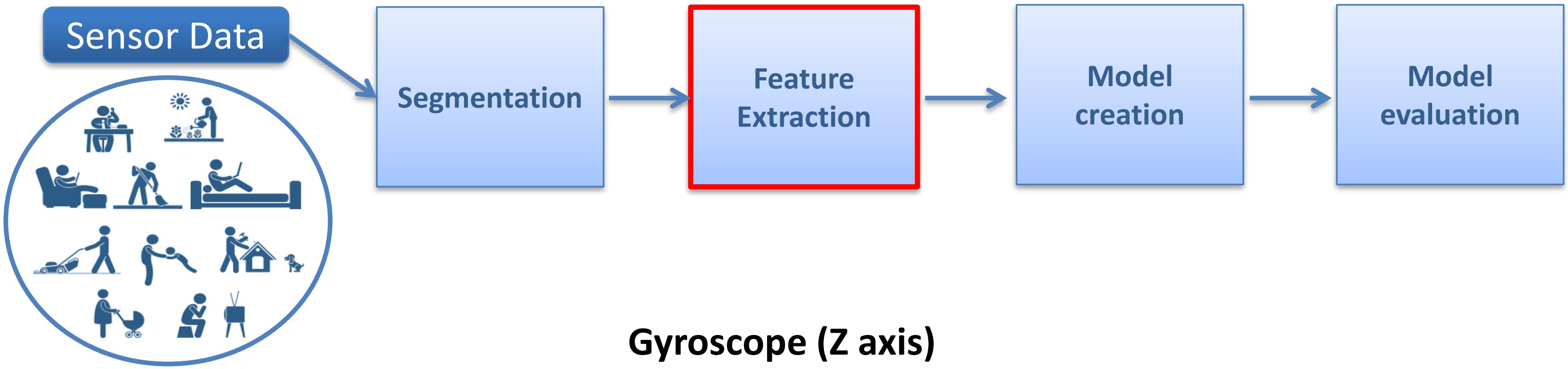
Sensor Data → Segmentation → **Feature Extraction** → Model creation → Model evaluation

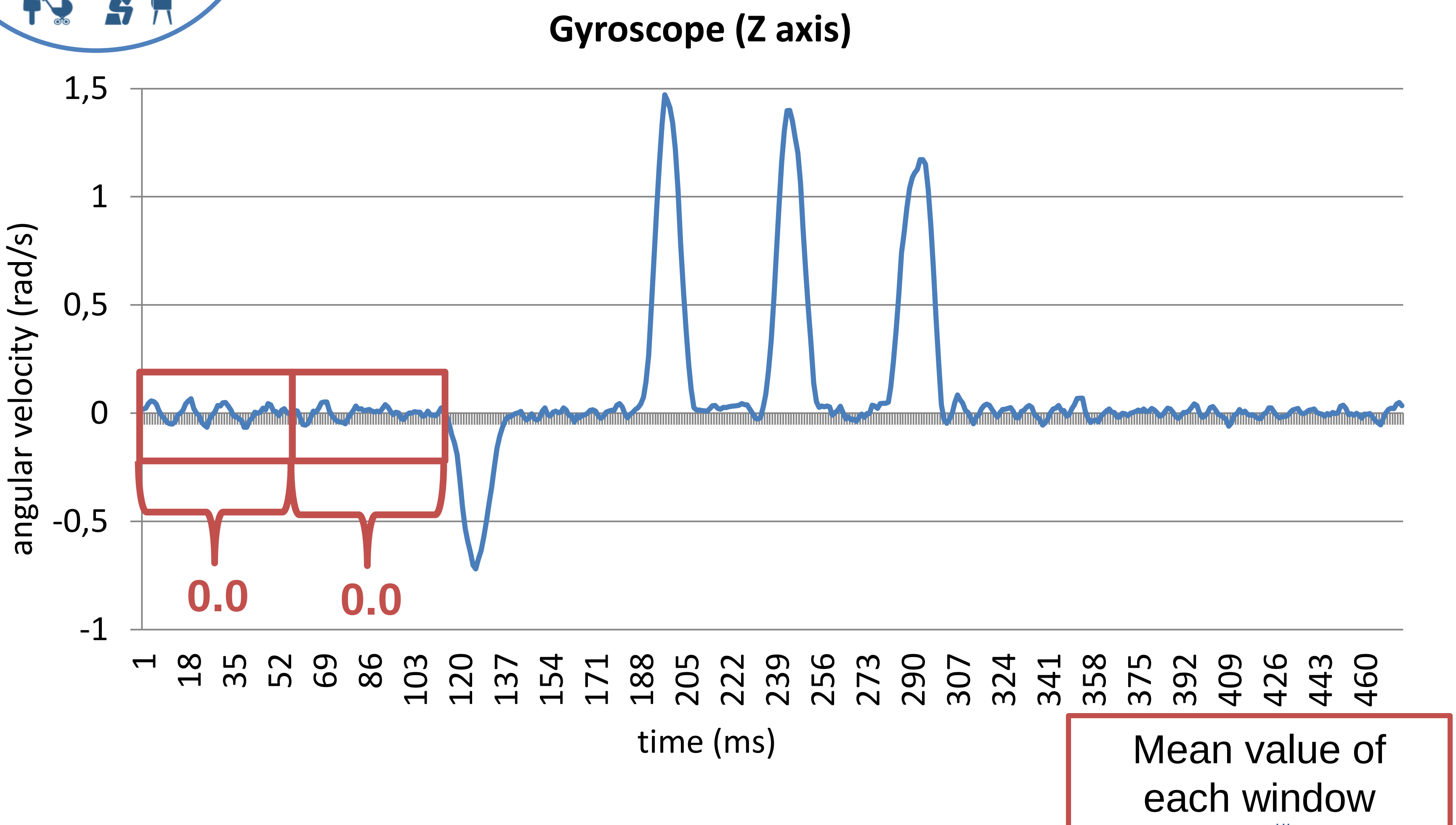
Gyroscope (Z axis)



Mean value of each window

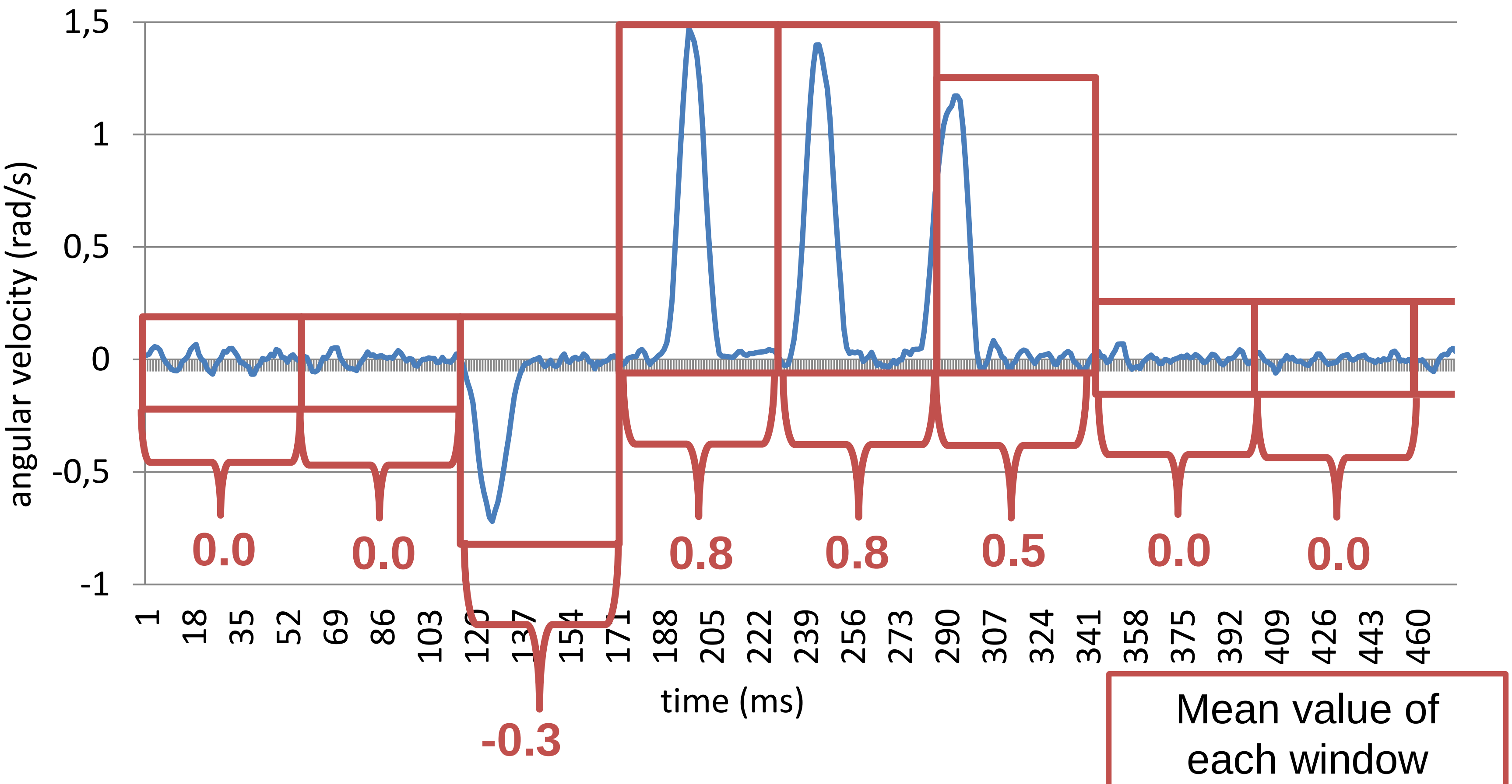
© ComTec 2019


Sensor Data
Segmentation
Feature Extraction
Model creation
Model evaluation


Gyroscope (Z axis)
angular velocity (rad/s)
1,5
1
0,5
0
-0,5
-1
1 18 35 52 69 86 103 120 137 154 171 188 205 222 239 256 273 290 307 324 341 358 375 392 409 426 443 460
time (ms)
0.0
0.0
Mean value of each window

Sensor Data → Segmentation → **Feature Extraction** → Model creation → Model evaluation

Gyroscope (Z axis)



© ComTec 2019

Data Features

ComTec

Statistical Features	
Mean	Averaged Derivatives
Median	Skewness
Mode	Zero Crossing Rate
Standard Deviation	Mean Crossing Rate
Variance	Pairwise Correlation
Covariance	Time between peaks
Root Mean square	Range Interquatile Range
Median Absolute Deviation (MAD)	Etc...

© ComTec 2019

Definition Central tendency (centrality)

ComTec

“The tendency of quantitative data to cluster around some central value. The central value is commonly estimated by the mean, median, or mode, whereas the closeness with which the values surround the central value is commonly quantified using the standard deviation or variance. The phrase ‘central tendency’ was first used in the late 1920s.“

[4]

© ComTec 2019

Central tendency

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

- Mean (arithmetic mean, average)

  Sum all the values in the sample and divide by the number of items.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Central tendency

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

- Mean (arithmetic mean, average)

Sum all the values in the sample and
divide by the number of items.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Mean = 7.1

© ComTec 2019

Central tendency

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Median

Sort the data from lowest to highest.
Pick up the middle number.

$$x_{med} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & for\ odd\ n \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & for\ even\ n \end{cases}$$

1, 2, 4, 4, 4, 4, 4, 7, 8, 8, 9, 30

© ComTec 2019

Central tendency

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

- Median

Sort the data from lowest to highest.
Pick up the middle number.

$$x_{med} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & for\ odd\ n \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & for\ even\ n \end{cases}$$

1, 2, 4, 4, 4, 4, 4, 7, 8, 8, 9, 30

Median = 4

© ComTec 2019

Central tendency

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

- Median

Sort the data from lowest to highest.
Pick up the middle number.

$$x_{med} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & for\ odd\ n \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & for\ even\ n \end{cases}$$

- *resistant or robust*
- $50\%\ of\ the\ data \leq x_{med}$
- $50\%\ of\ the\ data \geq x_{med}$

1, 2, 4, 4, 4, 4, 4, 7, 8, 8, 9, 30

Median = 4

Central tendency

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

- Mode

The mode of a sample is the element that occurs most often in the collection

1, 2, 4, 4, 4, 4, 4, 7, 8, 8, 9, 30

© ComTec 2019

Central tendency

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

- Mode

The mode of a sample is the element that occurs most often in the collection

1, 2, 4, 4, 4, 4, 4, 7, 8, 8, 9, 30

Mode = 4

© ComTec 2019

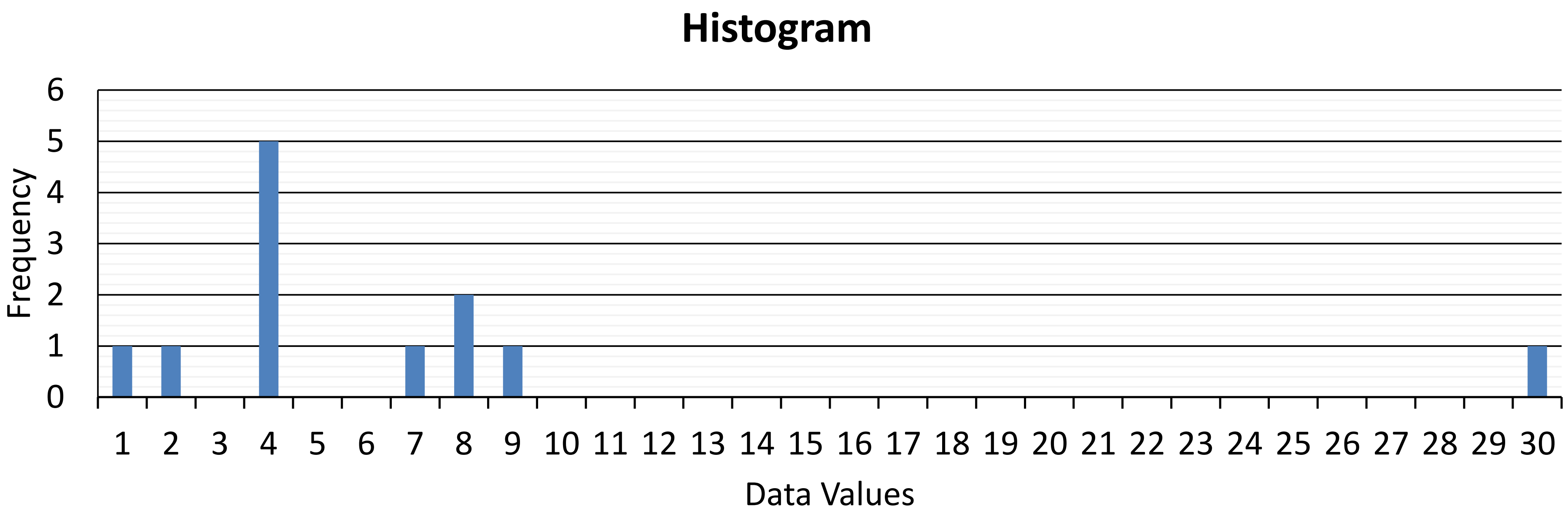
Central tendency

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Mode

The mode of a sample is the element that occurs most often in the collection



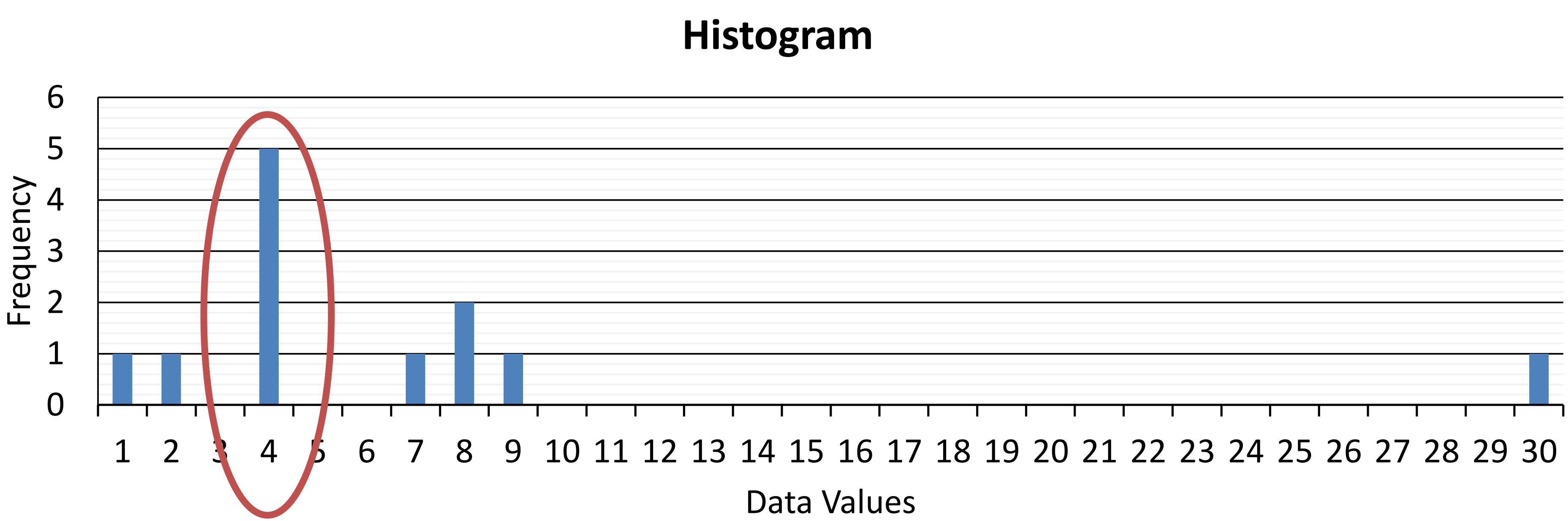
© ComTec 2019

Central tendency

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Mode

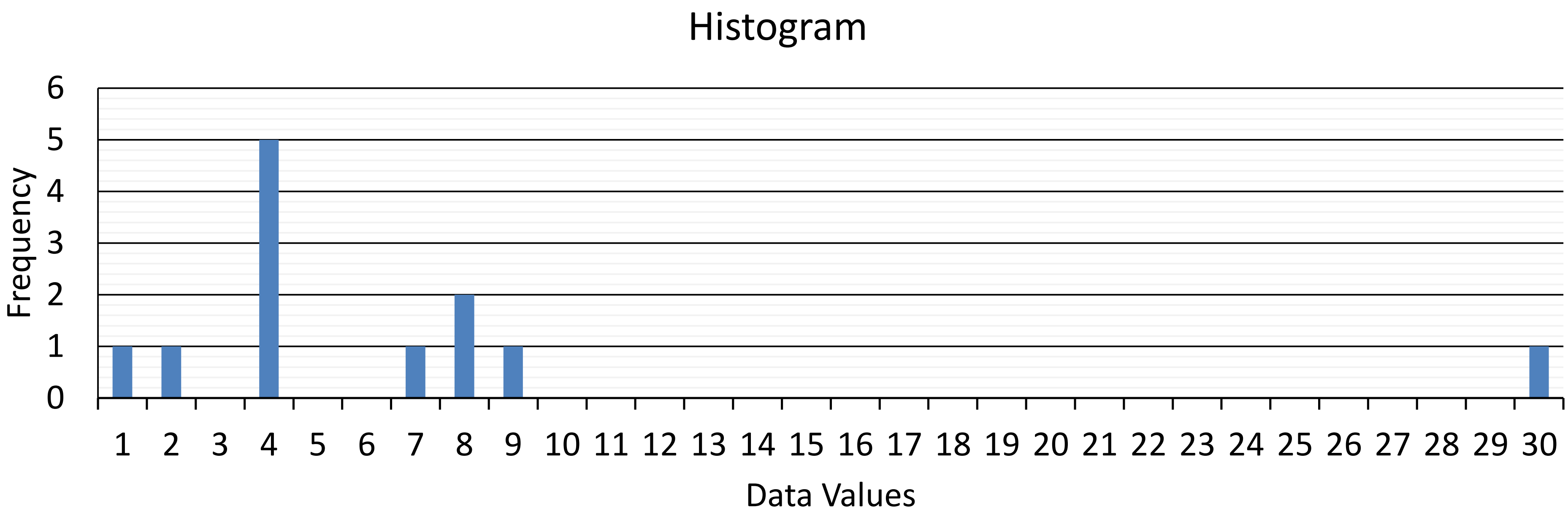
The mode of a sample is the element that occurs most often in the collection



Mode = 4

© ComTec 2019

Central tendency - Conclusion

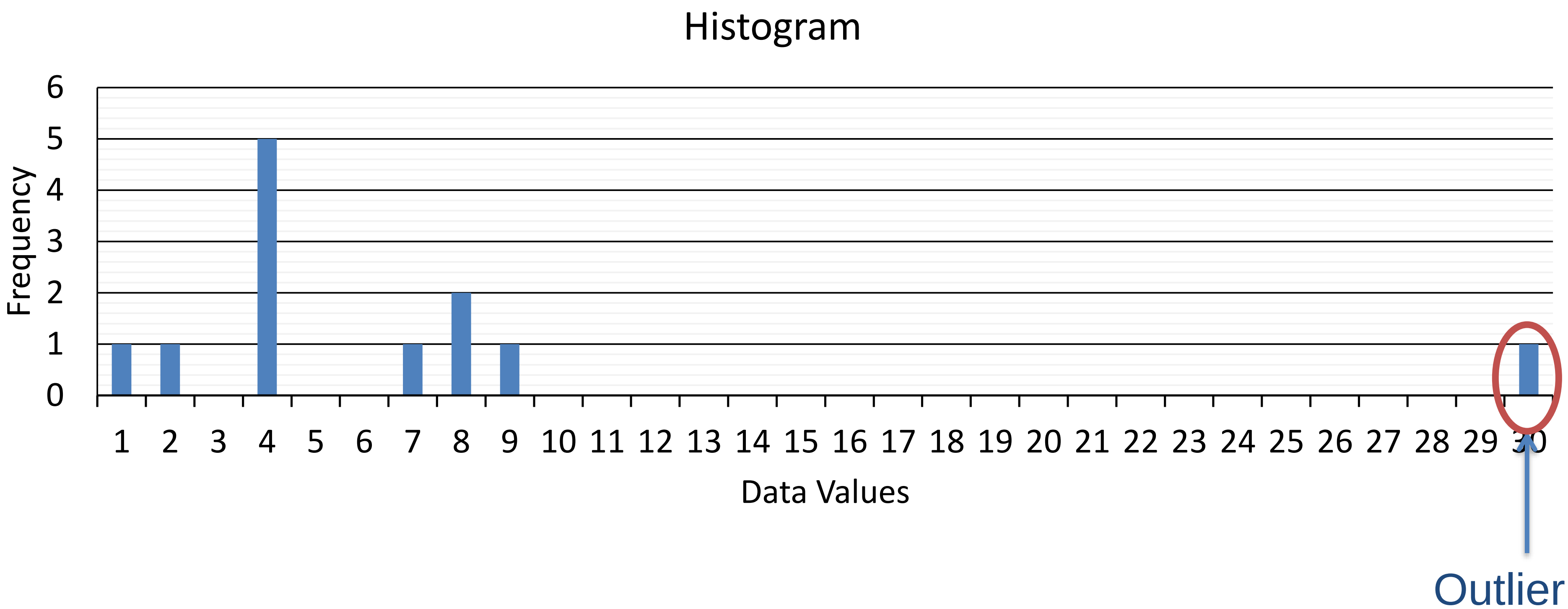


Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30)

Mean: 7.1
Median: 4
Mode: 4

Central tendency - Conclusion

ComTec



Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30)

Mean: 7.1
Median: 4
Mode: 4

„An observation that is very different to other observations in a set of data“ is called outlier

[4]

© ComTec 2019

Measures of Spread

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Range

The maximum minus the minimum value of the sample.

1, 2, 4, 4, 4, 4, 4, 7, 8, 8, 9, 30

© ComTec 2019

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Range

The maximum minus the minimum value of the sample.

1, 2, 4, 4, 4, 4, 4, 7, 8, 8, 9, 30

Range = 30-1=29

© ComTec 2019

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Quartile

  Use the Median to separate the data:

  1, 2, 4, 4, 4, 4   4, 7, 8, 8, 9, 30

  For each half we find the median:

© ComTec 2019

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Quartile

  Use the Median to separate the data:

  1, 2, 4, 4, 4, 4 4, 7, 8, 8, 9, 30

  For each half we find the median:

  Lower quartile (25%-Quantile): 4
  Upper quartile (75%-Quantile): 8

© ComTec 2019

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Quartile

  Use the Median to separate the data:

  1, 2, 4, 4, 4, 4 4, 7, 8, 8, 9, 30

  For each half we find the median:

  Lower quartile (25%-Quantile): 4
  Upper quartile (75%-Quantile): 8

- Interquartile range (IQR)

  $$IQR = '\mathit{Upper\ quartile}' - '\mathit{Lower\ quartile}'$$

Measures of Spread

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Quartile

Use the Median to separate the data:

The Median is the 50%-Quantile.

1, 2, 4, 4, 4, 4 4, 7, 8, 8, 9, 30

For each half we find the median:

Lower quartile (25%-Quantile): 4
Upper quartile (75%-Quantile): 8

- Interquartile range (IQR)

$$IQR = 'Upper\ quartile' - 'Lower\ quartile'$$

IQR = 4

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

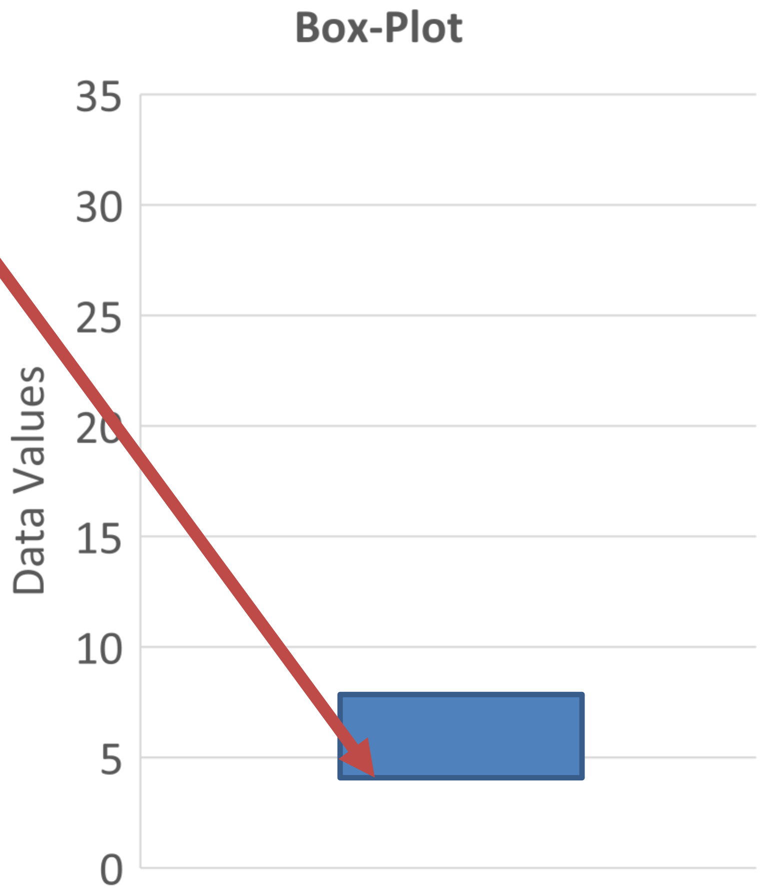
- Quartile

  Lower quartile (25%-Quantile): 4
  Upper quartile (75%-Quantile): 8

  Median (50%-Quantile): 4

- Interquartile range (IQR)

  IQR = 4



© ComTec 2019

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

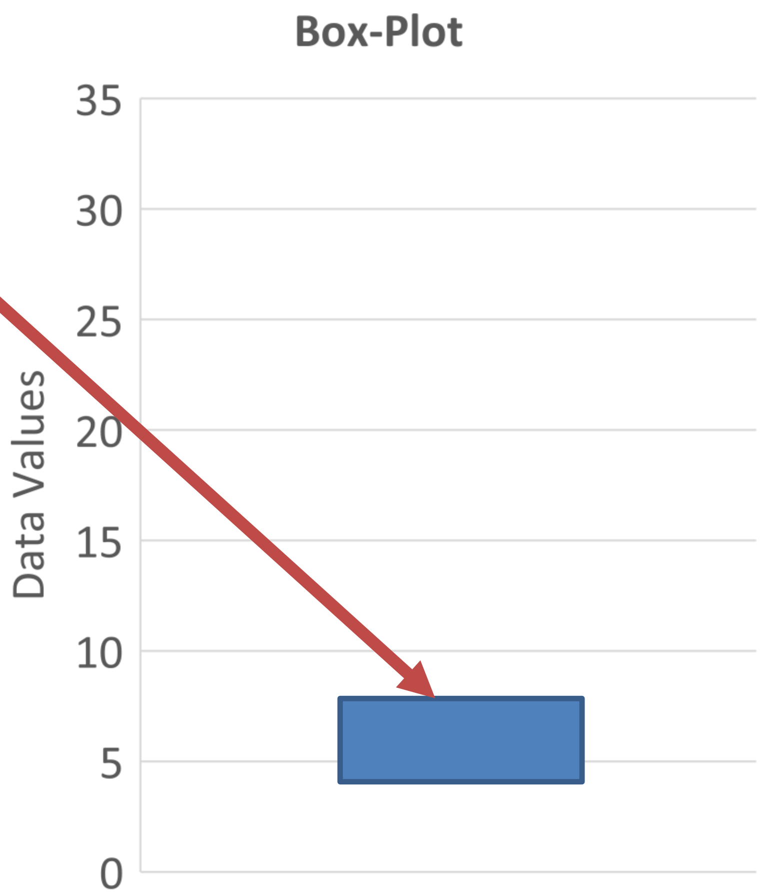
- Quartile

  Lower quartile (25%-Quantile): 4
  Upper quartile (75%-Quantile): 8

  Median (50%-Quantile): 4

- Interquartile range (IQR)

  IQR = 4



© ComTec 2019

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

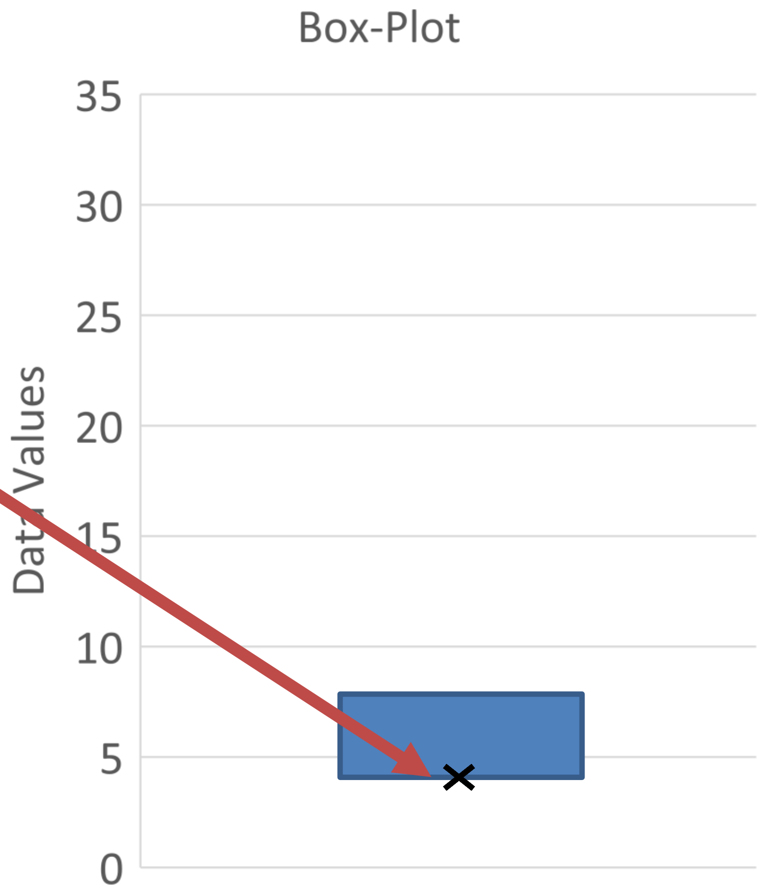
- Quartile

  Lower quartile (25%-Quantile): 4
  Upper quartile (75%-Quantile): 8

  Median (50%-Quantile): 4

- Interquartile range (IQR)

  IQR = 4



© ComTec 2019

Measures of Spread

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

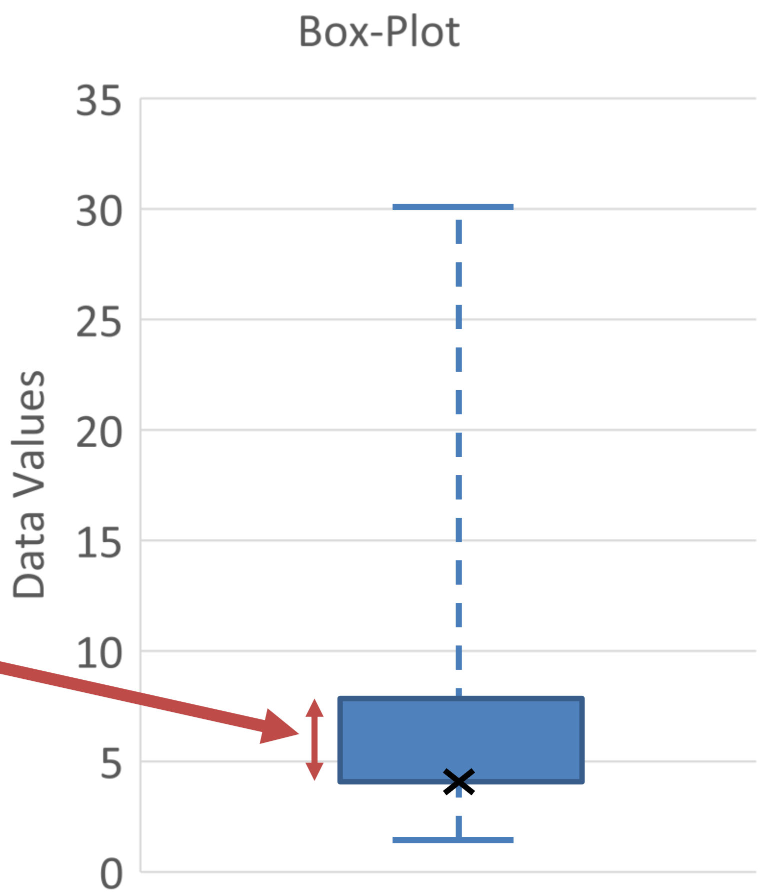
- Quartile

  Lower quartile (25%-Quantile): 4
  Upper quartile (75%-Quantile): 8

  Median (50%-Quantile): 4

- Interquartile range (IQR)

  IQR = 4



Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

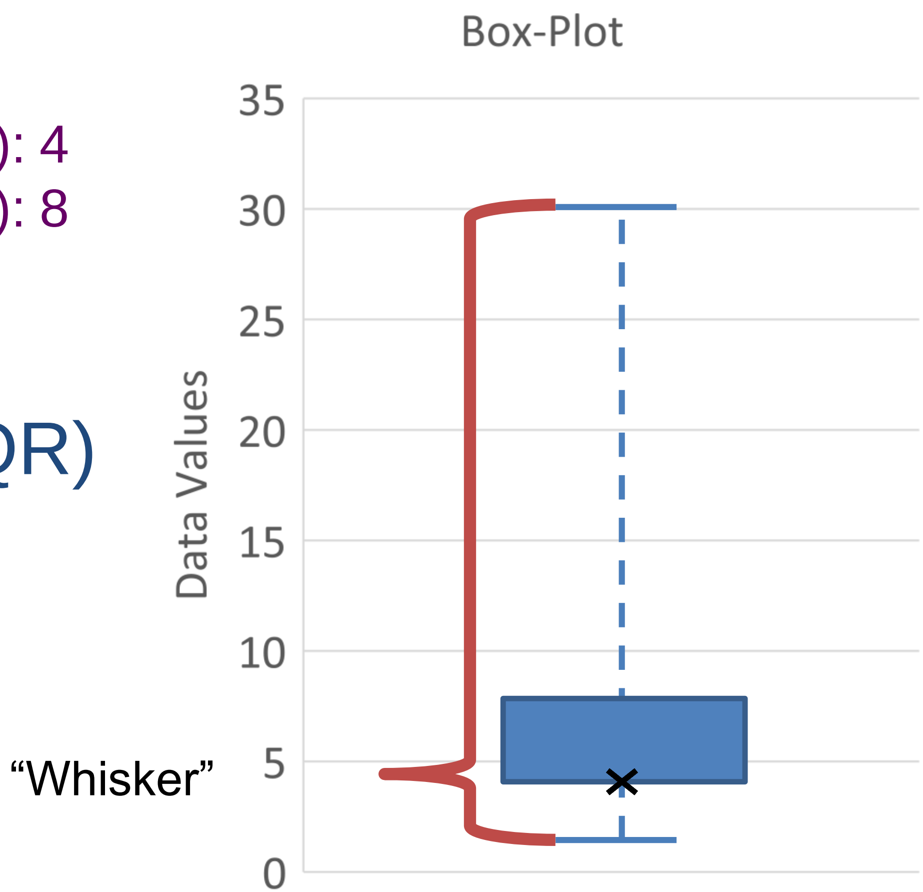
- Quartile

  Lower quartile (25%-Quantile): 4
  Upper quartile (75%-Quantile): 8

  Median (50%-Quantile): 4

- Interquartile range (IQR)

  IQR = 4



© ComTec 2019

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

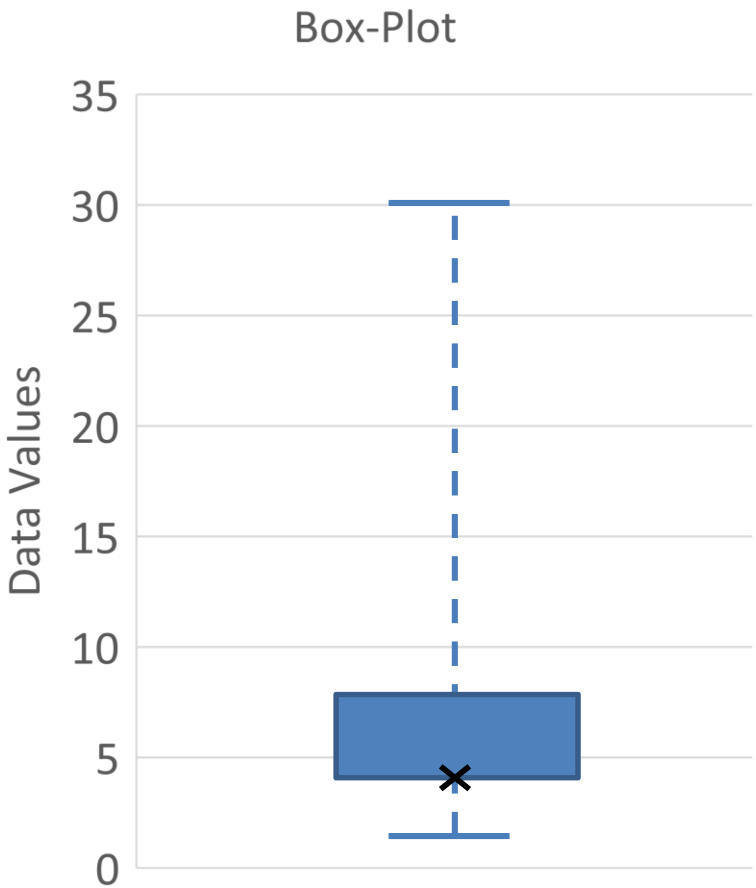
- Quartile

  Lower quartile (25%-Quantile): 4
  Upper quartile (75%-Quantile): 8

  Median (50%-Quantile): 4

- Interquartile range (IQR)

  IQR = 4



© ComTec 2019

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Standard Deviation

  The average distance from any point to the mean.

  Calculate the deviation from any point with respect to the mean:
  $(\bar{x} = 7.1)$

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Standard Deviation

  The average distance from any point to the mean.

  Calculate the deviation from any point with respect to the mean:
  (-3.1, 1.9, -3.1, -5.1, -3.1, 0.9, -3.1, -0.1, -6.1, -3.1, 0.9, 22.9)

  Square of the deviations:

Measures of Spread

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- ## Standard Deviation

The average distance from any point to the mean.

Calculate the deviation from any point with respect to the mean:
(-3.1, 1.9, -3.1, -5.1, -3.1, 0.9, -3.1, -0.1, -6.1, -3.1, 0.9, 22.9)

Square of the deviations:
(9.61, 3.61, 9.61, 26.01, 9.61, 0.81, 9.61, 0.01, 37.21, 9.61, 0.81, 524.41)

Calculate the mean and then the square root:

© ComTec 2019

Measures of Spread

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

- ## Standard Deviation

The average distance from any point to the mean.

Calculate the deviation from any point with respect to the mean:
(-3.1, 1.9, -3.1, -5.1, -3.1, 0.9, -3.1, -0.1, -6.1, -3.1, 0.9, 22.9)

Square of the deviations:
(9.61, 3.61, 9.61, 26.01, 9.61, 0.81, 9.61, 0.01, 37.21, 9.61, 0.81, 524.41)

Calculate the mean and then the square root:
Mean: 53.41

Standard deviation = 7.3

© ComTec 2019

Measures of Spread

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- ## Median Absolute Deviation (MAD)

  The average distance from any point to the Median.

  Calculate the deviation from any point with respect to the Median:
  $(x_{med} = 4)$

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \ldots, x_n)$

- Median Absolute Deviation (MAD)

  The average distance from any point to the Median.

  Calculate the deviation from any point with respect to the Median:
  (0, 5, 0, 2, 0, 4, 0, 3, 3, 0, 4, 26)

  Find the Median:

Measures of Spread

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

- ## Median Absolute Deviation (MAD)

  The average distance from any point to the Median.

  Calculate the deviation from any point with respect to the Median:
  (0, 5, 0, 2, 0, 4, 0, 3, 3, 0, 4, 26)

  Find the Median:
  0, 0, 0, 0, 0, 2, 3, 3, 4, 4, 5, 26

Measures of Spread

ComTec

Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30) = $(x_1, x_2, \dots, x_n)$

- ## Median Absolute Deviation (MAD)

The average distance from any point to the Median.

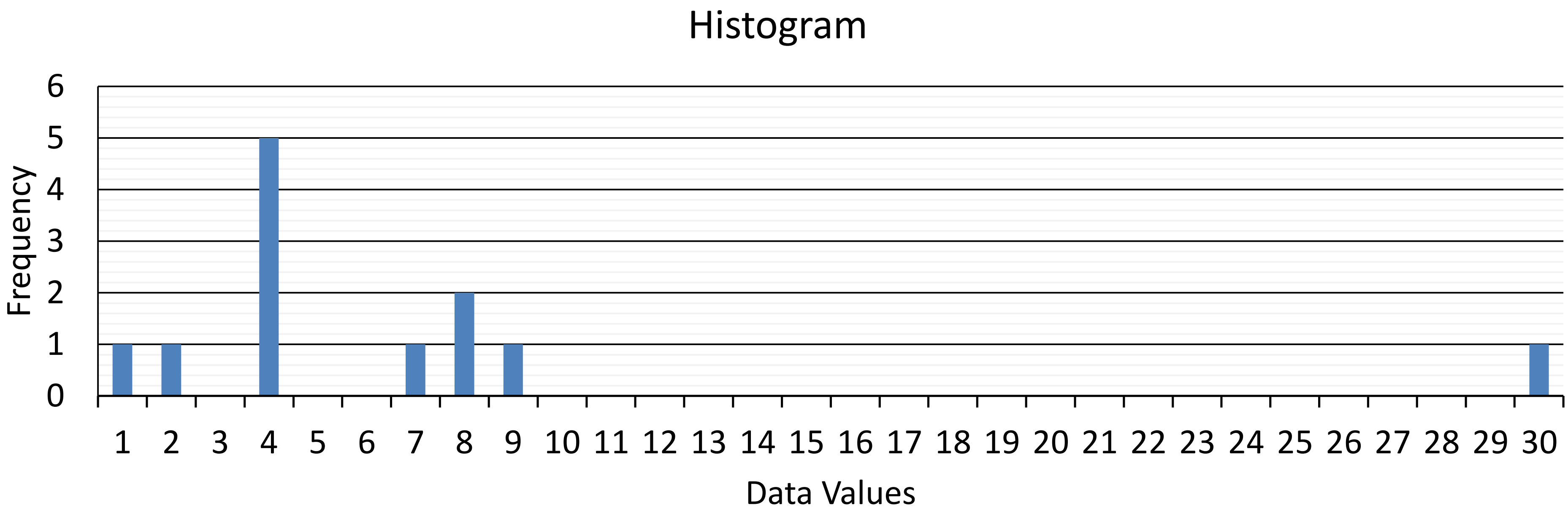
Calculate the deviation from any point with respect to the Median:
(0, 5, 0, 2, 0, 4, 0, 3, 3, 0, 4, 26)

Find the Median:
0, 0, 0, 0, 0, 2, 3, 3, 4, 4, 5, 26

MAD = 2.5

© ComTec 2019

Measures of Spread - Conclusion



Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30)

Range: 29

Lower Quartile: 4

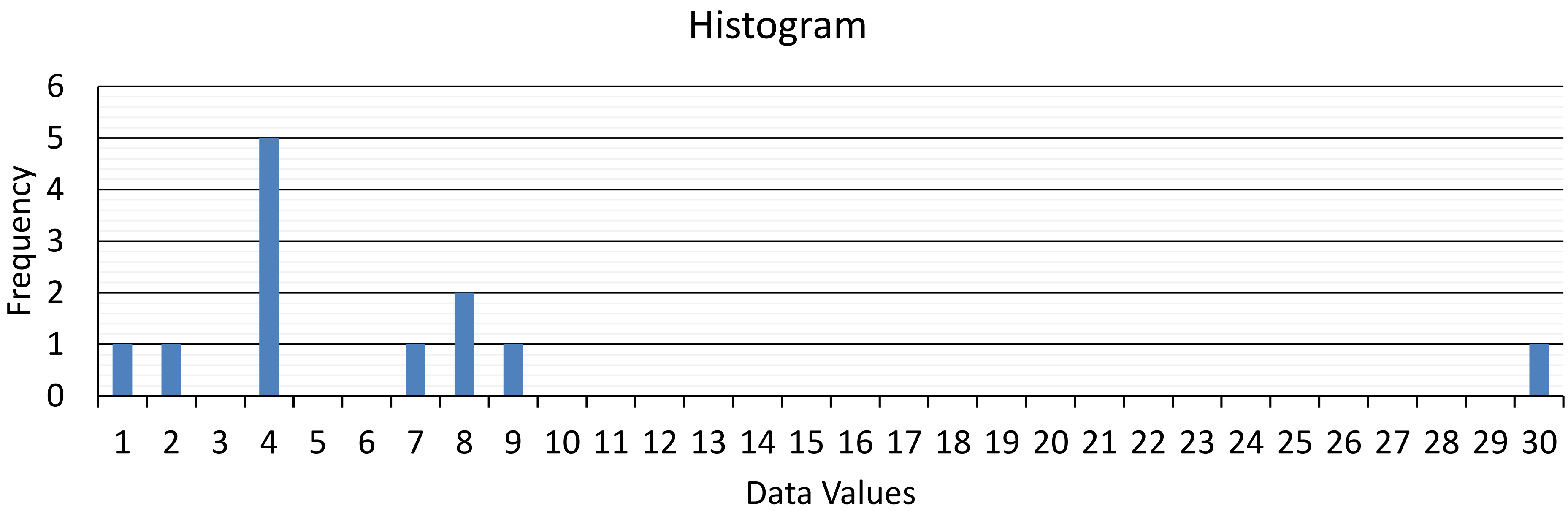
Upper Quartile: 8

Interquartile Range : 4

Standard Deviation: 7.3

© ComTec 2019

Measures of Spread - Conclusion



Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30)

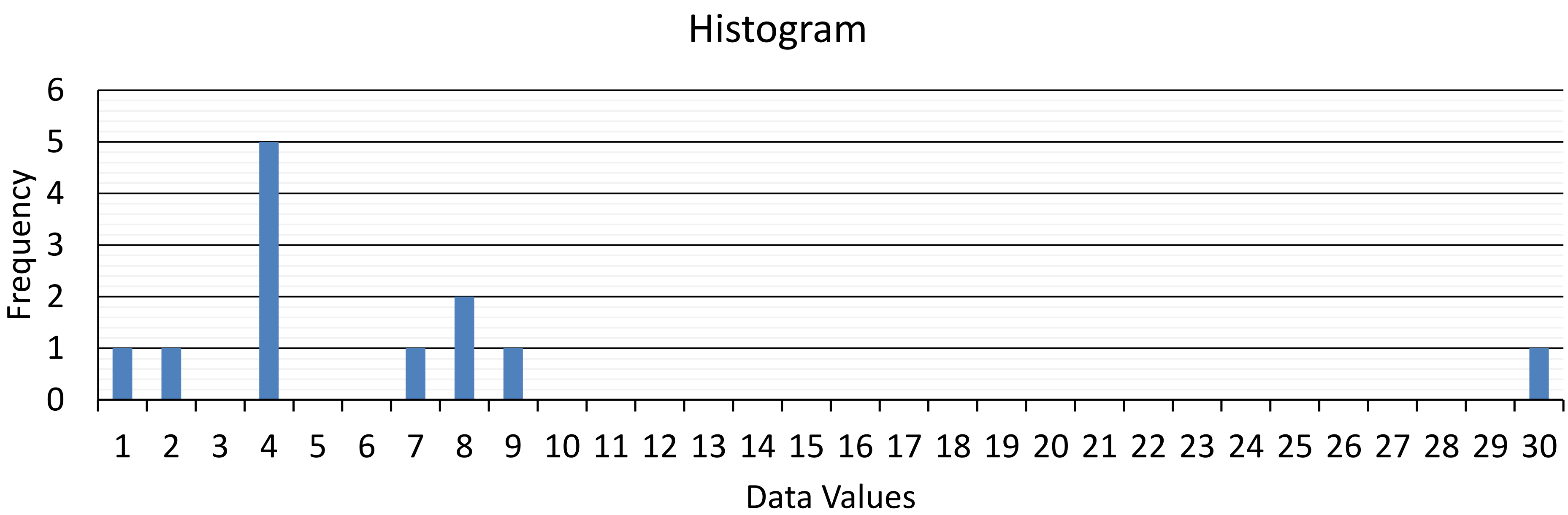
Range: 29
Lower Quartile: 4
Upper Quartile: 8
Interquartile Range : 4
Standard Deviation: 7.3

Interquartile Range (IQR) is usually 2-3 times the Standard Deviation (SD).

If SD is large compared to IQR, beware of outliers!

© ComTec 2019

Measures of Spread - Conclusion



Data: (4, 9, 4, 2, 4, 8, 4, 7, 1, 4, 8, 30)

Mean: 7.1
Median: 4
Mode: 4

Range: 29
Lower Quartile: 4
Upper Quartile: 8
Interquartile Range : 4
Standard Deviation: 7.3
Median Absolute Deviation: 2.5

What if *Mean* = *Median* = *Mode*?

Data is Gaussian-distributed

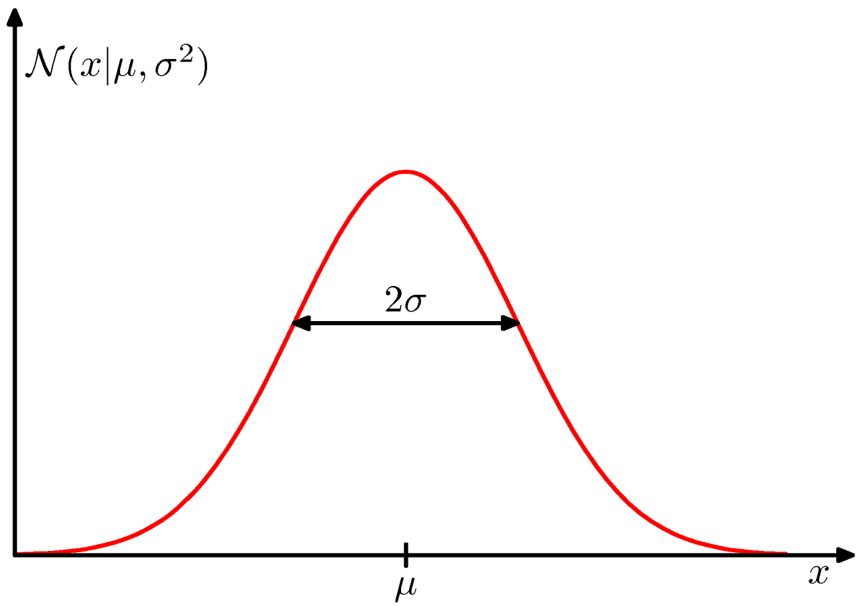
If Mean = Mode = Median, then the data is Gaussian-distributed:

$$\mathcal{N}(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Mean (expectation value): μ

Variance: σ^2

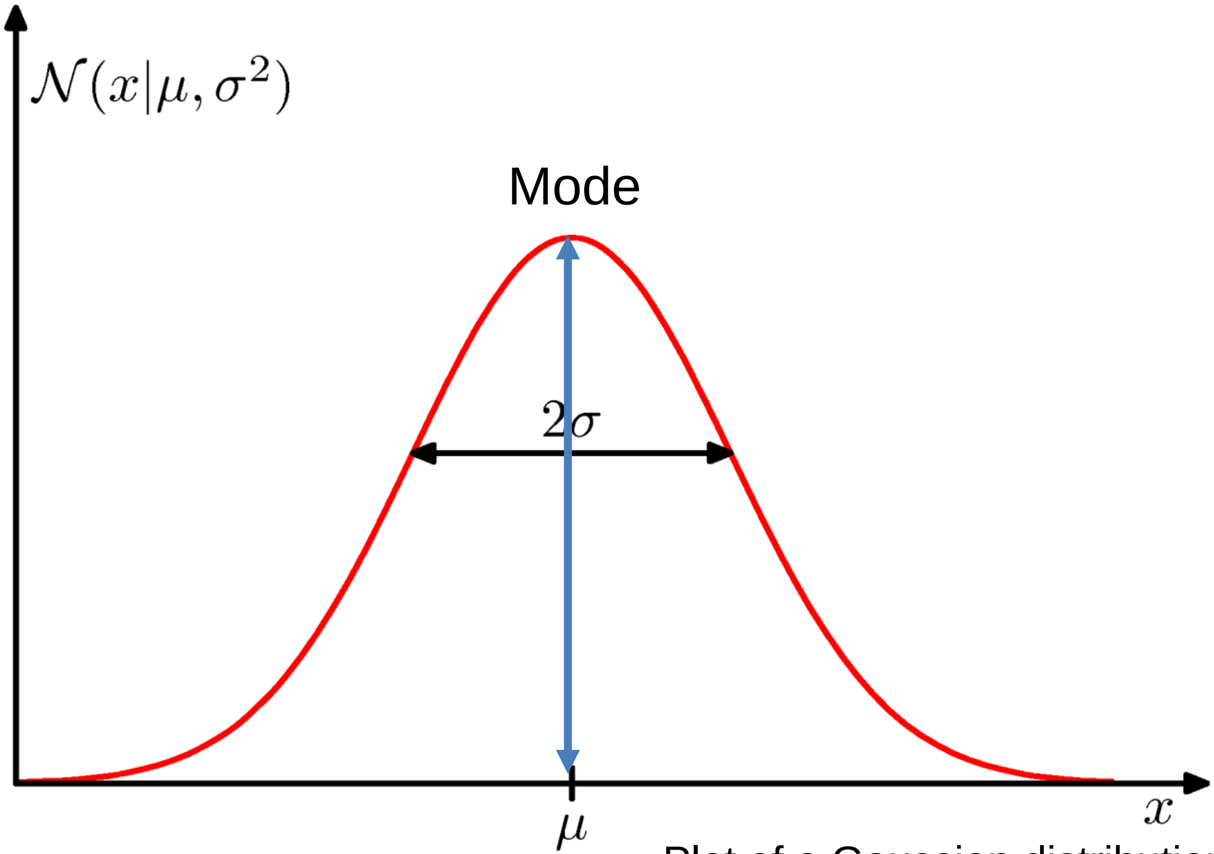
Standard deviation $\sigma = \sqrt{\sigma^2}$



Plot of a Gaussian distribution [5]

Data is Gaussian-distributed

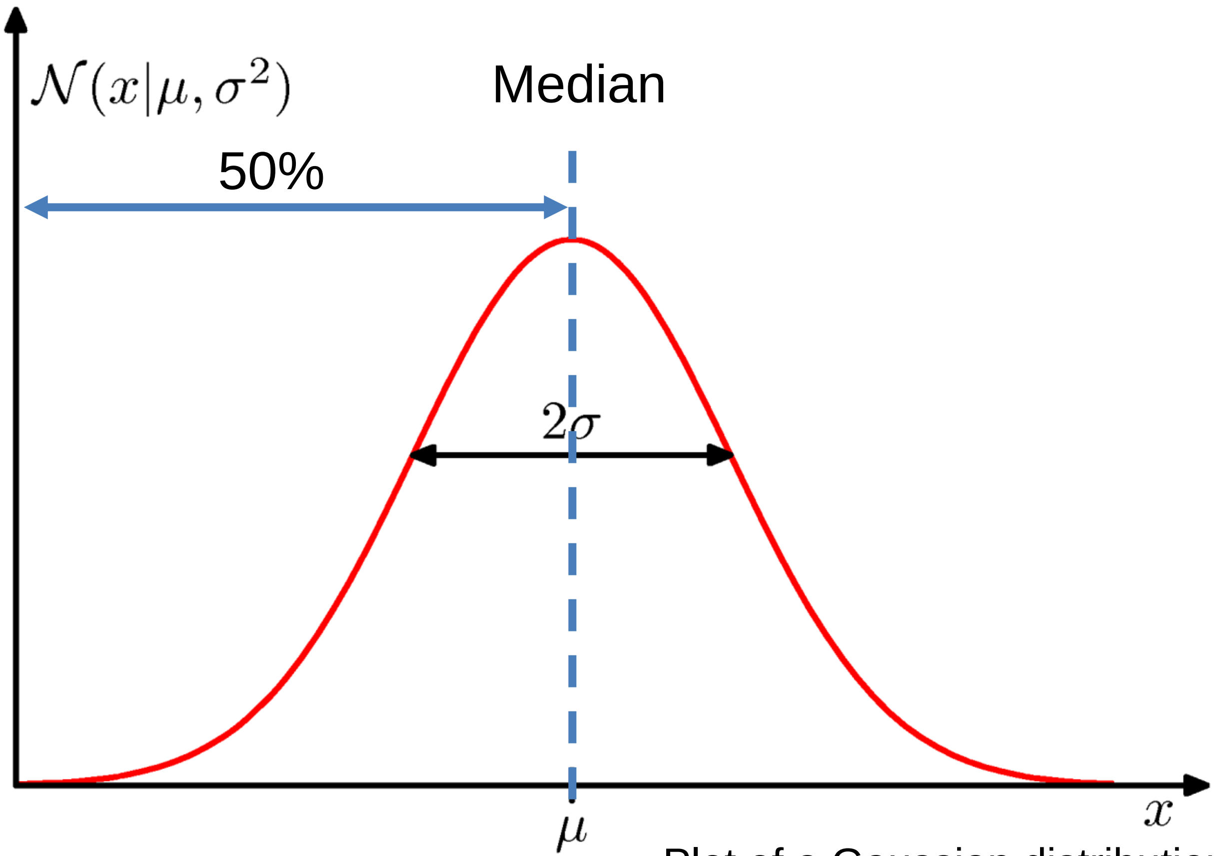
If Mean = Mode = Median, then the data is Gaussian-distributed:



Plot of a Gaussian distribution [5]

Data is Gaussian-distributed

If Mean = Mode = Median, then the data is Gaussian-distributed:

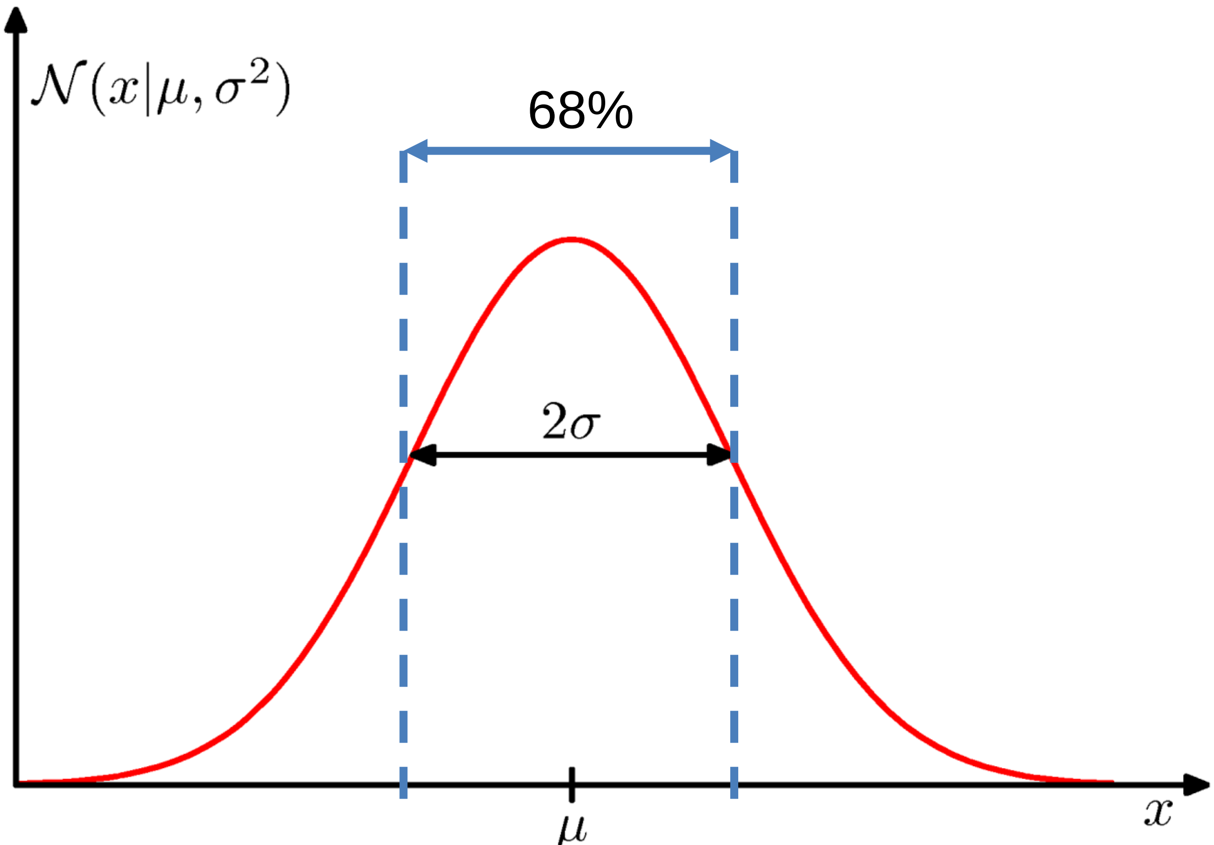


Plot of a Gaussian distribution [5]

© ComTec 2019

Data is Gaussian-distributed

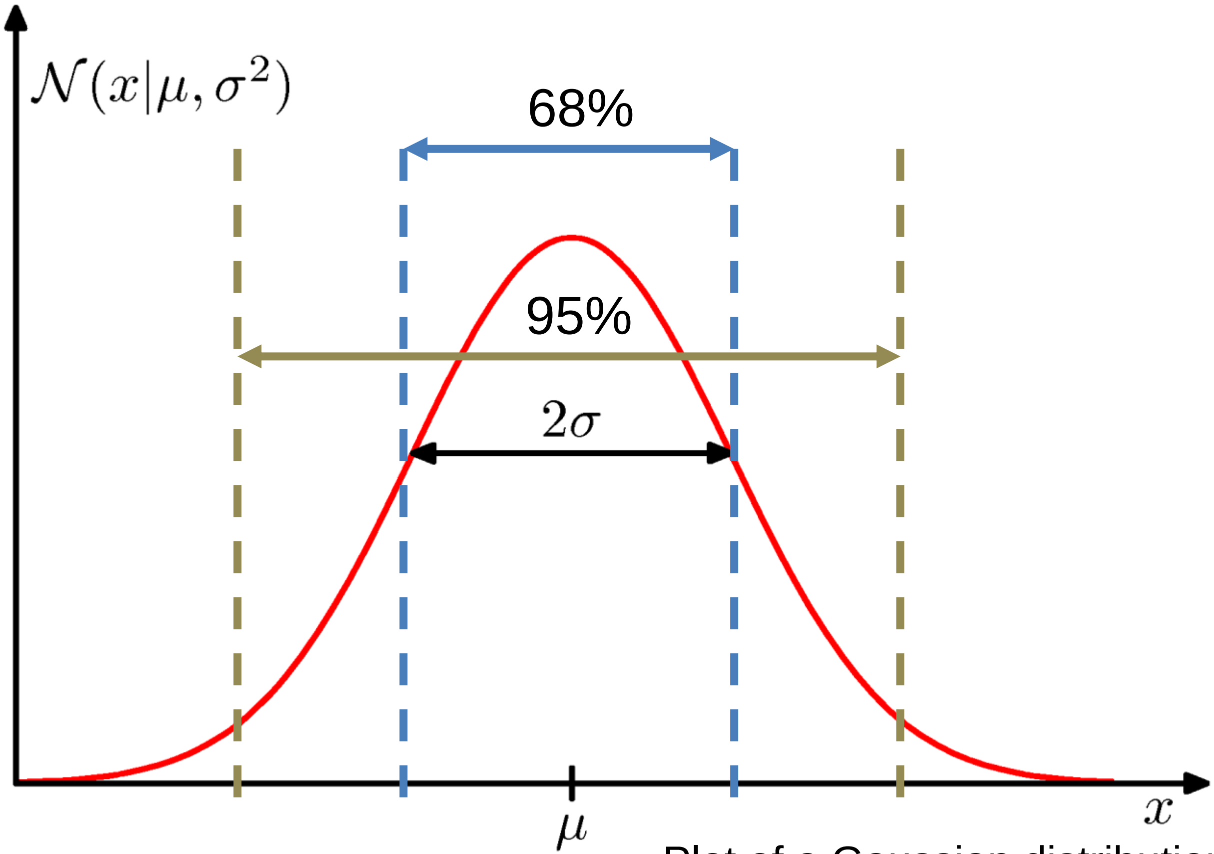
If Mean = Mode = Median, then the data is Gaussian-distributed:



Plot of a Gaussian distribution [5]

Data is Gaussian-distributed

If Mean = Mode = Median, then the data is Gaussian-distributed:

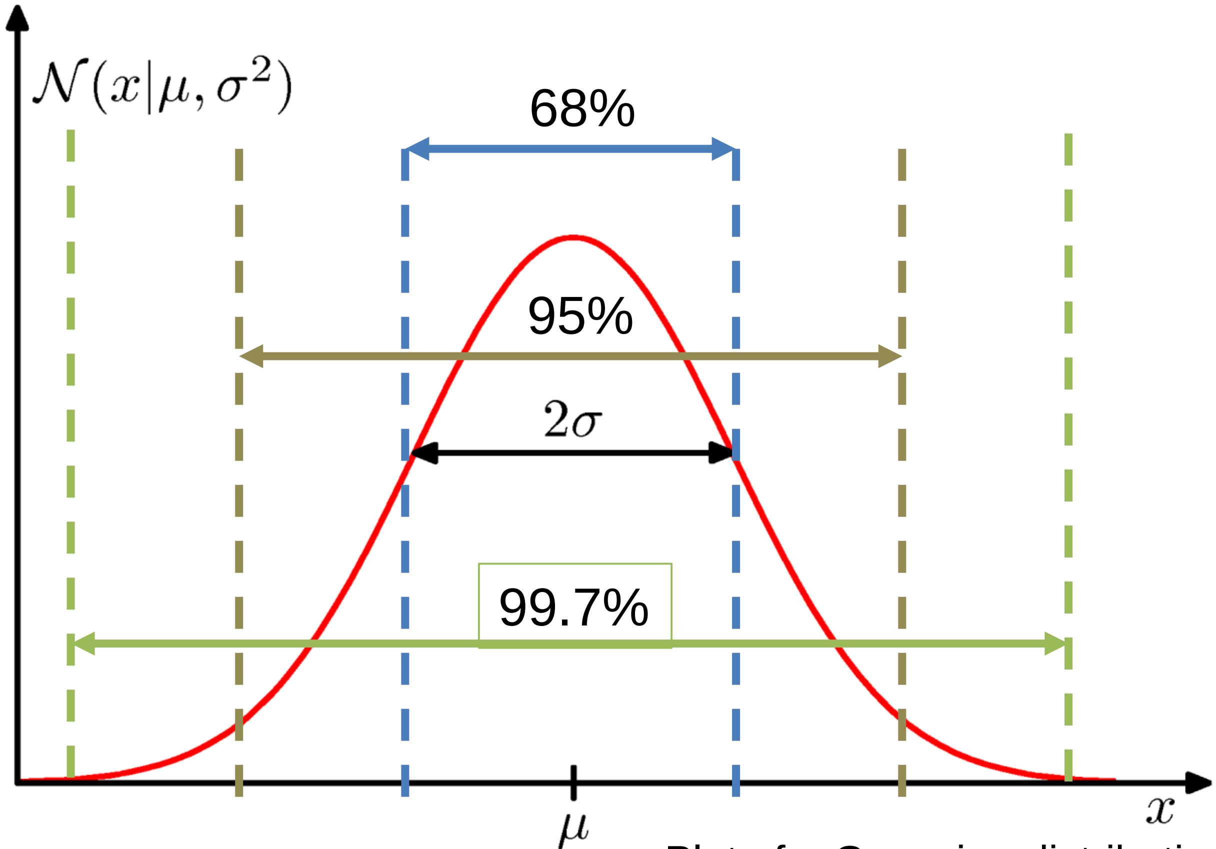


Plot of a Gaussian distribution [5]

© ComTec 2019

Data is Gaussian-distributed

If Mean = Mode = Median, then the data is Gaussian-distributed:



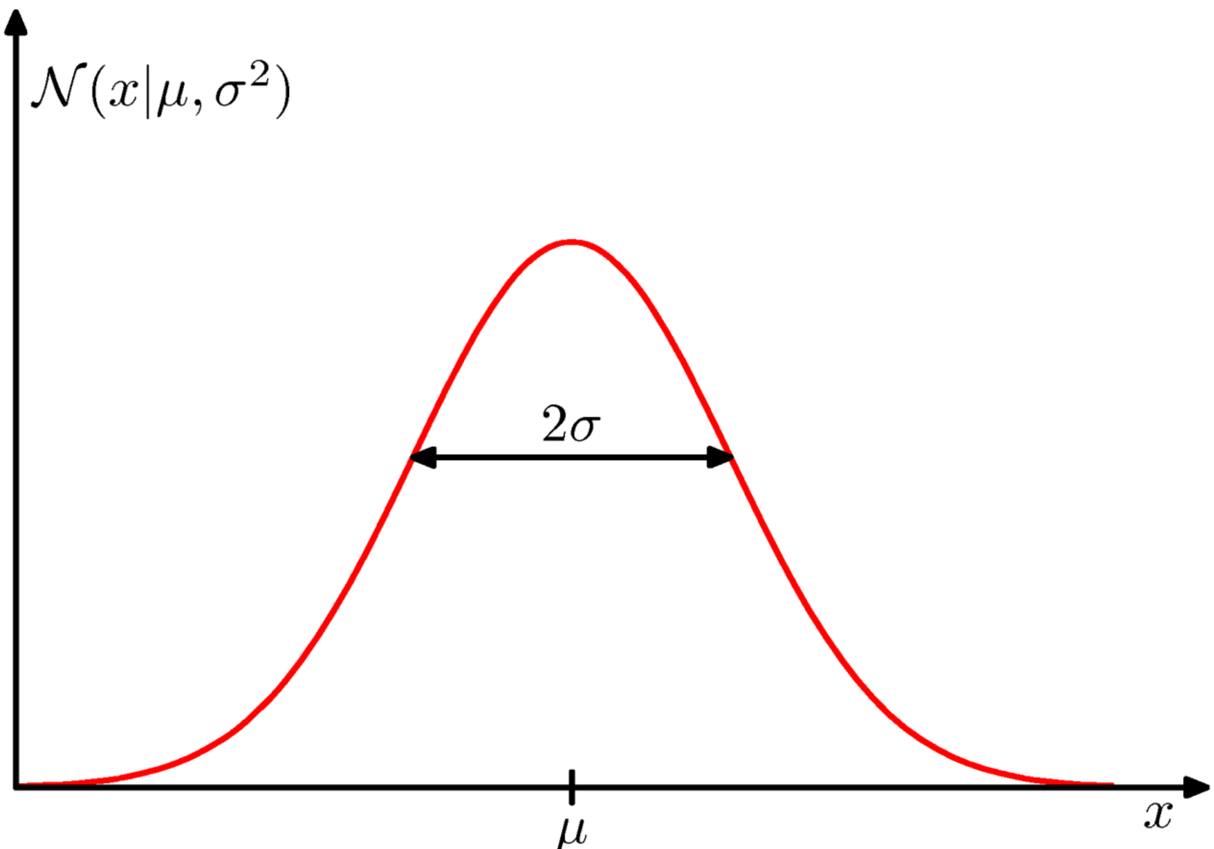
Plot of a Gaussian distribution [5]

© ComTec 2019

Skewness

ComTec

If Mean = Mode = Median, then the data is Gaussian-distributed and has a **skewness of zero**.

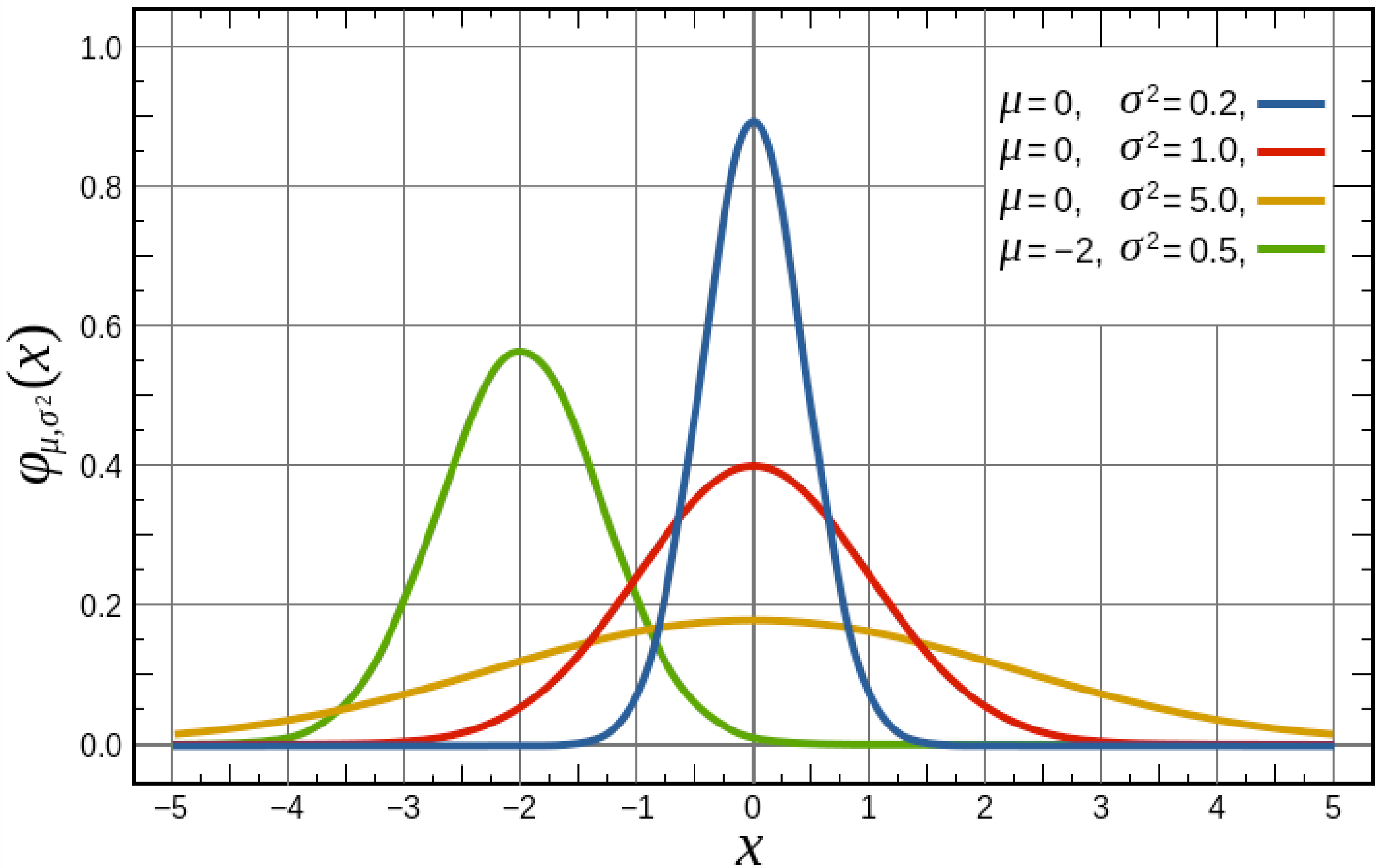


Plot of a Gaussian distribution [5]

© ComTec 2019

Skewness

ComTec



Zero Skewness

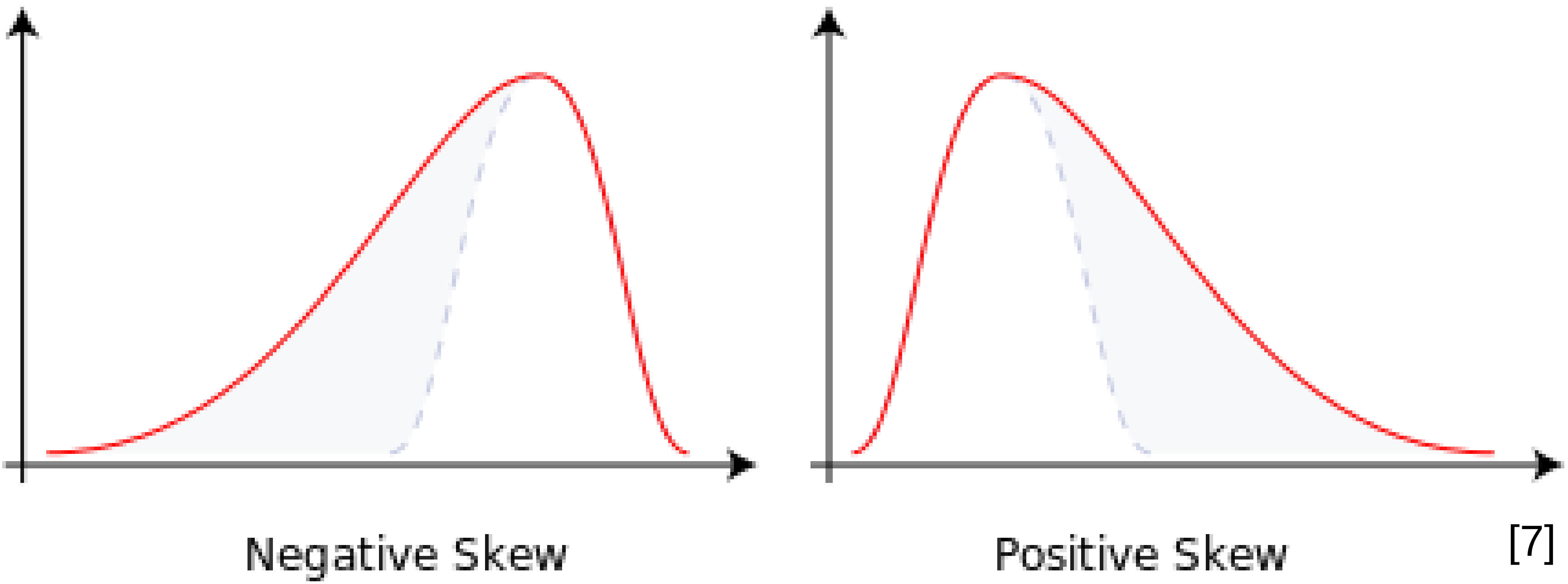
[6]

Skewness

Symmetric: $\bar{x} \approx x_{med} \approx x_{mod}$

Positive Skew: $\bar{x} > x_{med} > x_{mod}$

Negative Skew: $\bar{x} < x_{med} < x_{mod}$



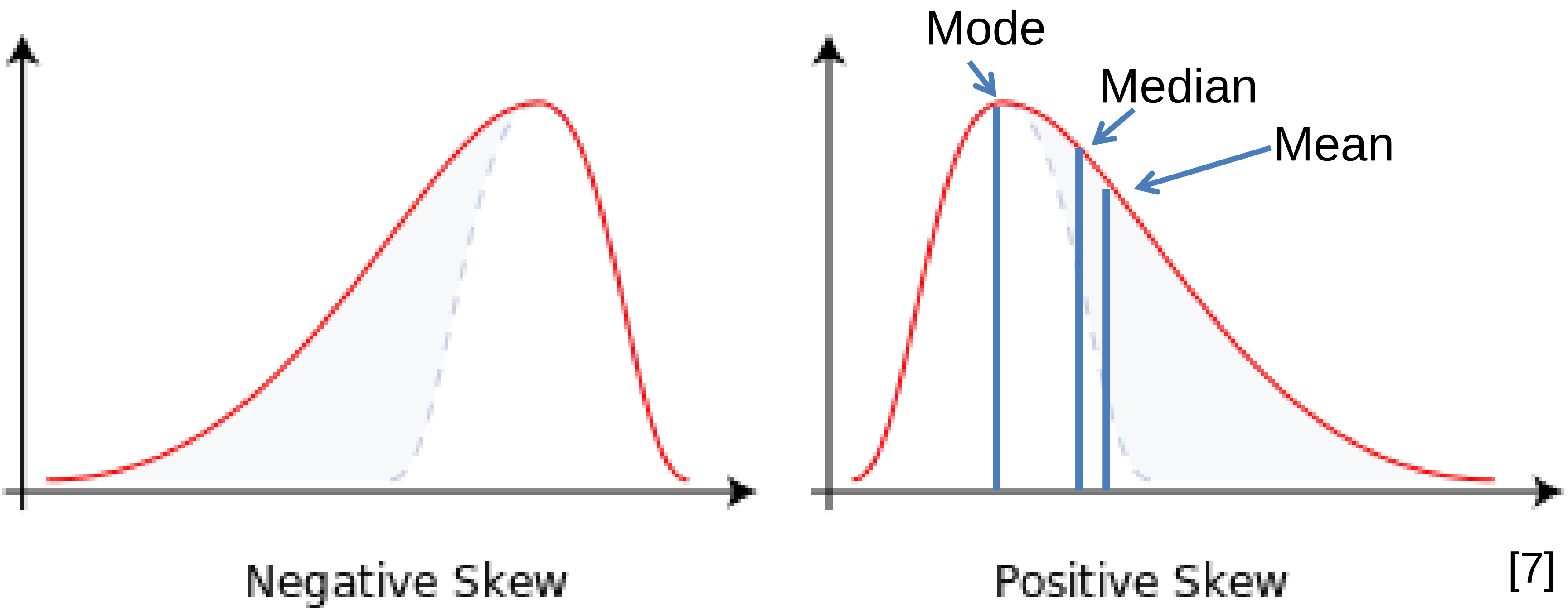
[7]

Skewness

Symmetric: $\bar{x} \approx x_{med} \approx x_{mod}$

Positive Skew: $\bar{x} > x_{med} > x_{mod}$

Negative Skew: $\bar{x} < x_{med} < x_{mod}$



[7]

Skewness

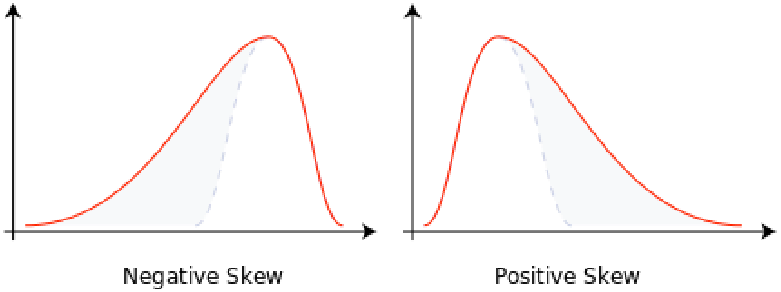
- Pearson's coefficient of skewness

$$\frac{\bar{x} - x_{mod}}{standard\ deviation}$$

or

$$3 * \frac{\bar{x} - x_{med}}{standard\ deviation}$$

If the coefficient is positive, the distribution is positively skewed, usually.



[7]

Skewness

ComTec

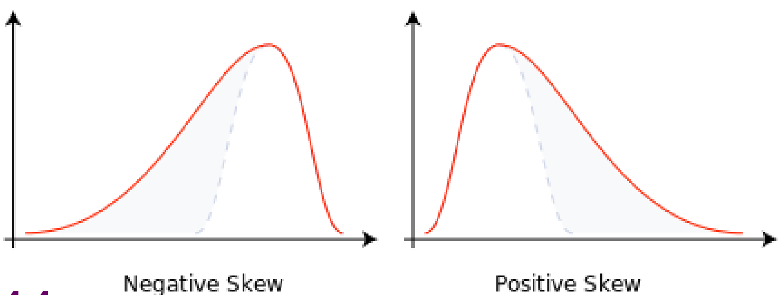
- Pearson's coefficient of skewness

$$\frac{\bar{x} - x_{mod}}{standard\ deviation}$$

or

$$3 * \frac{\bar{x} - x_{med}}{standard\ deviation}$$

If the coefficient is positive, the distribution is positively skewed, usually.



[7]

Pearson's coefficient of skewness: 0.44
=> Positive Skew

Feature Extraction

Sensor Data

Segmentation → Feature Extraction → Model creation → Model evaluation

Segmentation (JH)

Features (JH)

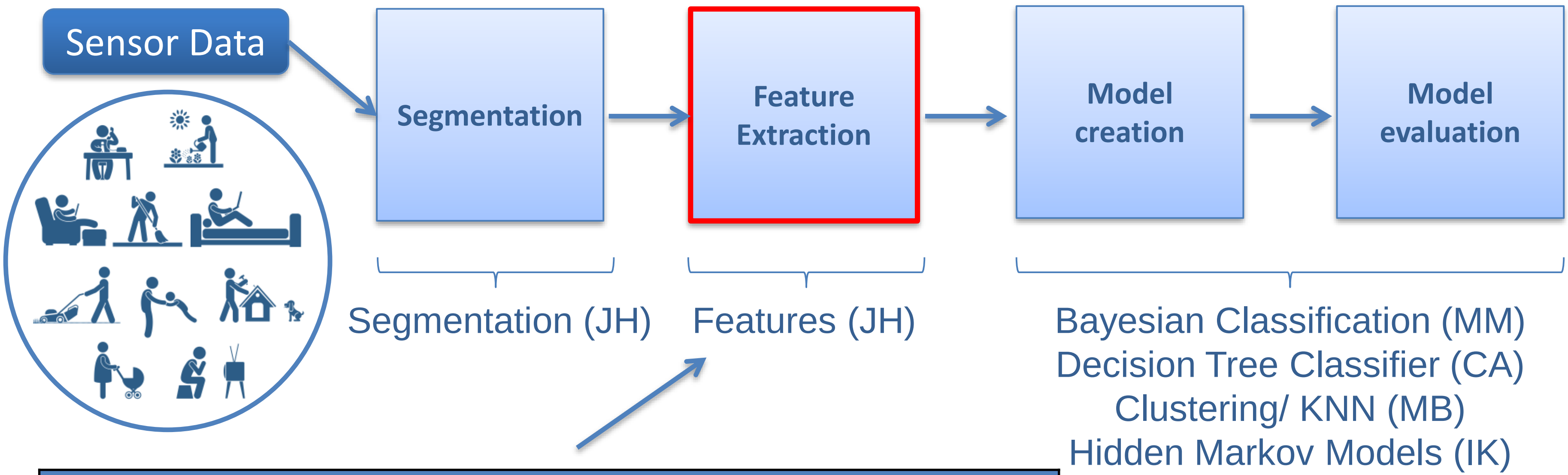
Bayesian Classification (MM)
Decision Tree Classifier (CA)
Clustering/ KNN (MB)
Hidden Markov Models (IK)

Statistical Features	
Mean	Averaged Derivatives
Median	Skewness
Mode	Zero Crossing Rate
Standard Deviation	Mean Crossing Rate
Variance	Pairwise Correlation
Covariance	Time between peaks
Root Mean square	Range Interquatile Range
Median Absolute Deviation (MAD)	Etc...

© ComTec 2019

Feature Extraction

ComTec



Statistical Features	
Mean	Averaged Derivatives
Median	Skewness
Mode	Zero Crossing Rate
Standard Deviation	Mean Crossing Rate
Variance	Pairwise Correlation
Covariance	Time between peaks
Root Mean square	Range Interquatile Range
Median Absolute Deviation (MAD)	Etc...

15 Features *
5 Sensors
= 75

© ComTec 2019

Why do we need feature selection?

- Facilitating data visualization
- Data understanding
- Reducing the measurement and storage requirements
- Reducing training and utilization times
- Defying the curse of dimensionality to improve prediction performance

[8,9,11]

Feature Selection Algorithms

- Supervised
- Unsupervised
- Semi-supervised

- Feature subset selection (FS)
- Dimensionality reduction (DR)

- Filter
- Wrapper
- Embedded

- Hybrid Methods

[8,9,11]

© ComTec 2019

Feature Selection Algorithms

- Supervised
- Unsupervised
- Semi-supervised

- Feature subset selection (FS)
- Dimensionality reduction (DR)

- Filter
- Wrapper
- Embedded
- Hybrid Methods

[8,9,11]

© ComTec 2019

Feature Selection Algorithms: **Filter**

- Ranking based on the ability to discriminate between different classes / clusters
- Usage independent of data modelling algorithm

Univariate feature filters: rank individual features or

Multivariate filters: evaluate entire feature subsets

Examples: Information Gain, Correlation, Fisher score, etc.

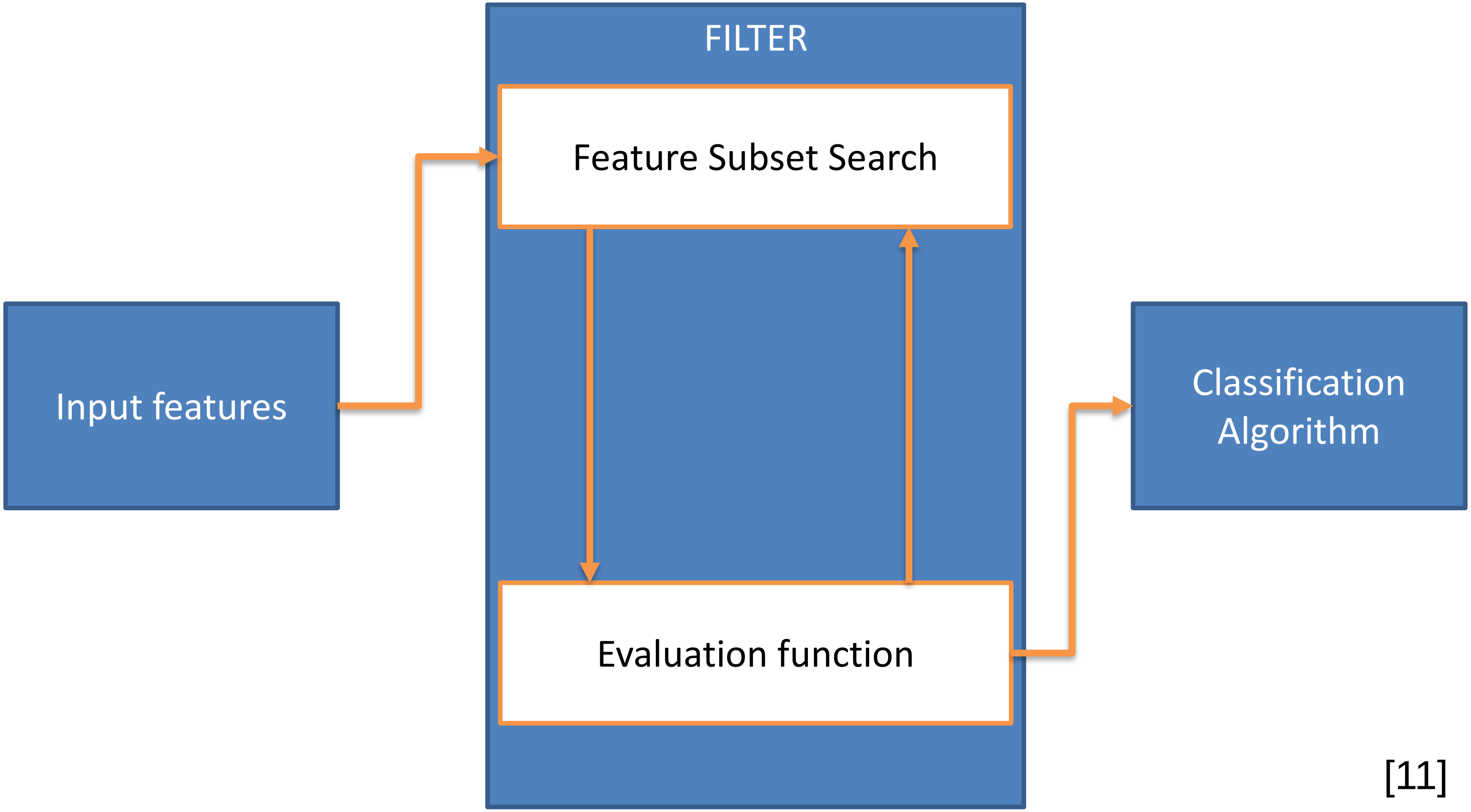
Dependent on: classification, regression or clustering

[8,9,11]

© ComTec 2019

Feature Selection Algorithms: **Filter**

ComTec



[11]

© ComTec 2019

Feature Selection Algorithms: **Wrapper**

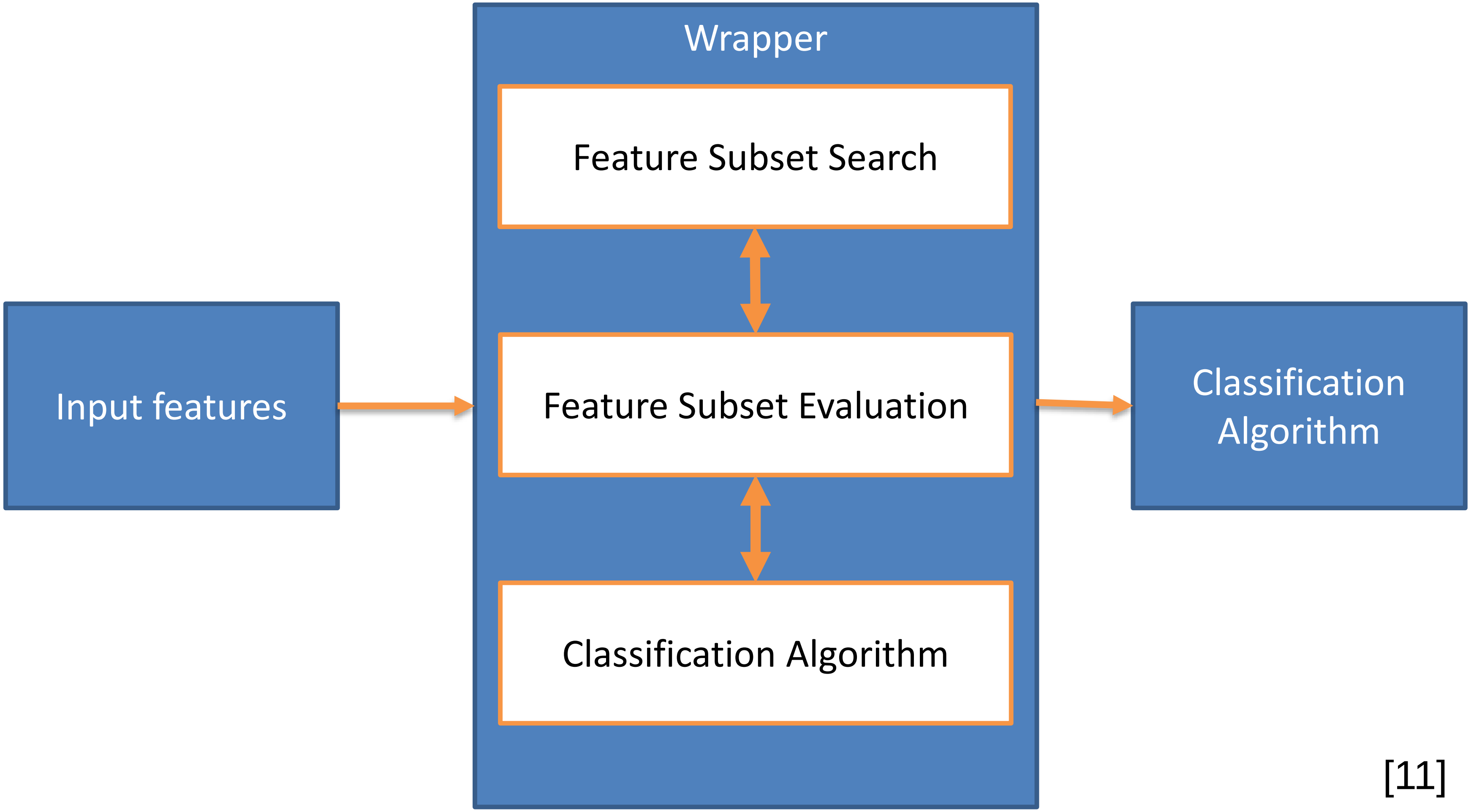
“The feedback of the classification algorithm is used to measure the quality of selected features” [9]

1. Find feature subsets (Forward selection, backward elimination, random,…)
2. Evaluate feature subsets with chosen algorithm
3. Continue step 1. and 2. until feature subset is optimized

- Dependent on the performance of the chosen algorithm
  - Independent validation sample
  - Another modelling algorithm for classification / clustering / regression

[11]

Feature Selection Algorithms: **Wrapper**



[11]

Feature Selection Algorithms: **Embedded Method**

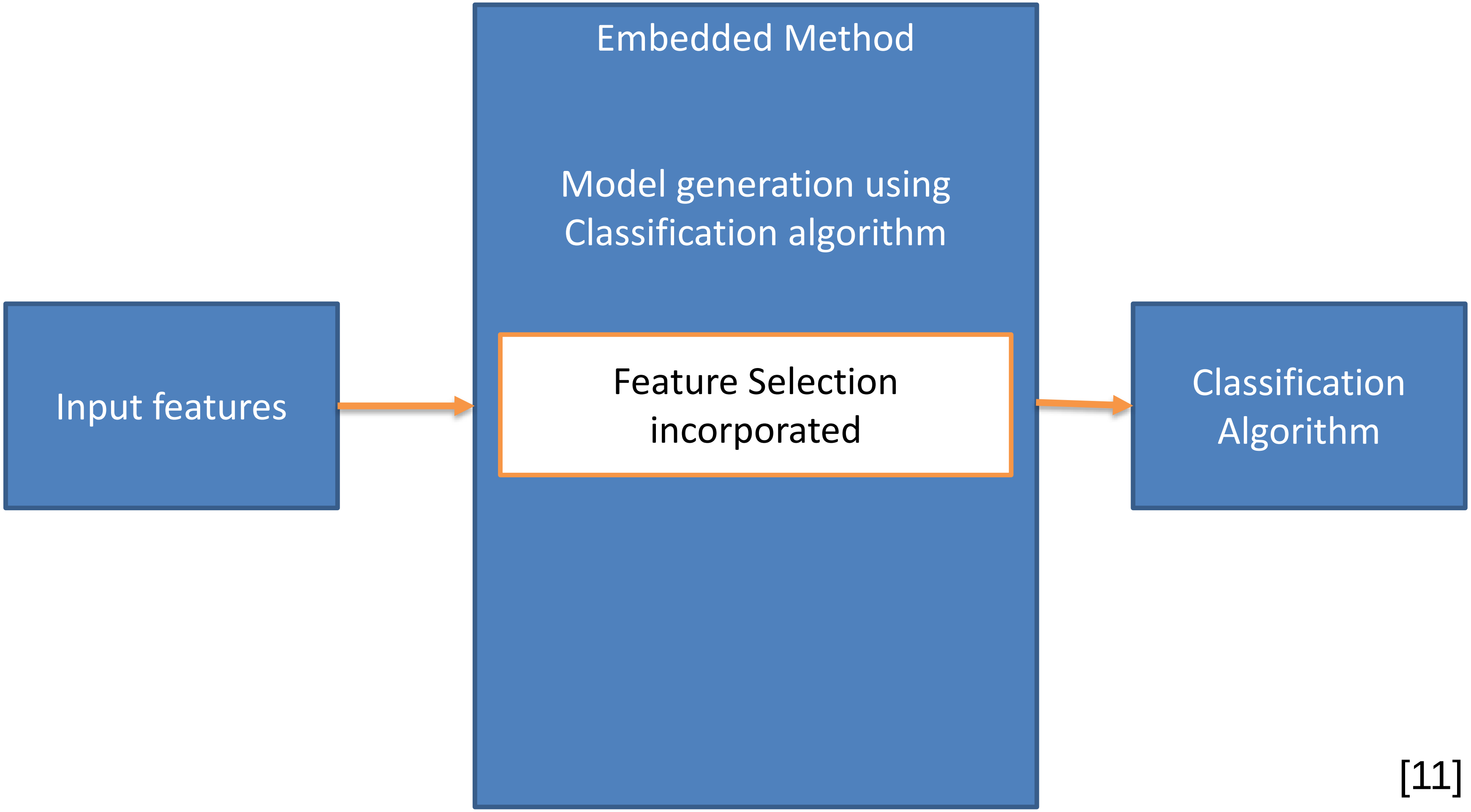
ComTec

- Feature selection is integrated in the classification algorithm itself
  1. Initialize feature subset (Forward selection, backward elimination, random,…)
  2. Evaluate the subset using independent measure, if criteria is fulfilled go to 3., else find new subset
  3. Evaluate with chosen algorithm

- Examples: Decision Tree, SVM, etc.

[11]

Feature Selection Algorithms: **Embedded Method**



[11]

© ComTec 2019

Feature Selection Algorithms:
Hybrid method

- Combination of filter and wrapper algorithms:
  1. Filter algorithm to find subsets
  2. Evaluation of the best subset with wrapper algorithm

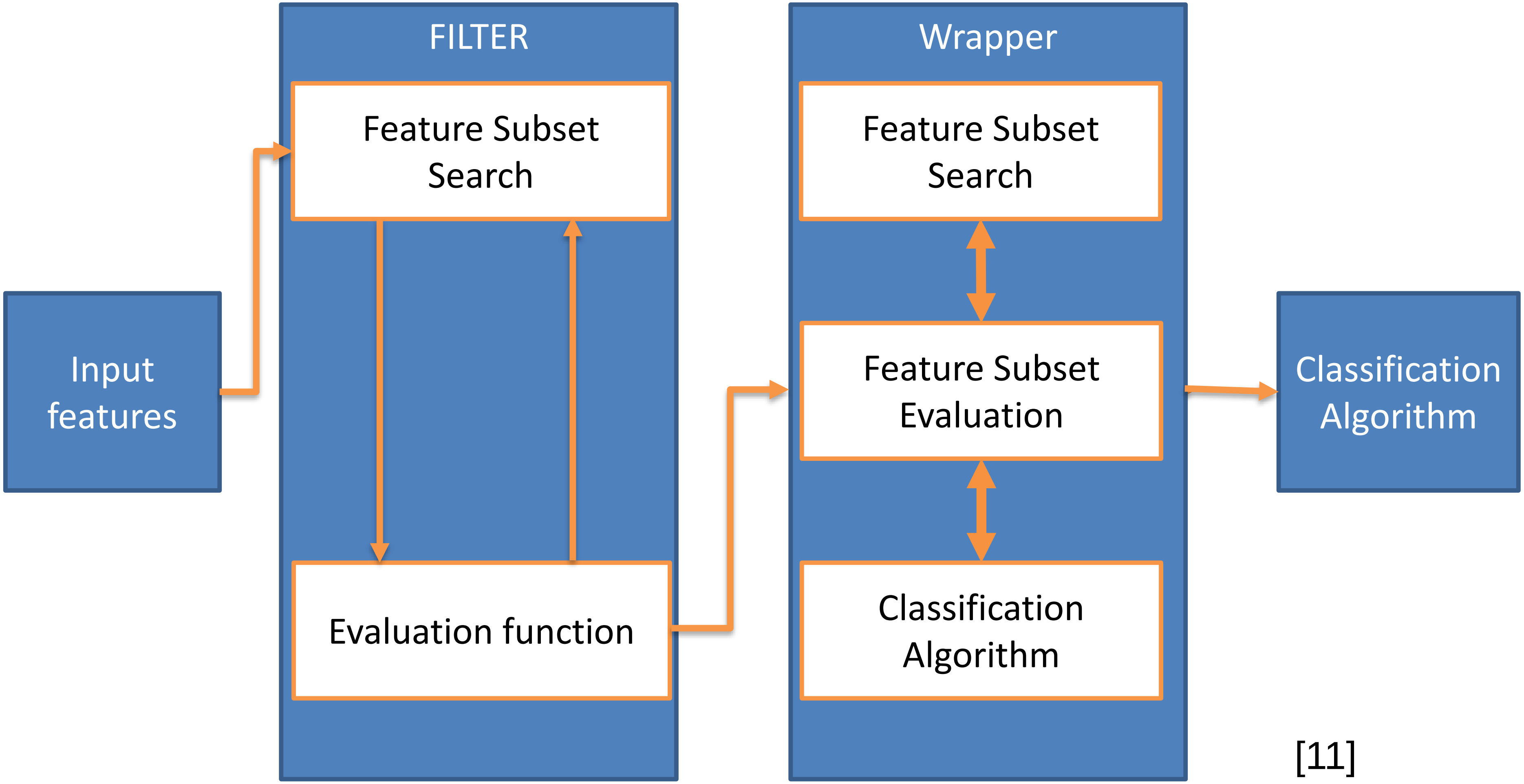
- Dependent on the performance of the chosen algorithm for wrapper
- Ranking based on the ability to discriminate between different classes / clusters

[11]

© ComTec 2019

Feature Selection Algorithms: **Hybrid method**

ComTec



[11]

Feature Selection: **Comparison**

Feature selection method	Advantages	Disadvantages
Filter	+ Simple approach + Computationally less expensive	– No interaction with classifier
Wrapper	+ Considers feature dependencies + High-quality features	– Computationally expensive – Overfitting
Embedded	+ Combines advantages of filter and wrapper	– Specific to chosen algorithm
Hybrid	+ High accuracies + Short evaluation time	– Dependency on algorithm

Feature Extraction

Sensor Data

Segmentation → Feature Extraction → Model creation → Model evaluation

Segmentation (JH)

Features (JH)

Bayesian Classification (MM)
Decision Tree Classifier (CA)
Clustering/ KNN (MB)
Hidden Markov Models (IK)

– Time Series
– Time intervals
– Segmentation Algorithms
    • Sliding Window Algorithm

- Calculate Features
- Feature Selection:
    • Filter
    • Wrapper
    • Embedded methods
    • Hybrid methods

© ComTec 2019

References

[1] TimeAndDate: Weather in Kassel, last access 2016/04/19, http://www.timeanddate.com/weather/germany/kassel

[2] P. S. P. Cowpertwait and A. V. Metcalfe, *Introductory Time Series with R*. New York, NY, USA: Springer Science+Business Media, 2009.

[3] E. Keogh, S. Chu, D. Hart and M. Pazzani, "An online algorithm for segmenting time series," in *Proceedings of IEEE Int. Conf. on Data Mining*, San Jose, CA, 2001, pp. 289-296.

[4] G. Upton and I. Cook, *A Dictionary of Statistics*, 2nd ed. Oxford University Press, 2008.

[5] C. J. Bishop, *Pattern recognition and machine learning*, 8. corr. print. ed. (Information science and statistics). New York: Springer-Verlag, 2009.

[6] By Inductiveload - self-made, Mathematica, Inkscape, Public Domain, https://commons.wikimedia.org/w/index.php?curid=3817954

[7] By Rodolfo Hermans (Godot) at en.wikipedia. - Own work; transferred from en.wikipedia by Rodolfo Hermans (Godot)., CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=4567445

References

[8] I. Guyon and A. Elisseeff, “An Introduction to Variable and Feature Selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[9] A. Jovic, K. Brkic, and N. Bogunovic, “A review of feature selection methods with applications,” in *38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, pp. 1200–1205, 2015.

[10] M. S. Raza and U. Qamar, *Understanding and Using Rough Set Based Feature Selection: Concepts, Techniques and Applications*. Springer, 2017.

[11] D. Mladenić, “Feature Selection for Dimensionality Reduction,” in *Subspace, Latent Structure and Feature Selection*, vol. 3940, C. Saunders, M. Grobelnik, S. Gunn, and J. Shawe-Taylor, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 84–102.

© ComTec 2019