Time series analysis: fundamentals

Eliana Pastor
Politecnico di Torino

Time Series Definition

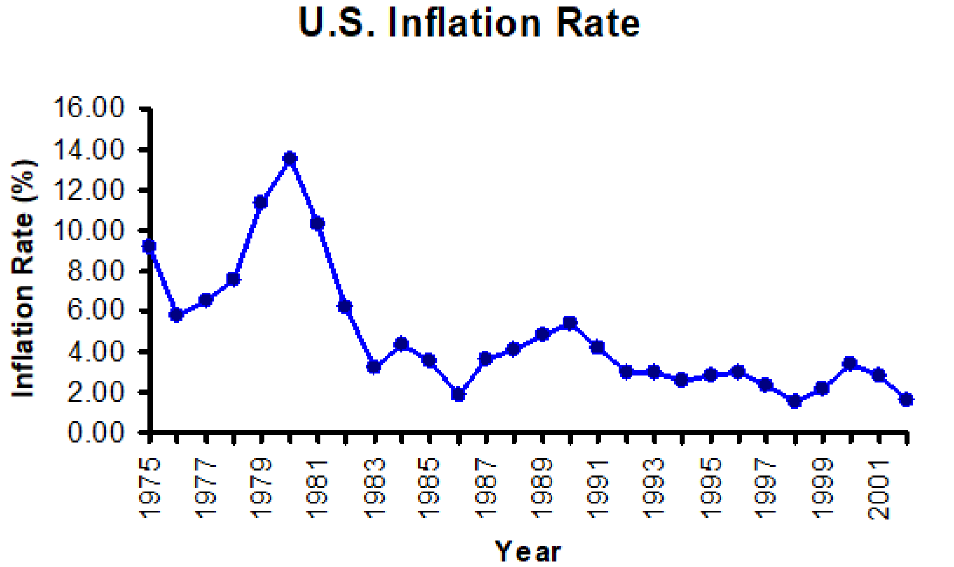
- A time series is a **sequential** set of data points, measured typically over successive times

- A time series containing values of a single variable is termed as **univariate**. But if values of more than one variable are considered, it is termed as **multivariate**

- A time series can be **continuous** or **discrete**.
  - **continuous** time series: observations are measured at every instance of time
    - e.g., temperature readings, flow of a river
  - **discrete** time series: observations are measured at discrete points of time
    - e.g., city population, production of a company, exchange rates

A time series plot

- A time-series plot (time plot) is a two-dimensional plot of time series data
  - the vertical axis measures the **variable of interest**
  - the horizontal axis corresponds to the **time** periods



Time series analysis

- Two main kinds of analysis can be performed on time series
  - Characterizing the nature of the phenomenon represented by the sequence of observations,
    - Time series components
  - Classification task vs. forecasting future values
    - Classification
      - speech recognition
      - classification of machine failures
      - ...
    - Forecasting
      - energy demand prediction
      - weather forecasting
      - traffic prediction
      - ...

Time series characterization

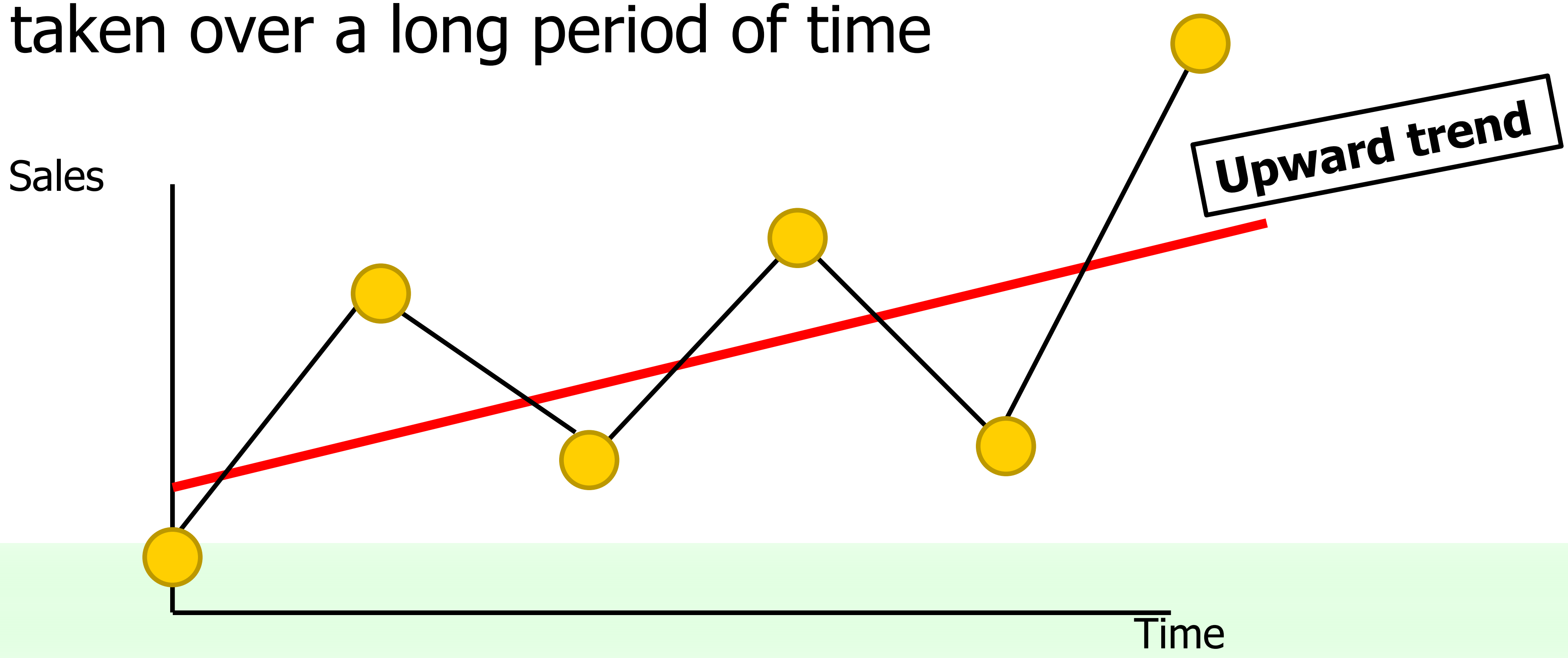
Time Series Components

Time Series

- **Trend** *Component*: Overall, persistent, long-term movement
- **Seasonal** *Component*: Regular periodic fluctuations
- **Cyclical** *Component*: Repeating swings or movements
- **Irregular / Random** *Component*: Erratic or residual fluctuations

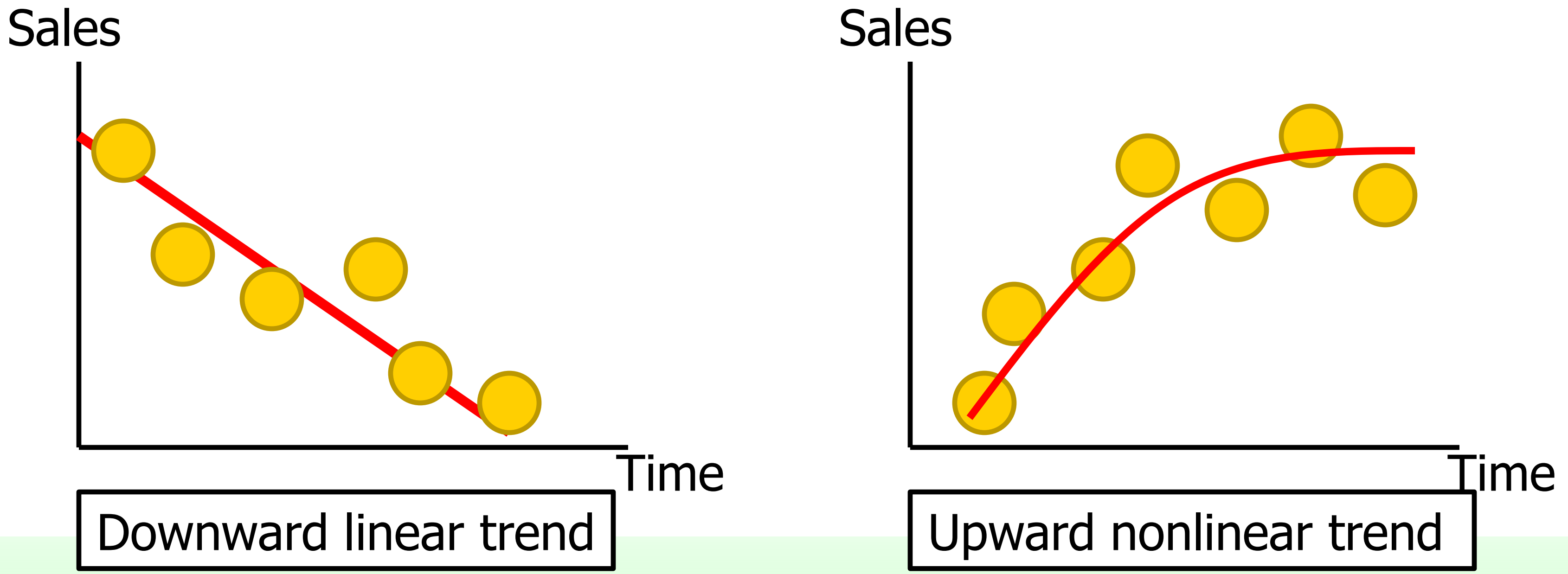
Trend Component

- Overall, persistent, long-term movement
  - increasing or decreasing over time
  - overall upward or downward movement
  - Data taken over a long period of time



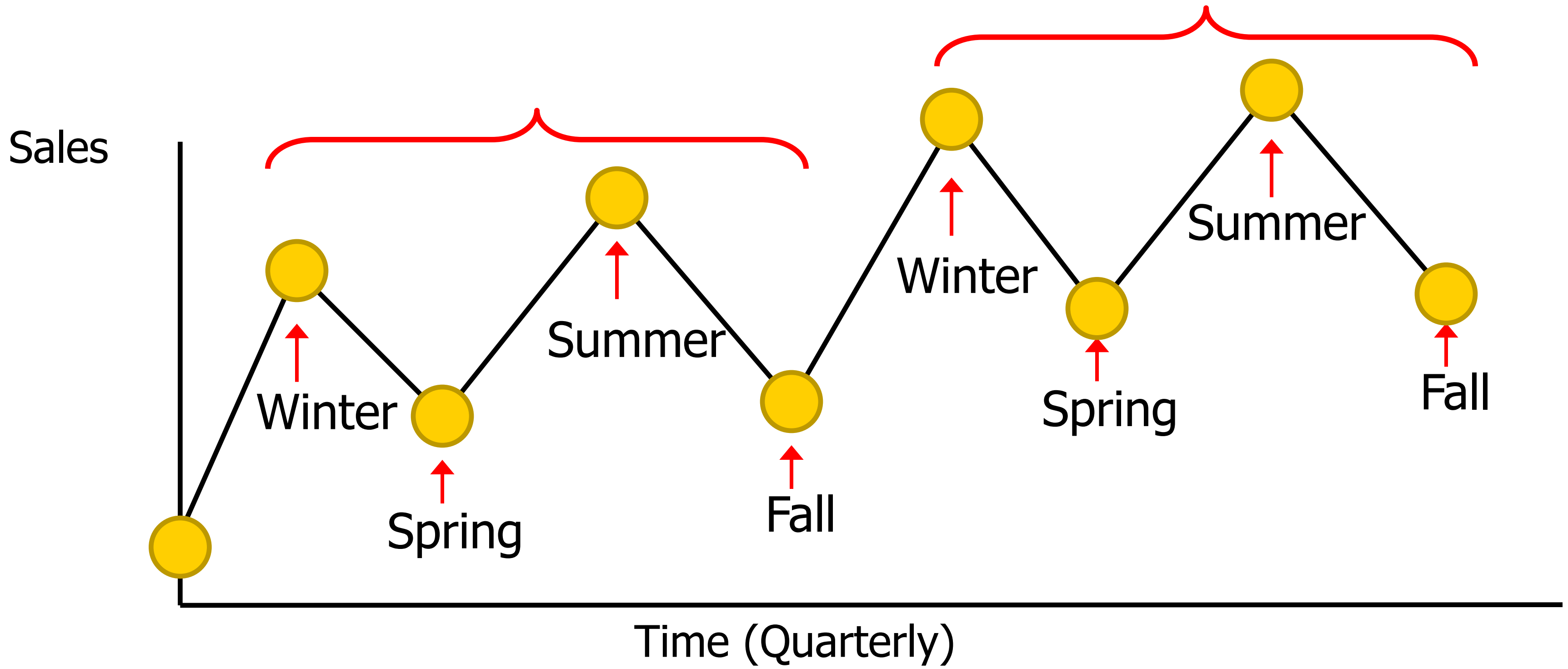
Trend Component

- Different trends
  - Trend can be upward or downward
  - Trend can be linear or non-linear



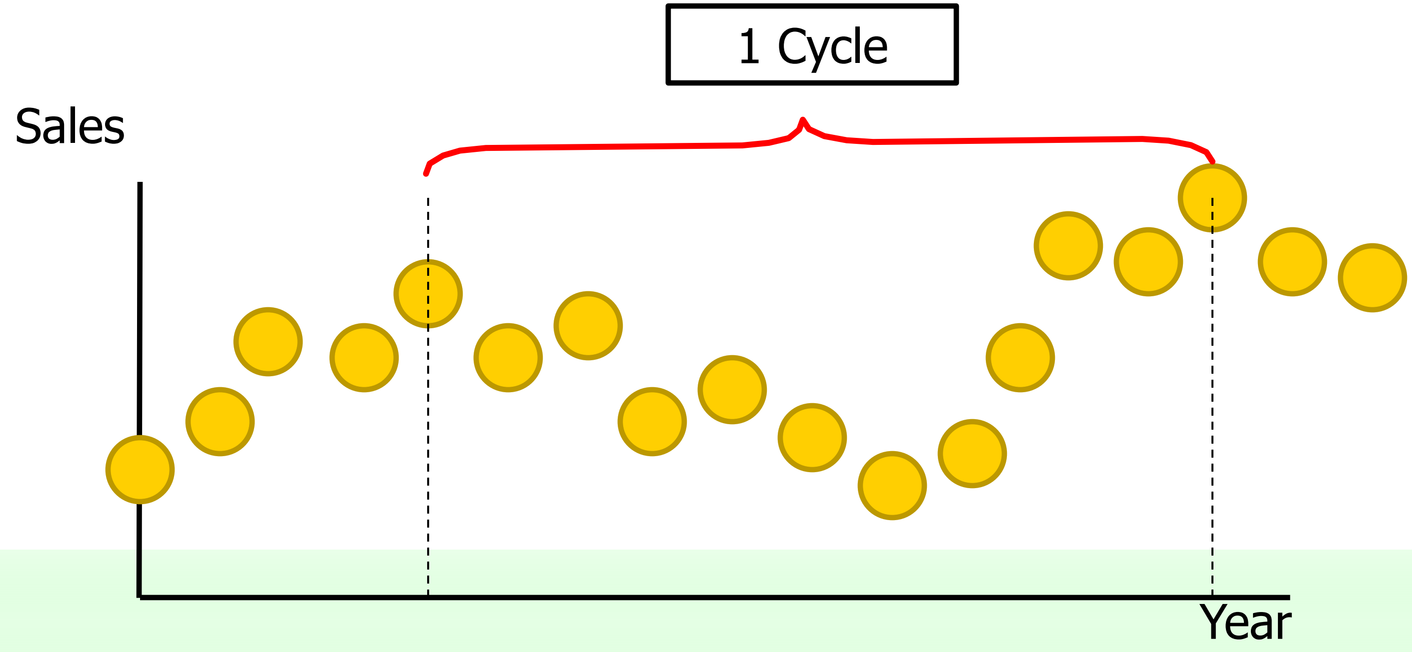
Seasonal Component

- Regular periodic fluctuations
  - Short-term regular wave-like patterns



Cyclical Component

- Repeating swings or movements
  - Long-term wave-like patterns
  - Regularly occur but may vary in length
  - Often measured peak to peak

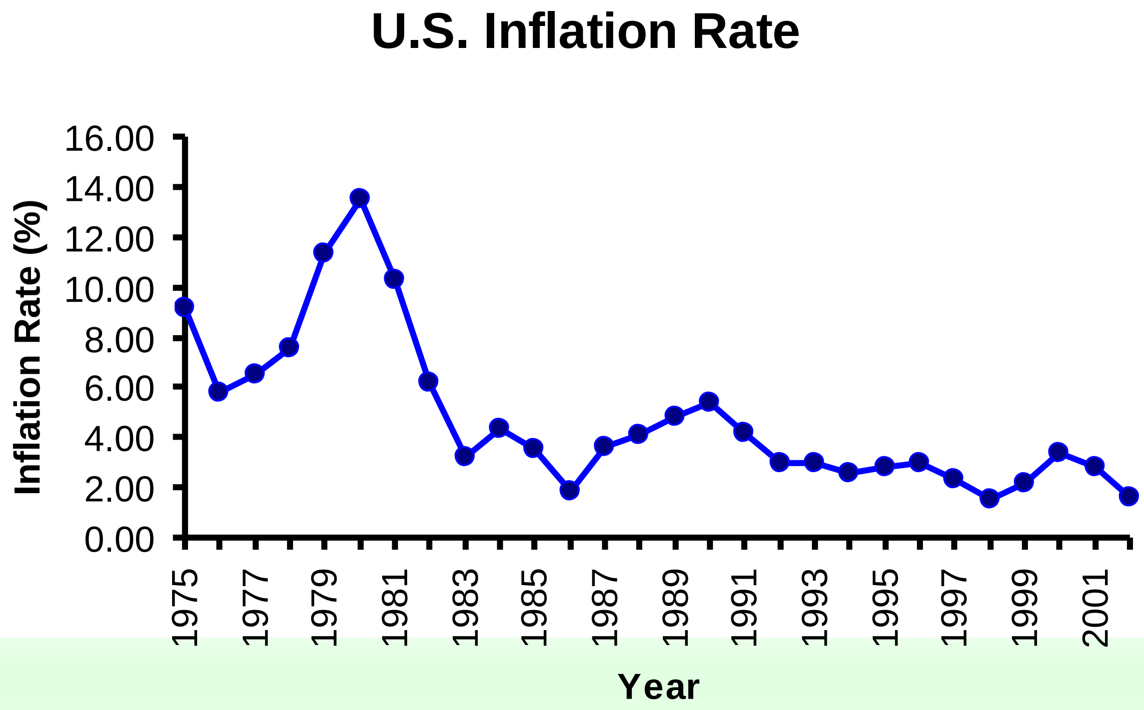


Irregular/Random Component

- Erratic or residual fluctuations
  - Caused by **unpredictable** influences
    - Influences are not regular and also they do not repeat in a specific pattern
  - This component usually represents “**noise**” in the time series

Discrete Time Series

- **Discrete time series**
  - Numerical sequence of data obtained at regular time intervals
    - e.g. time intervals can be annually, quarterly, monthly, weekly, daily, hourly.

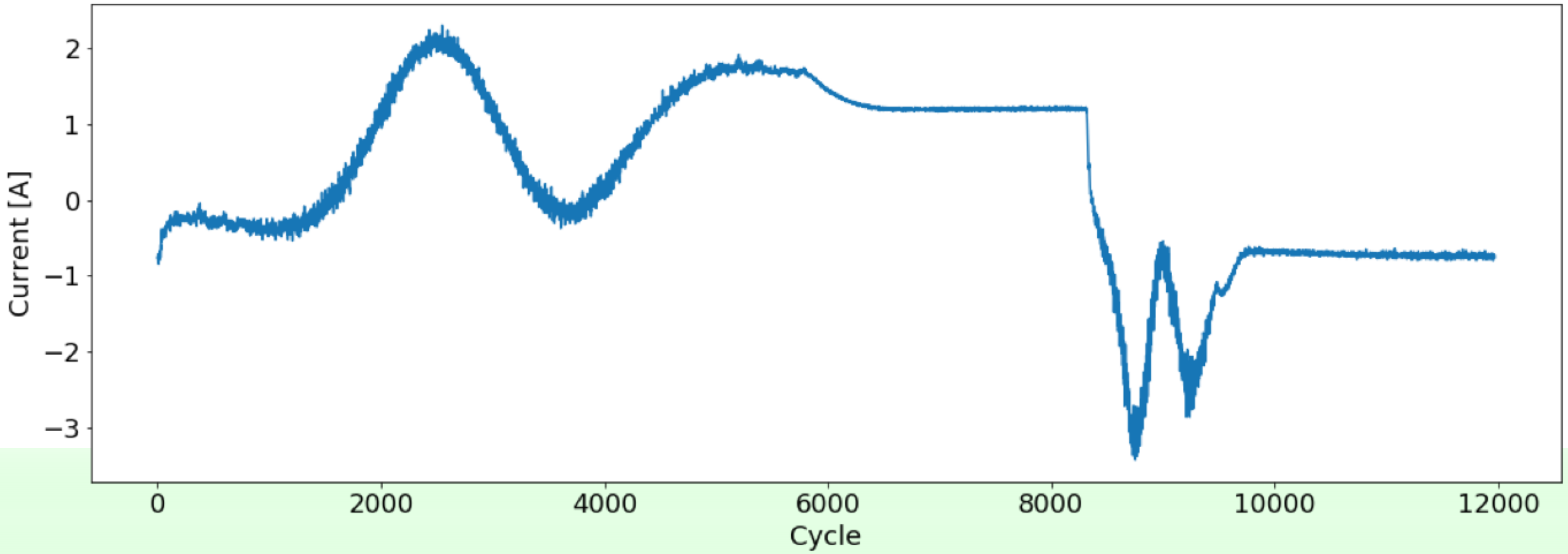


Continuous Time Series

Continuous time series

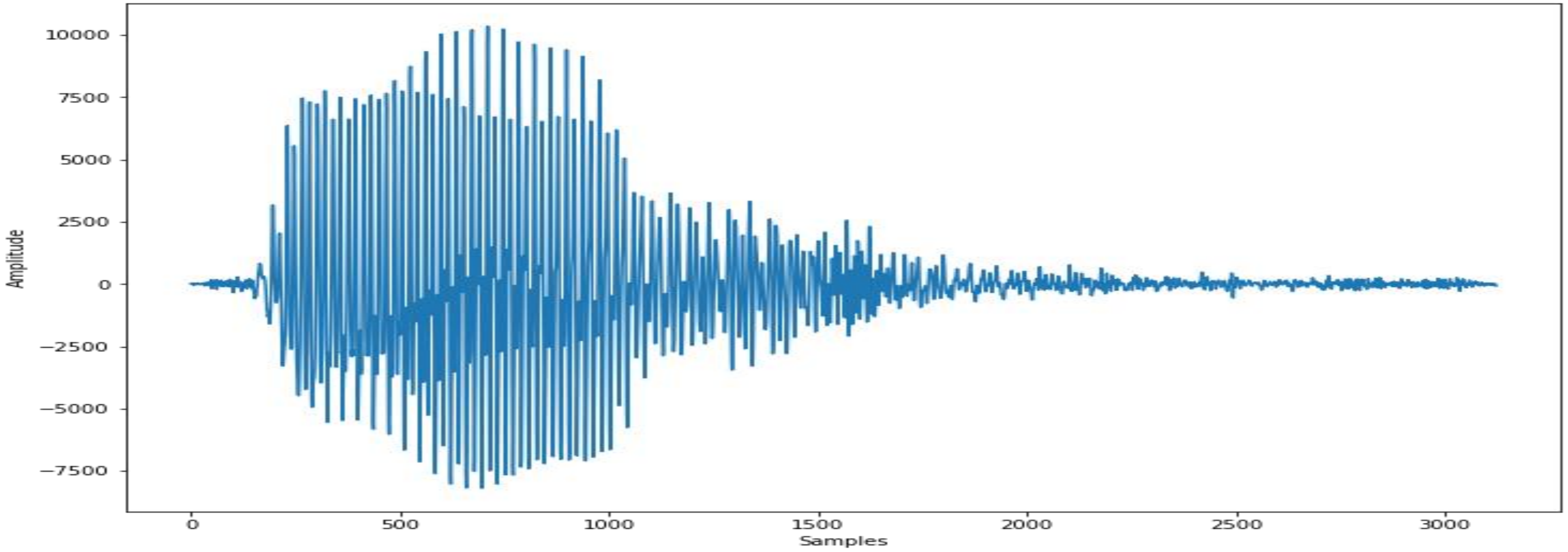
- observations are measured at every instance of time

- Example
  - The plot shows the current (Ampere) trend of a robotic arm over time.
  - Robot Cycle duration: about 24 s
  - Sampled every 2 ms (around 11,972 samples)



Continuous Time Series

- Example of **continuous** time series
  - Audio signal
  - The speaker said numbers from 0 to 9
  - Classification task: classifier the number said by the speaker



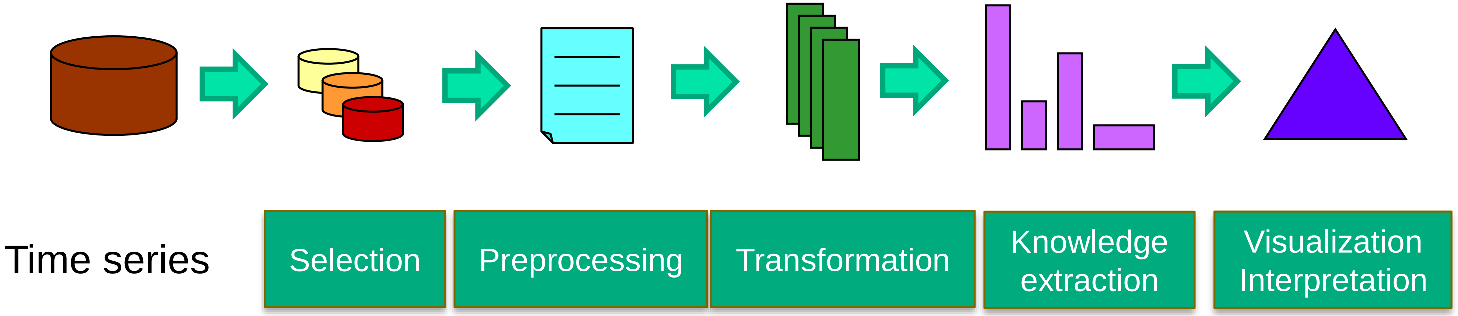
Time series Analytics tasks

DBMG

Politecnico di Torino

1859

KDD: Time series analysis



- In the preprocessing step
  - A time series **alignment** technique might be required
    - e.g., padding technique
  - In case of multivariate problem, **correlated time** series should be **identified** and **removed**
    - Correlation-based approach
    - Domain-driven knowledge
    - Mixed approach
- Transformation
  - Feature engineering
  - Feature embedding

Analytics tasks: Classification and Forecasting

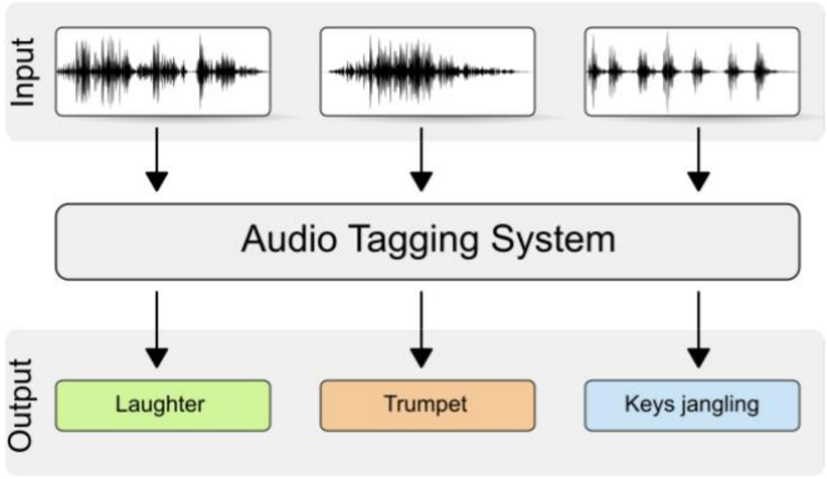
- Many algorithms can be applied to address classification and forecasting tasks
  - Machine learning algorithms such as
    - Random forest classifier/regressor
    - SVM
    - Neural networks, etc.
  - Statistical approaches
    - e.g., ARIMA (Autoregressive integrated moving average) models
- Given an analytics goal different methods can be exploited
  - The algorithm selection is driven by
    - Application requirements: accuracy, human-readable model, scalability, noise and outlier management
    - The complexity of the analytics task

Artificial Neural Networks

- Based on the analytics goal common neural network architectures can be used to analyze time series
  - Classification task
    - CNN (Convolutional Neural Network)
  - Forecasting task
    - RNN (Recurrent Neural Network)

Convolutional Neural Networks

- VGGish is a convolutional neural network to extract the relevant features from audio signals
  - The inputs of the network are log mel spectrogram audios
  - The output is an audio embedding
    - It can be used for further analytics tasks like classification



https://github.com/tensorflow/models/tree/master/research/audioset/vggish

Recurrent Neural Networks

- RNN for time series forecasting
  - Connections between nodes form a directed graph along a temporal sequence
    - This allows the network to exhibit temporal dynamic behavior.
  - RNNs can use their internal state memory to process sequences of inputs

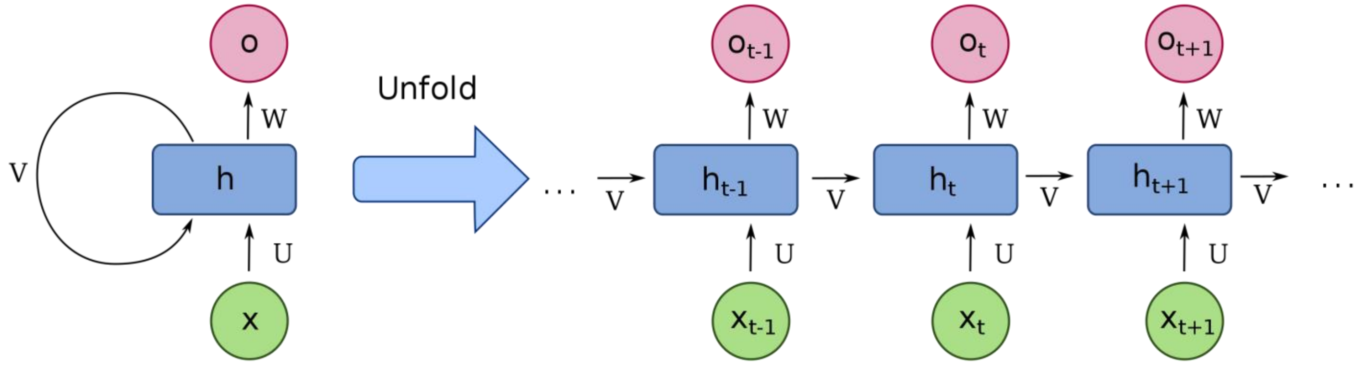
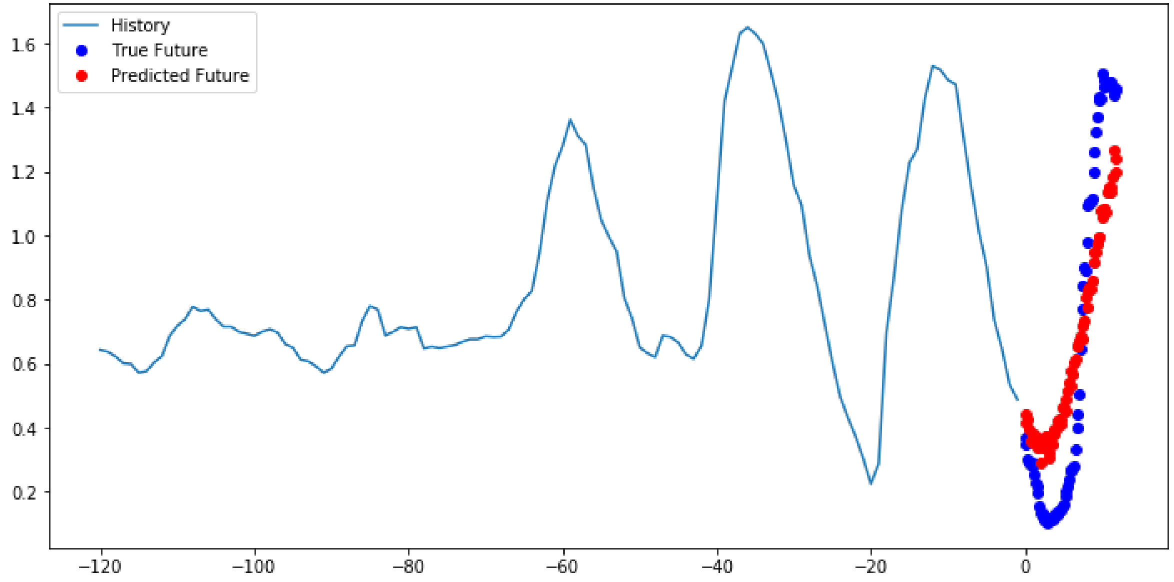


Diagram of a single unit RNN

Recurrent Neural Networks

- RNN for time series forecasting in the context of meteorological data
- 72 predicted values



https://www.tensorflow.org/tutorials/structured_data/time_series

Strong and weak points of Artificial Neural Networks

- Main advantages
  - Powerful algorithms to train accurate models
  - Able to deal with different complex analytics tasks

- Main drawbacks
  - A huge amount of data is required for the training phase
  - Training is an heavy task in terms of both computational time and hardware resources
  - The <u>feature learning</u> step is hidden in the network
    - The user is not able to understand the key features driving the prediction task
    - Some specific analytics tasks might require an ad-hoc feature engineering step to better characterize the input data and train more accurate models

Feauture engineering

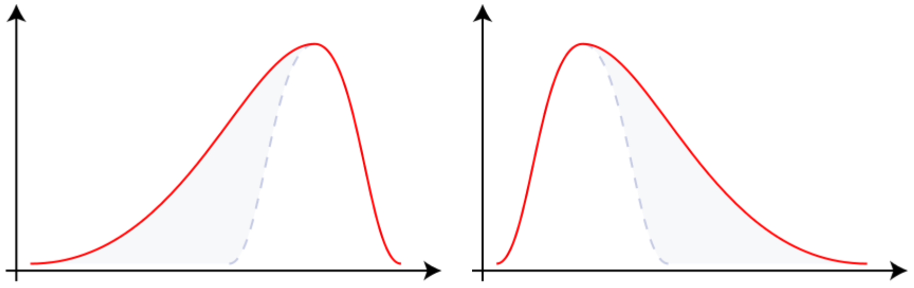
- Feature computation over a time series
  - Basic statistics
    - Min,Max,Mean,Standard Deviation
  - Indices
    - Kurtosis
    - Skewness
  - Time series summarization
    - Percentile
    - Joint approach based on CDF + percentile
    - Technique based on Derivate + CDF + percentile
  - Linear Regression
- Different combinations of features can be evaluated
- Correlated features are identified and removed
- Selected features model the time series under analysis
- Selected features will feed the next analytics tasks

Basics statistics

- Given a time series
  - Minimum value
  - Maximum value
  - Mean value
  - Number of samples
  - Stardard deviation
  - ...

Skewness

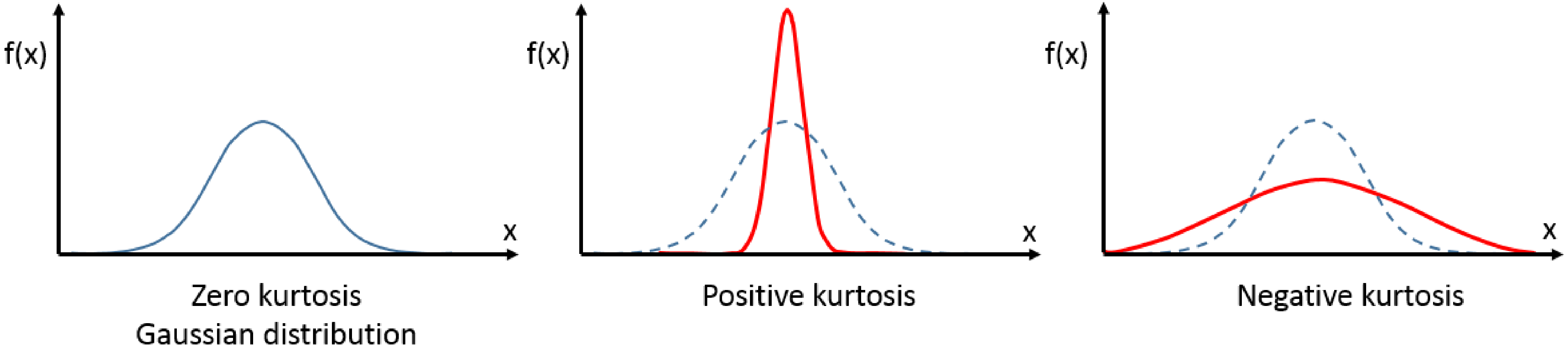
- **Skewness** is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.



  - where μ is the mean, σ is the standard deviation, E is the expectation operator, μ_3 is the third central moment, and κ_t are the t-th cumulants.

Kurtosis

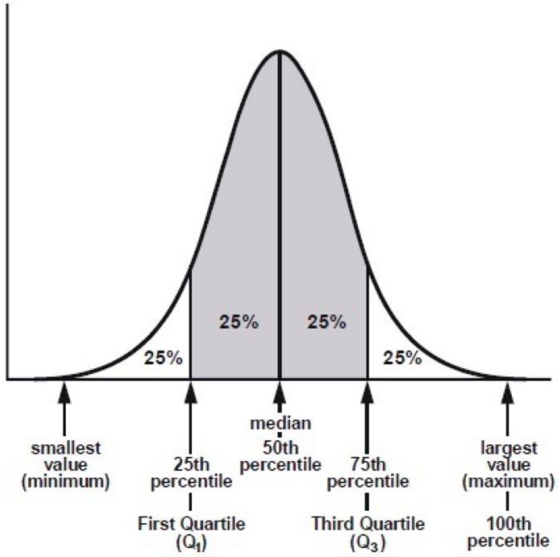
- **kurtosis** is a measure of the "tailedness" of the probability distribution of a real-valued random variable.



- where $\boldsymbol{\mu_4}$ is the fourth central moment and $\boldsymbol{\sigma}$ is the standard deviation.
- A very common choice is κ, which is fine as long as it is clear that it does not refer to a cumulant.
- Other choices include γ_2, to be similar to the notation for skewness, although sometimes this is instead reserved for the excess kurtosis.

Time Series Summarization

- **Percentile** indicates the value below which a given percentage of observations in a group of observations falls
- Representing a time series through percentiles allow representing the entire distribution
  - Selecting the four percentile
  - Selecting the ten percentiles
    - selected **10 percentiles**: 10, 20, 30, 40, 50, 60, 70, 80, 90, 99
    - remove outliers by removing the last percentile of the distribution
- The temporal sequence is lost
- The percentiles are the **features** describing the time series



Time Series Summarization

- **The Cumulative Distribution Function** of a real-valued random Variable X is the function given by

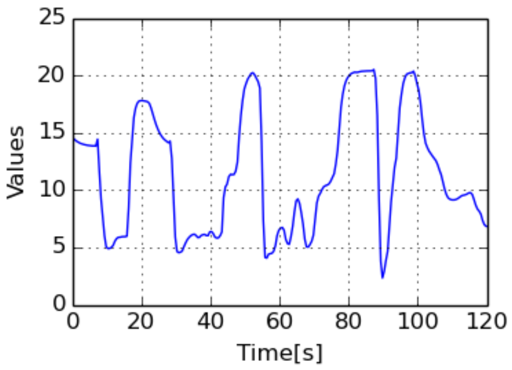
$$F_x(x) = P(X \leq x)$$

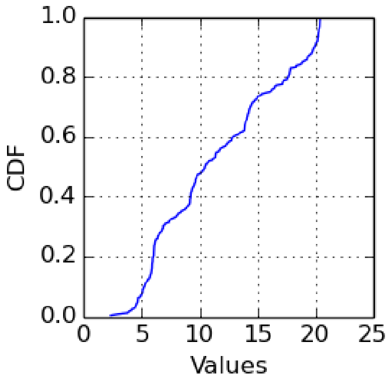
where the right-hand side represents the probability that the random variable X takes on a value less than or equal to x. The probability that X lies in the semi-closed interval $(a, b]$, where $a < b$, is therefore

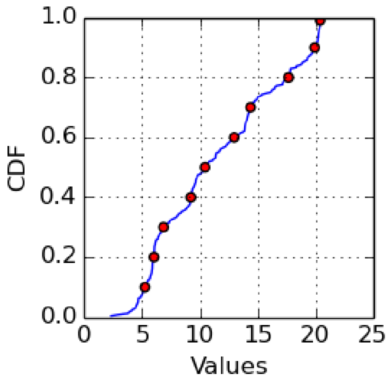
$$P(a < X \leq b) = F_x(b) - F_x(a)$$

Time Series Summarization

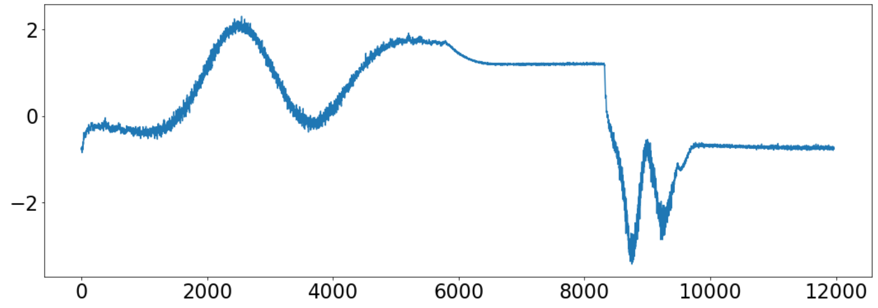
- **Compute the Cumulative Distribution Function** (CDF) to represent time series samples
  - To compute a reliable CDF at least 100 samples are required
- Summarize the CDF with a few **percentiles**
- The percentiles are the **features** describing the time series



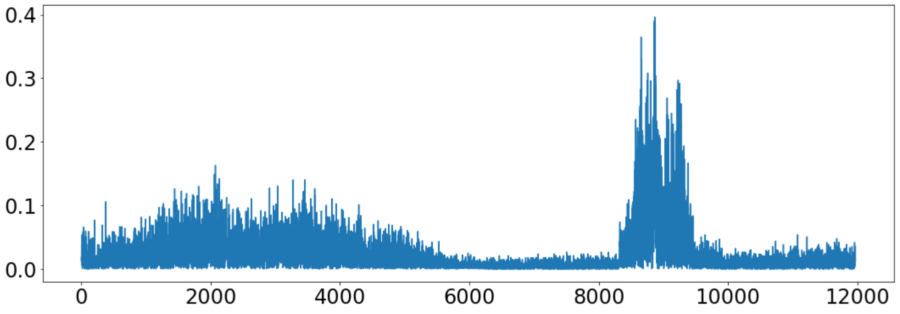




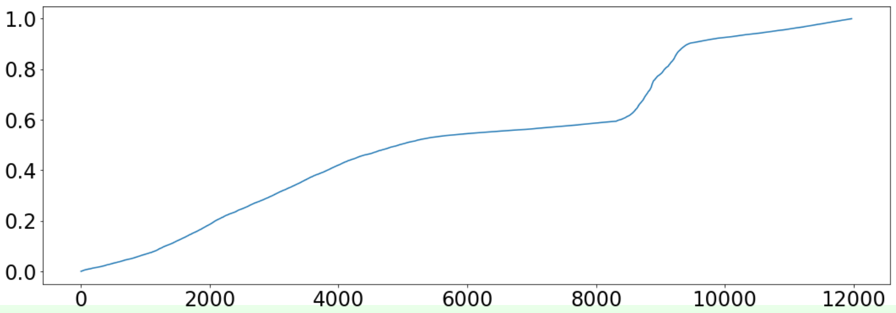
Derivate + CDF + Percentile



- Current (Ampere) of a robotic arm



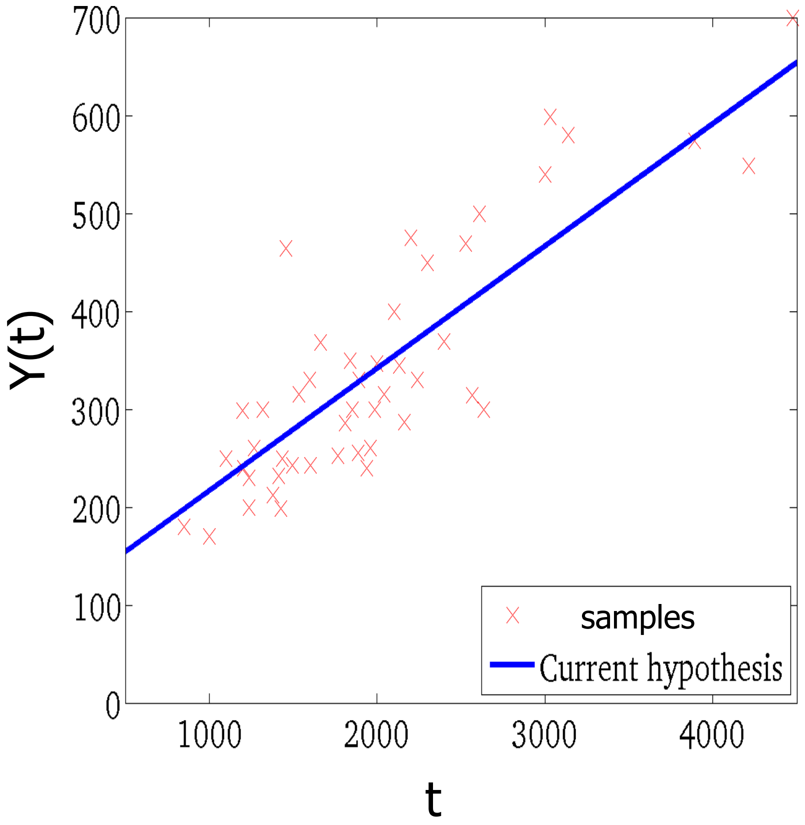
- Derivate



- CDF
- Summarize the CDF with a few **percentiles**

Linear Regression

$$Y(t) = \beta_0 + \beta_1 t + \varepsilon$$



- β0: The **intercept** represents the estimated value of y when t assumes 0
  - β0 is the portion of y not explained by t

- β1: the **slope** measures the estimated change in the y value as for every one-unit change in t
  - The average value of a t change

Time series Characterization

- Feature engineering can be calculated **on the entire time series**
  - e.g., on a signal representing the current consumption
    - Statistics on the entire robot cycle can be useful to characterize the overall time series trend

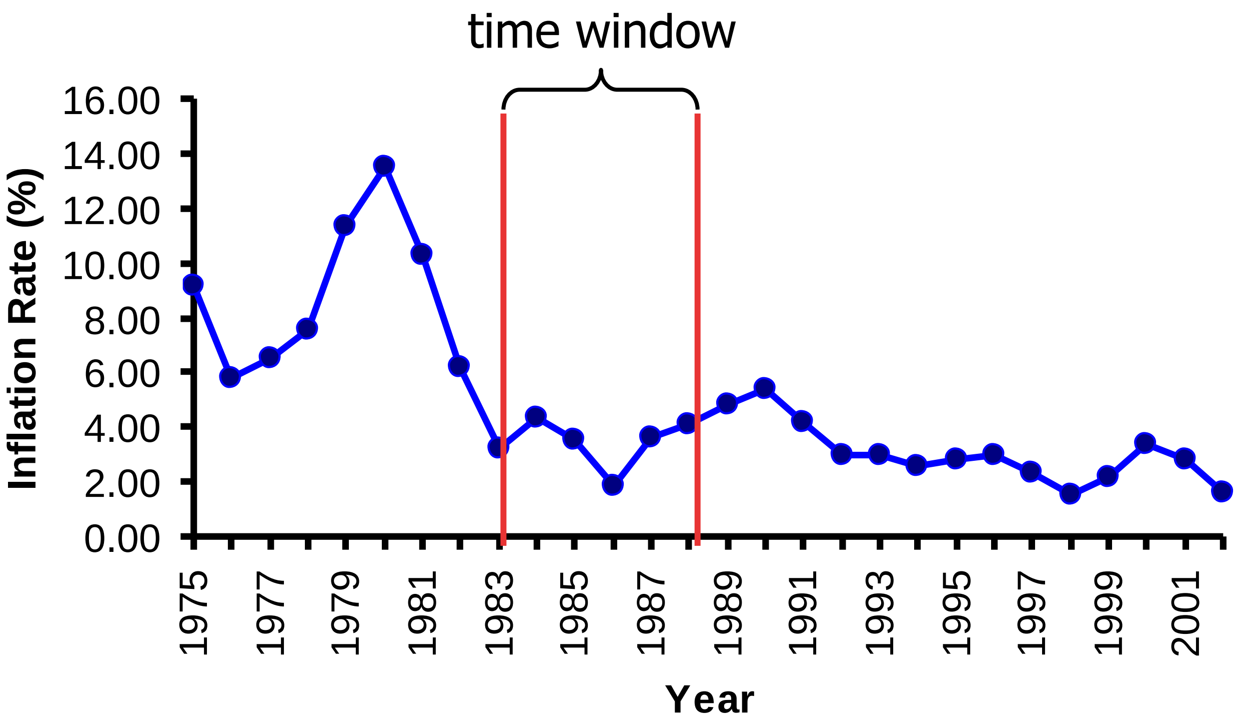
- Feature engineering can be also calculated **in local parts** of the time series
  - **Time windows**
  - For each time window, statistical features summarize the local time series trend

Time Window

- Time windows are defined by
  - **Window length**: size (in time units) of each window
    - Domain-driven (e.g. parts of the speech, actions of a mechanical arm)
    - Data-driven (e.g. time windows on seasonal features of a signal)
  - **Window shift**: window position with respect to contiguous windows
    - Not-overlapped (jumping): all windows are independent and do not share any data
    - Overlapped (sliding): two consecutive windows share a portion of data

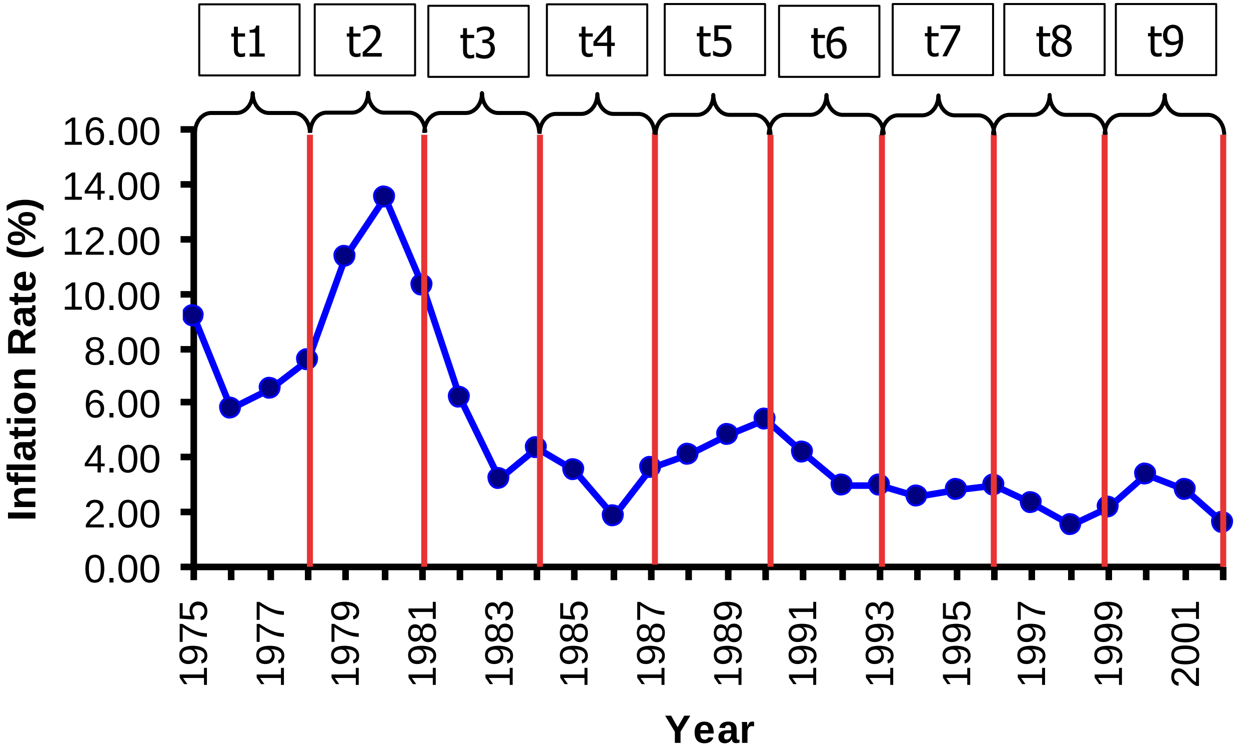
Time Window

- A time-window is a fixed interval of time when the data stream is processed for query and mining purposes



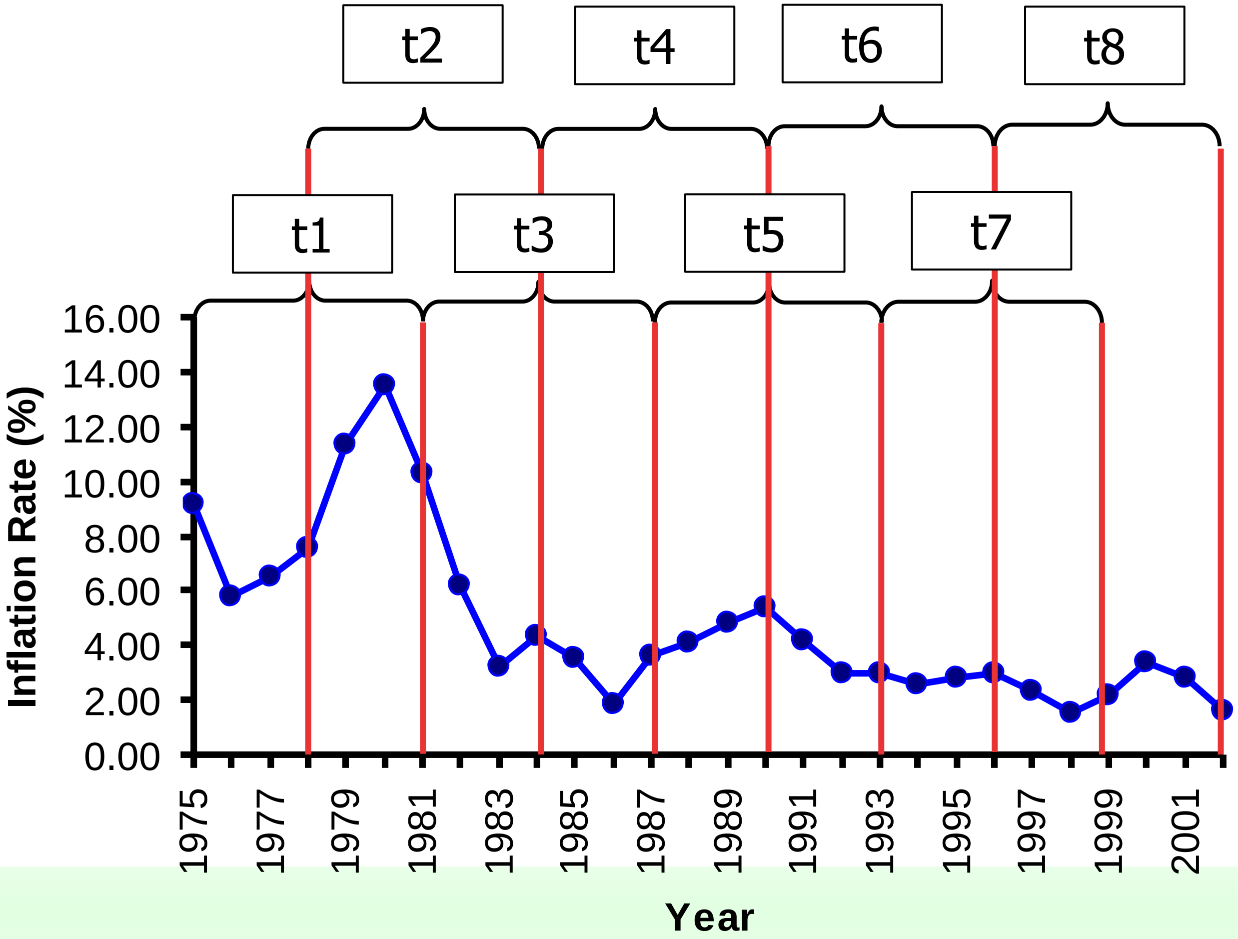
Time Window Not-overlapped

- All windows are independent and do not share any data



Time Window Overlapped

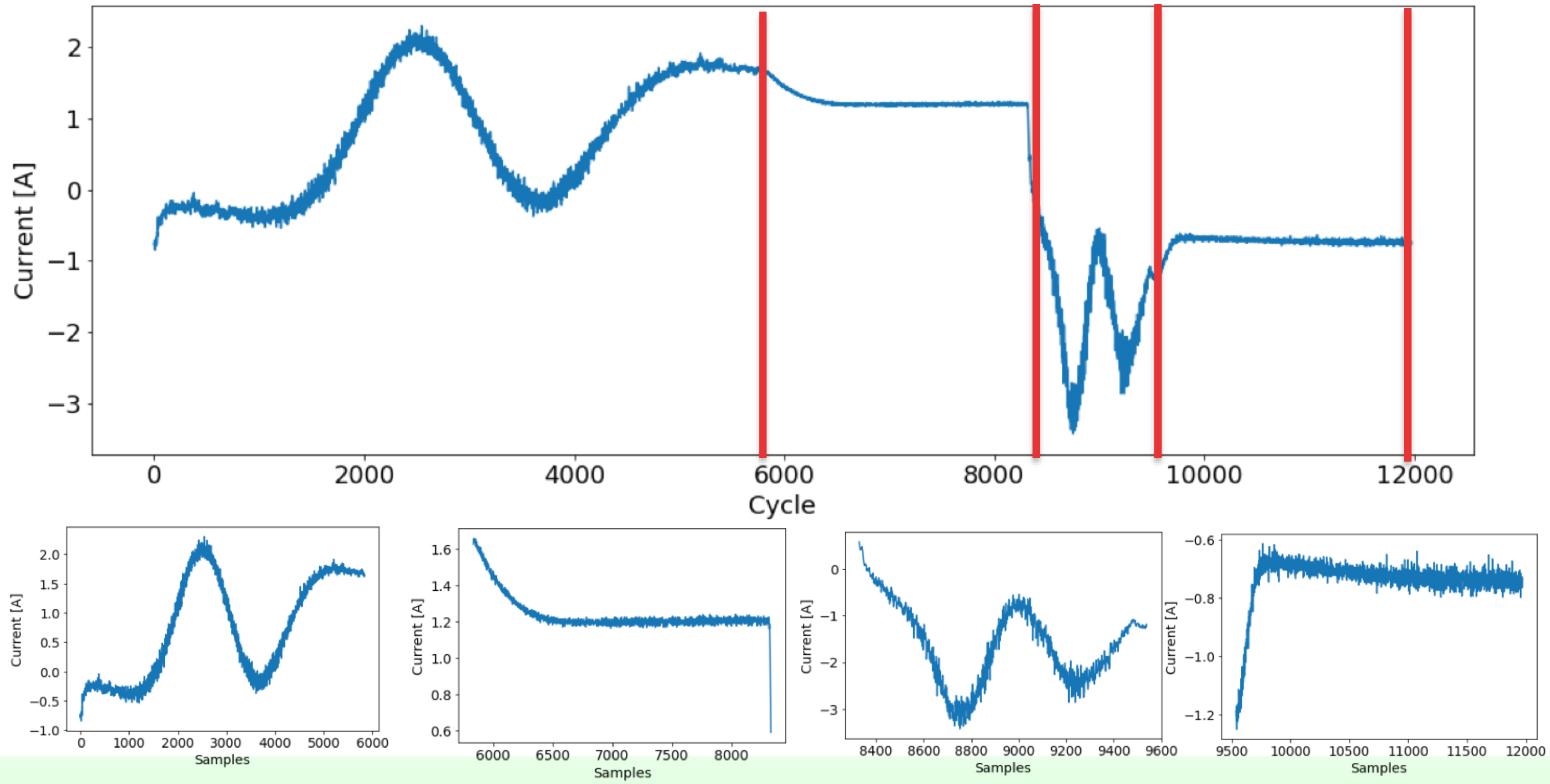
- Two consecutive windows share a portion of data



Time window: an example

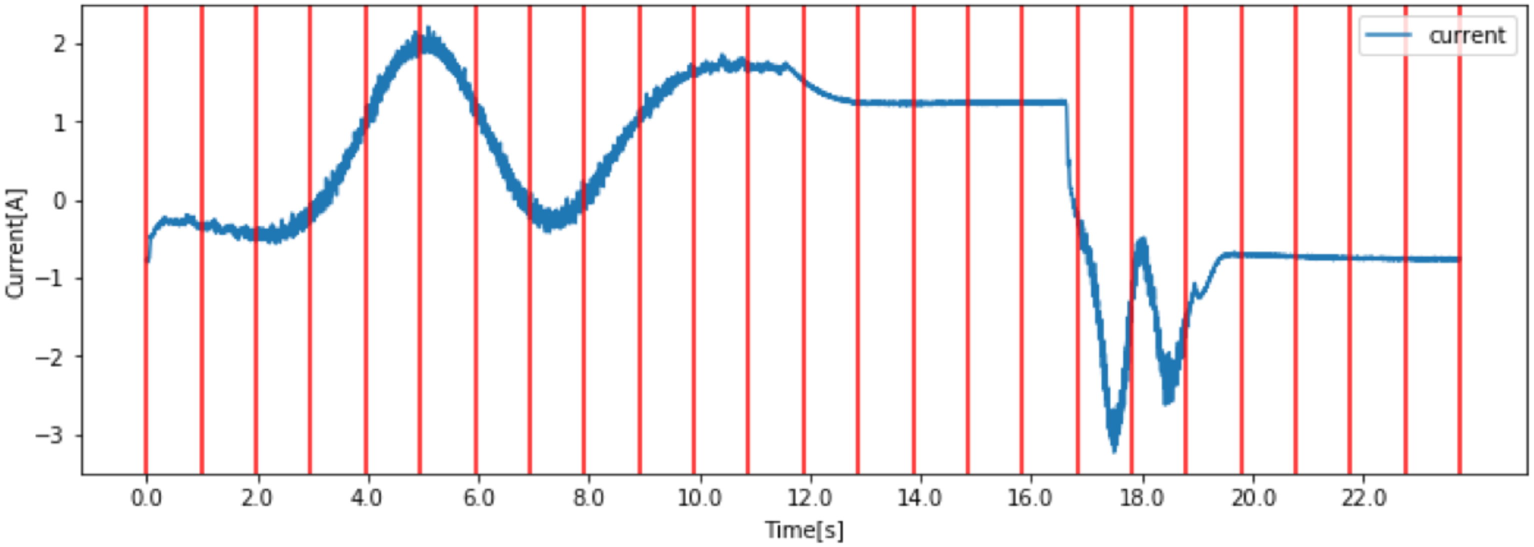
- Domain Driven

The plot shows the current (Ampere) trend of a robotic arm over time.



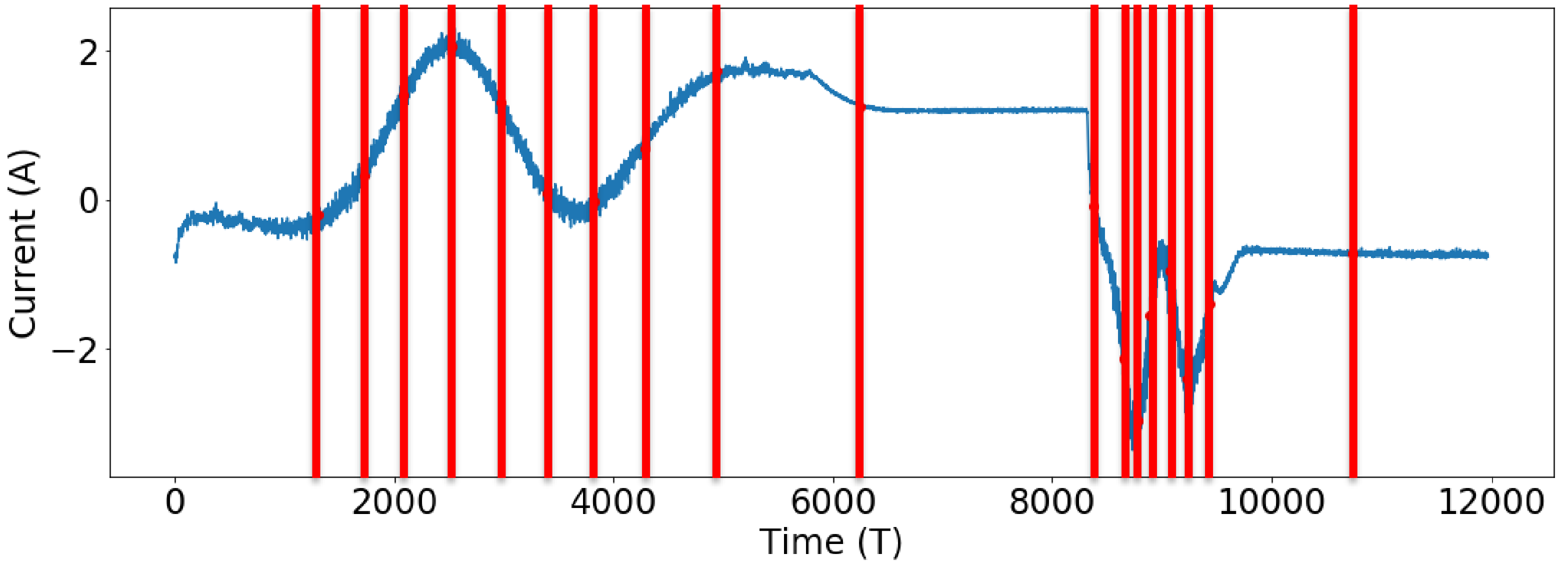
Time window: an example

- Data Driven

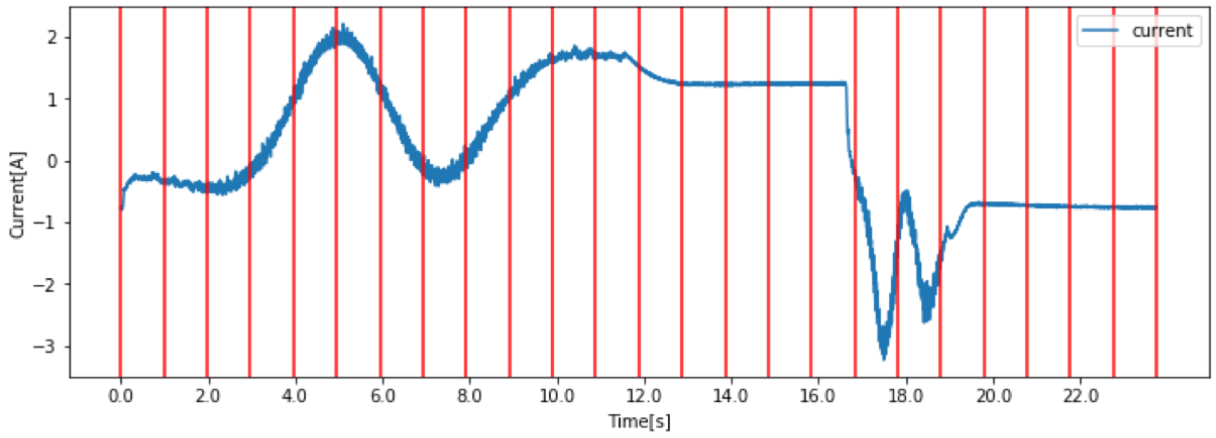


Time window: an example

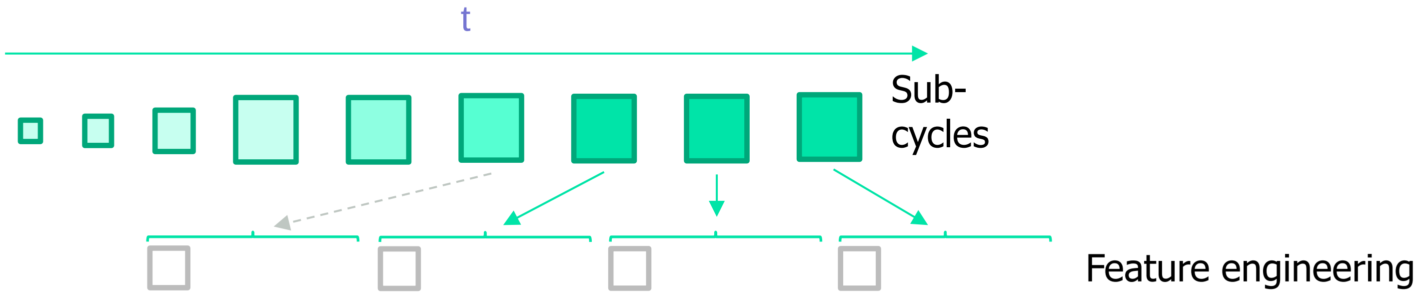
- Longer time window in those parts where the time series is more stable
- Shorter time window in those parts where the time series varies most



Time Series summarization



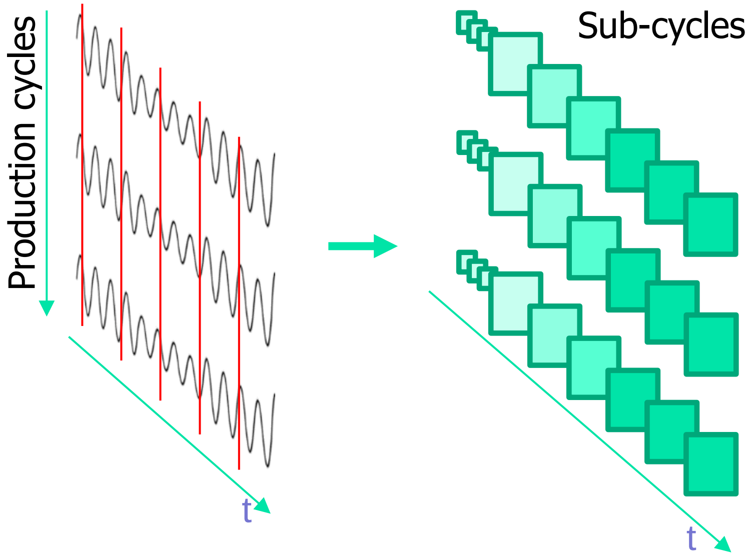
Each sub-cycle corresponds to a time windows



The time series trend can be captured through the features extracted from each sub-cycle of each time series

Time Series Aggregation

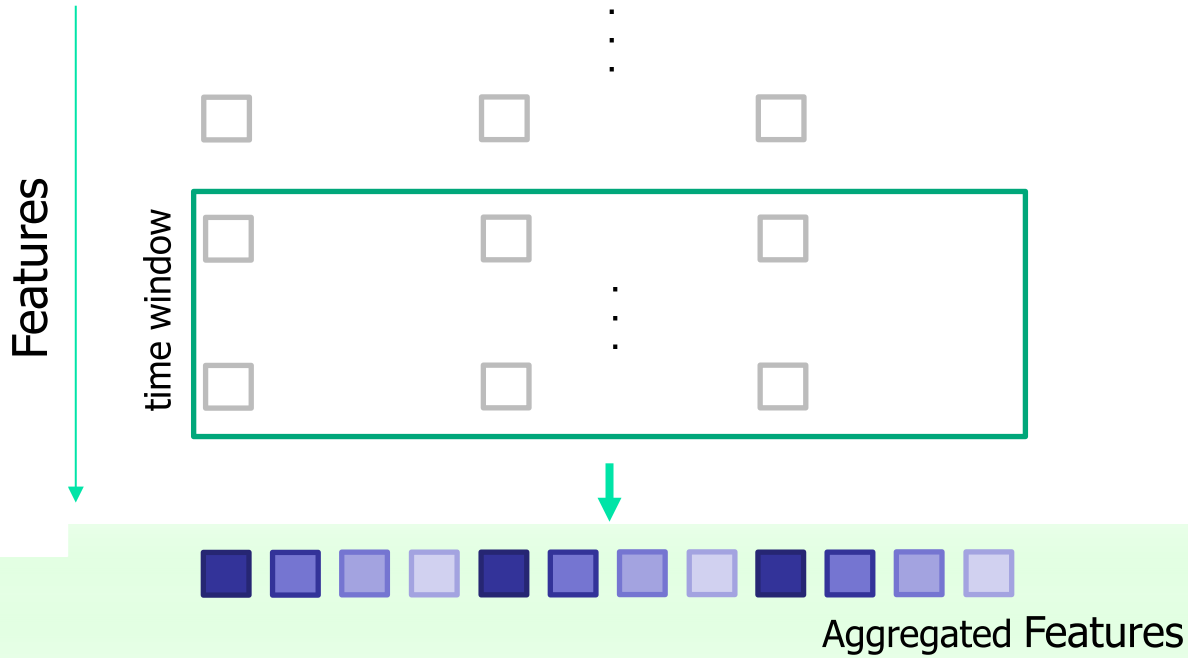
- In slowly-degrading environments single time-series (cycle) predictions have a too short horizon.
- To deal with long horizon prediction
  - The multi-cycle time-based aggregation step could be based on time series aggregation over a time window.



Cycles divided into time window to extract the variability of each sub-cycle

Time Series Aggregation

- The main characteristics of the window is captured through feature computation over a time window of features
- The feature aggregation preserves the meaning of the time series, keeping the process transparent.
  - Different feature computation can be exploited



Feature selection and removal

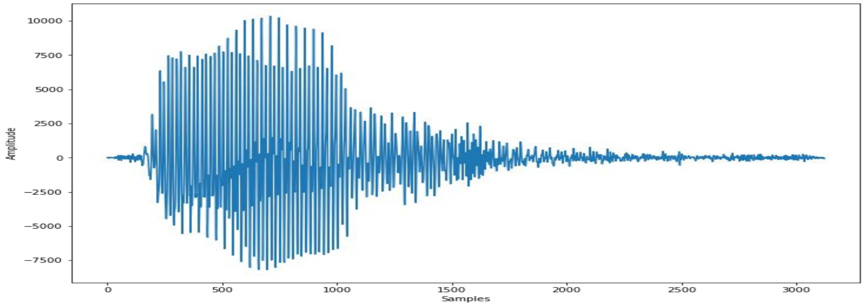
- In case of a large number of features modeling the time series, **some of them** might provide **redundant information**.
- **Feature selection and removal** simplifying the model computation
  - improving the model performance
  - Enhancing the model interpretation (i.e., better explainability of the dependent variables)
- Feature selection based on correlation-based approach)
  - Features highly-correlated with other features could be discarded from the analysis
  - having dependence or association in any statistical relationship, whether causal or not

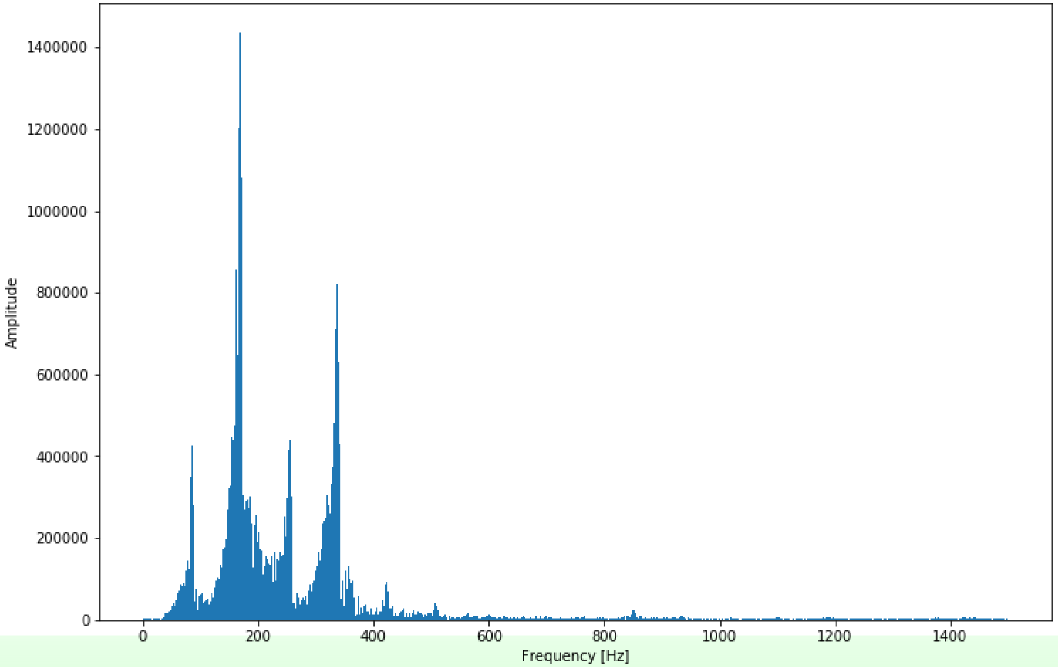
Time series in frequency domain

- In some cases, it may be useful to analyze a signal in the frequency domain.
  - e.g., audio, video, etc...

- The Fourier transformation can transform a time series in the frequency domain

Time series in frequency domain





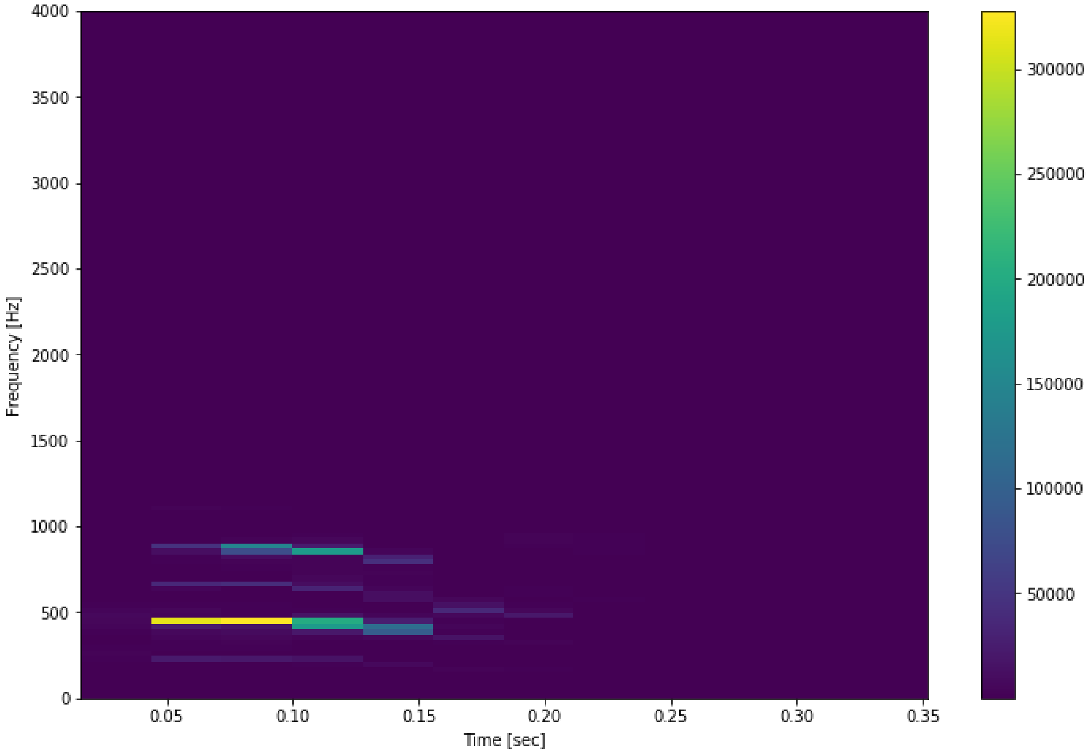
- Audio Signal in time domain
- Audio Signal in the frequency domain through the Fourier Transformation

Time series in frequency domain

- To analyze an audio signal in the frequency domain spectrograms are often usually used

- A **spectrogram** is a visual representation of the spectrum of frequencies of a time series as it varies with time.
  - In the case of audio, spectrograms are sometimes called sonographs, voiceprints, or voicegrams.

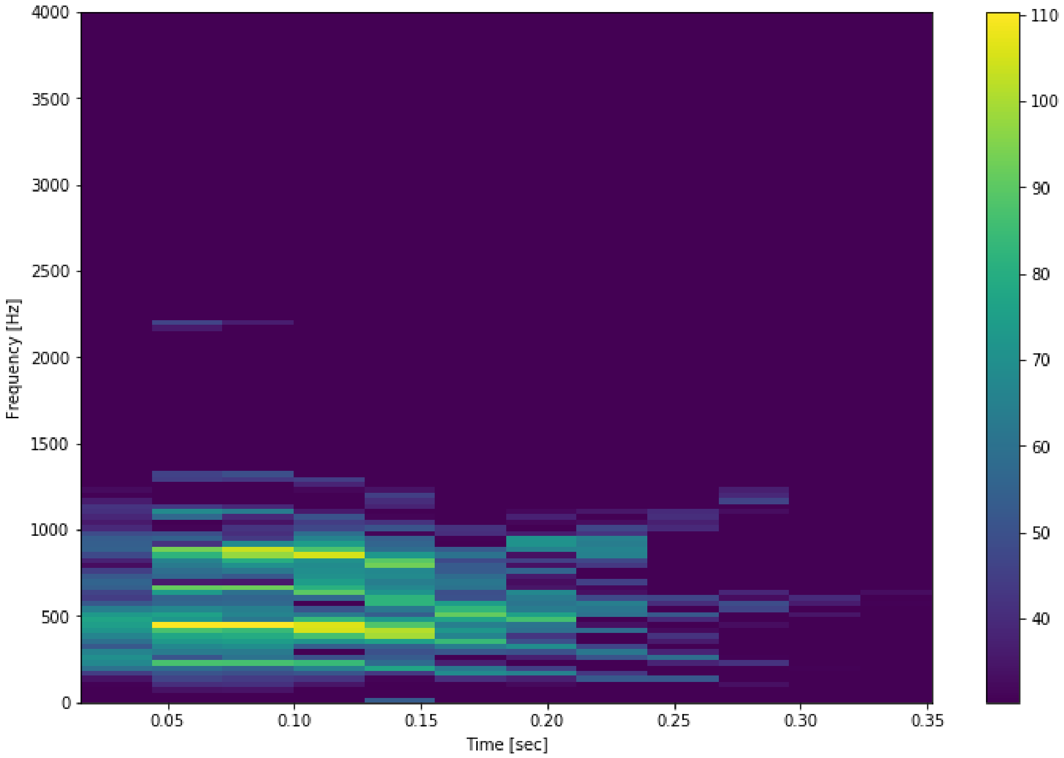
Time series in frequency domain



- In the **spectrogram** the colour intensity corresponds to the signal amplitude.
- If the amplitude is linear, it is difficult to identify the components because the audio follows logarithmic trends
- A data transformation is needed

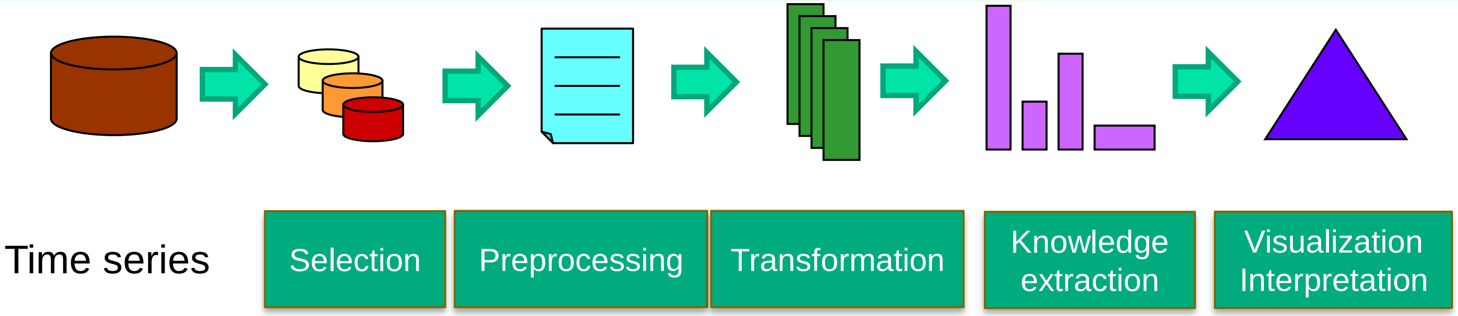
Usually obtained by dividing the total time interval into equal subintervals and calculating the Fourier transform of waveform in each window which gives the amplitude depending on the frequency

Time series in frequency domain



- In this plot the amplitude has been transformed from linear to logarithmic in order to give more emphasis to musical, tonal relationships

KDD: Time series analysis



- Knowledge extraction
  - Different algorithms can be exploited to address the analytics tasks
  - Selected features feed the knowledge extraction algorithm

- Visualization and interpretation
  - Help the domain expert correctly understand the extracted knowledge items to effectively support the decision-making process