Decomposition of Time Series and Data Streams

Rodrigo Fernandes de Mello

Invited Professor at Télécom ParisTech (until July 2019)

Associate Professor at Universidade de São Paulo, ICMC

mello@icmc.usp.br

March 21th, 2019

DigiCosme, Paris-Saclay

Some information on the University of São Paulo, Brazil

- Founded in 1934 in the State of São Paulo
  - 46 million people
  - 12 million in the city of São Paulo (capital)
    - 21 million in the capital and neighborhoods
- 11 Campi
  - Annual budget 1.2 billion Euros
- 96.364 students
  - 58.823 Undergraduation
  - 14.106 Masters
  - 15.894 PhD

DigiCosme, Paris-Saclay

Some information on the University of São Paulo, Brazil

- Students:
  - 52.28% men and 47.72% women
- Staff members
  - 5.844 Professors
  - 14.866 Administrative positions
- World University Ranking of the Times Higher Education

Rank	Name	Overall	Teaching	Research	Citations	Industry Income	International Outlook
251–300	University of São Paulo Brazil	46.4–49.4	55.9	53.5	37.0	39.5	32.7

Some Definitions

- What is a data stream?
  - Unbounded sequence of observations collected along time
- What is a time series?
  - Sequence of observations along time
    - Each observation is typically single dimensional, but that is not a law
      - Ex: The Lorenz System with 3 partial derivatives
    - They can be uniformly or non-uniformly spaced along time
    - It does not assume anything about being unbounded

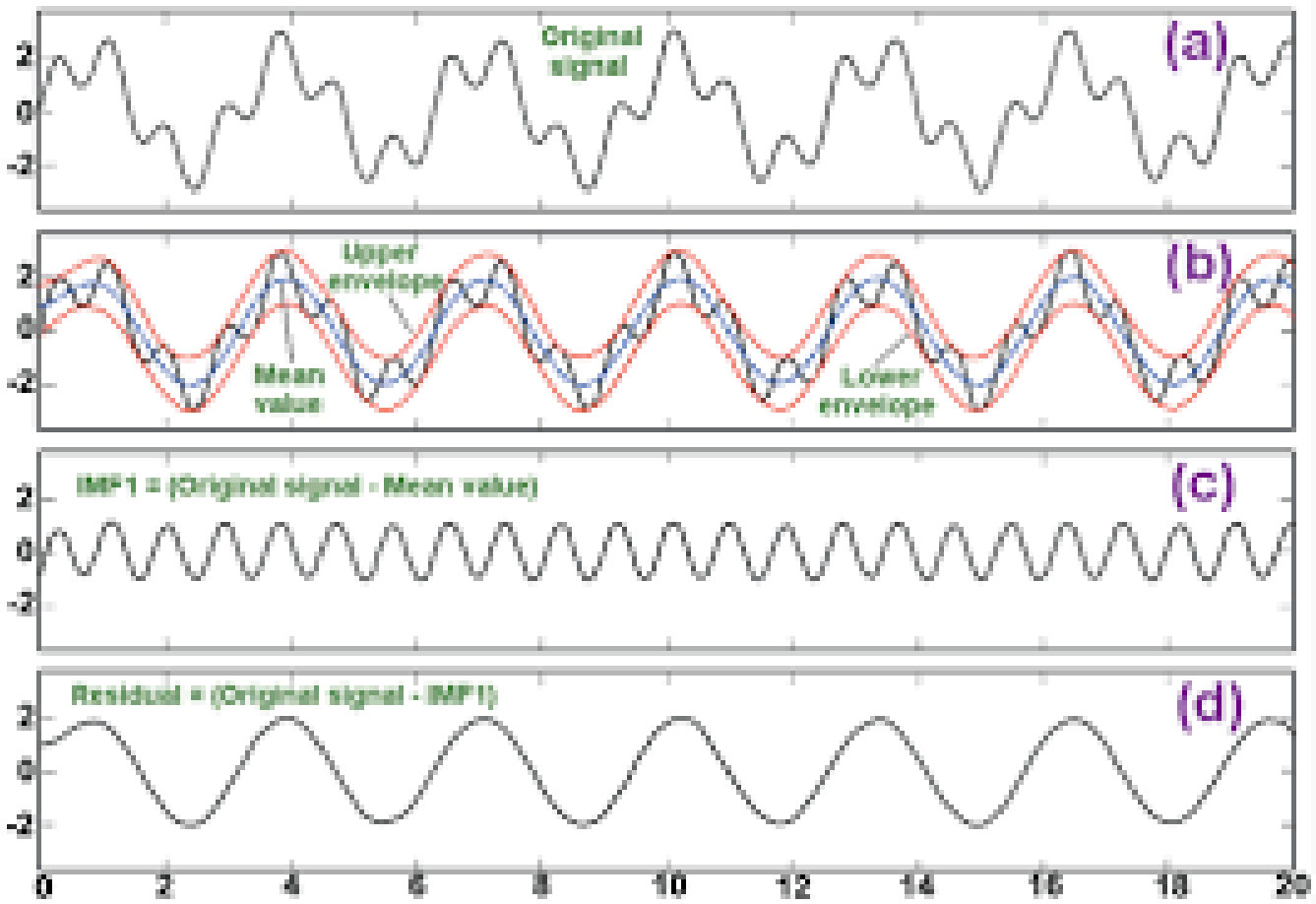
Some Definitions

- What is the Empirical Mode Decomposition?
  - It is a way to decompose a signal into Intrinsic Mode Functions (IMF)
  - It is designed to work with nonstationary and nonlinear data
  - It is a.k.a. the Hilbert-Huang transform (HHT)

Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1971). "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis". Proceedings of the Royal Society of London A. 454: 903–995.

Some Definitions

- What is the Empirical Mode Decomposition?



Some Definitions

- We decided to apply EMD to decompose time series and data streams into their:
  - Stochastic and Deterministic components
  - Allowing to take advantage of both branches

Ricardo Araújo Rios, Rodrigo Fernandes de Mello. Applying Empirical Mode Decomposition and mutual information to separate stochastic and deterministic influences embedded in signals. Signal Processing 118: 159-176 (2016)

Conclusions

- After such decomposition:
  - We can employ classical time series modeling strategies to understand data
    - From Statistics
      - ARIMA models
    - From Dynamical Systems
      - Takens' theorem
      - False Nearest Neighbors
      - Auto-Mutual Information
  - We can obtain an i.i.d. input space for ML algorithms
    - Ensuring learning guarantees according to the Statistical Learning Theory

      Lucas de Carvalho Pagliosa, Rodrigo Fernandes de Mello. Applying a kernel function on time-dependent data to provide supervised-learning guarantees. Expert Syst. Appl. 71: 216-229 (2017)

Decomposition of Time Series and Data Streams

Rodrigo Fernandes de Mello

Invited Professor at Télécom ParisTech (until July 2019)

Associate Professor at Universidade de São Paulo, ICMC

mello@icmc.usp.br

March 21th, 2019