TIME SERIES

Introduction

Data representation refers to the process in which data is stored, processed and outputted. Data can be interpreted in a time series representation. The subdivisions in the types of time series representation are data-adaptive time series and non-data adaptive time series. Piecewise Aggregate Approximation (PAA) and Symbolic Aggregate Approximation (SAX) are two different types of data-adaptive time series representations.

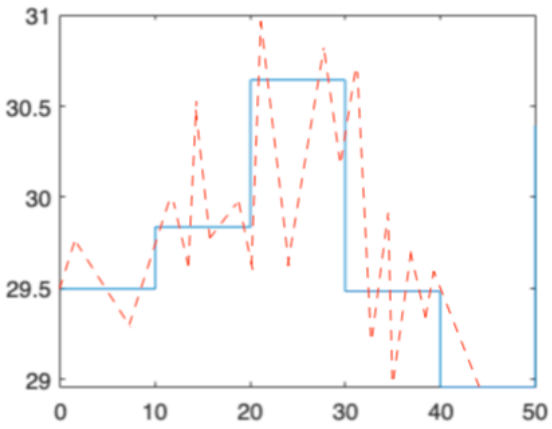
PAA corresponds to the downsampling of the original time series and each segment with fixed segment size. The mean value is retained.

SAX builds upon the implementation of PAA by quantizing the mean value. Those quantization boundaries are then represented using symbols, under standard normal distribution assumption.

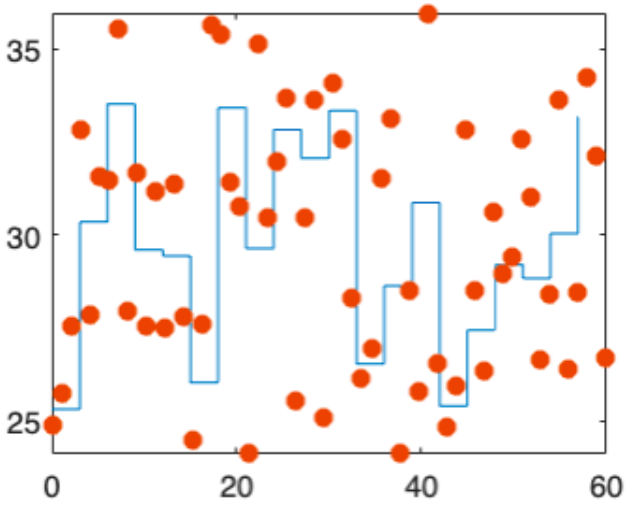
Implementation of PAA

The general idea of this implementation is to reduce dimensionality of the inputted time series by splitting them into equal-sized segments, which are computed by averaging the values in those segments.

Hypothesis of PAA with one time series:



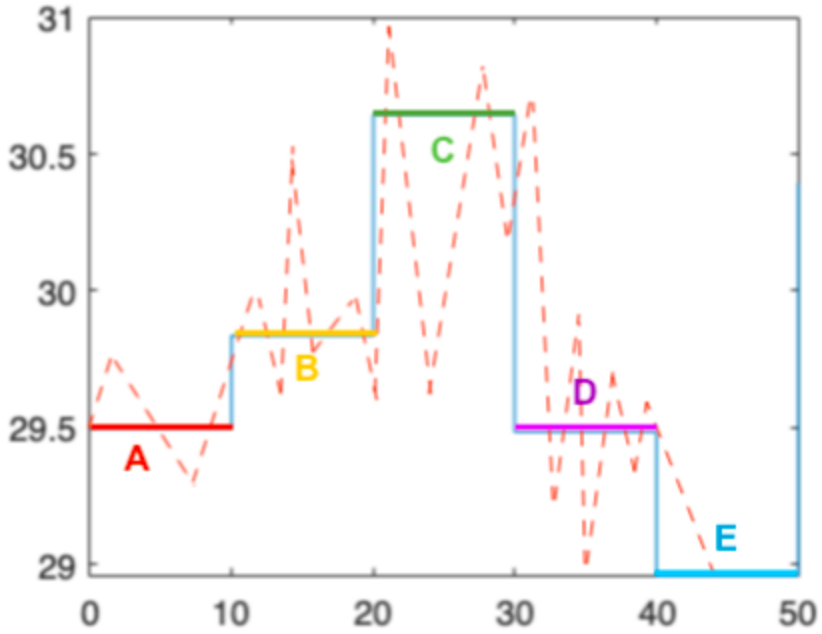
Actual implementation of PAA with one time series:



Implementation of SAX

SAX is a symbolic representation for time series that allows for dimensionality reduction and indexing with a lower-bounding distance measure.

Hypothesis of SAX with one time series:



Euclidean and Manhattan Distances

Algorithms, be it supervised or unsupervised, use distance measures. These measures, including Euclidean, Hamming, Minkowski and Manhattan can be found in algorithms like k-NN.

k-NN is a technique that is often used for supervised learning, and uses Euclidean distance as its default.

Euclidean Distance

Euclidean distance is a measure that can be explained as the length of a segment connecting two points.

The formula is represented by:

$$D(x, y) = \sqrt{\sum_{i=1}^{n} ((x_i - y_i)^2)}$$

Manhattan Distance

The Manhattan distance calculates the distance between real-valued vectors. Manhattan distance refers to the distance between the two vectors if they could move at right angles.

The formula is represented by:

$$D(x, y) = \sum_{i=1}^{k} |x_i - y_i|$$