

Eric Yang

CSE 454 Project 2

Time Series Data Analysis

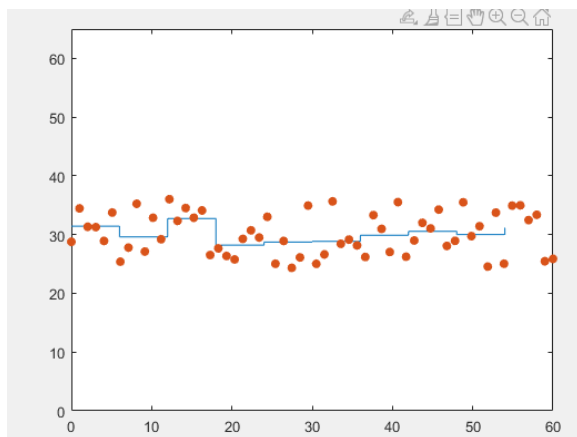
Fall 2021

Description:

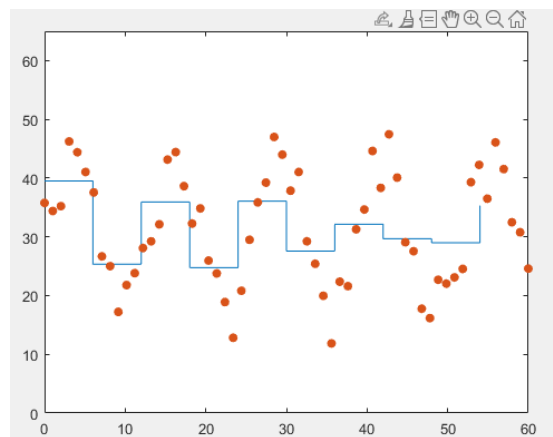
Creates a representation of a temporal dataset using Piecewise Aggregate Approximation and Symbolic Aggregate Approximation. Classification of the datasets are done using PAA and the original unmodified data.

PAA Implementation:

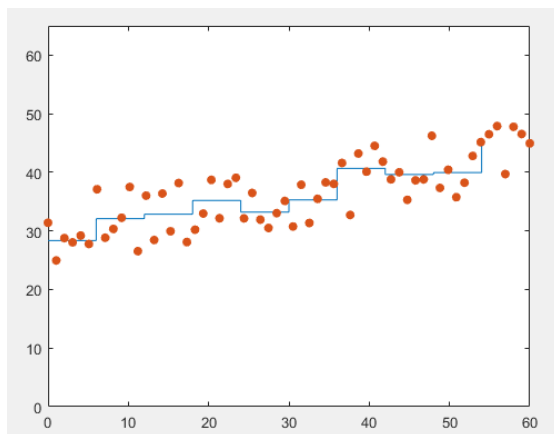
Each row/set of data is split along the temporal axis by the number of columns. The data points in each segment is averaged together and sequentially stored in a matrix with each row containing the other segments from the set of data. Each row of the resulting matrix would contain the PAA values of the respective time series.



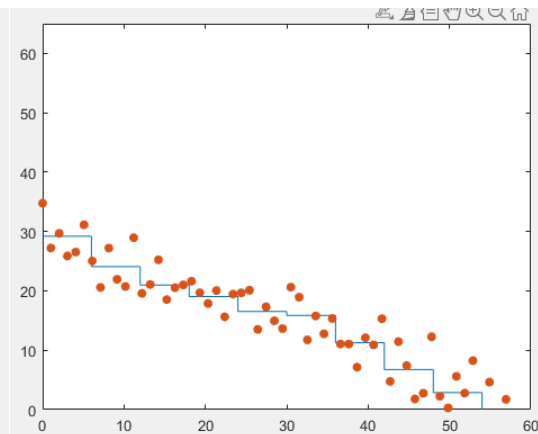
Index 1 – Normal



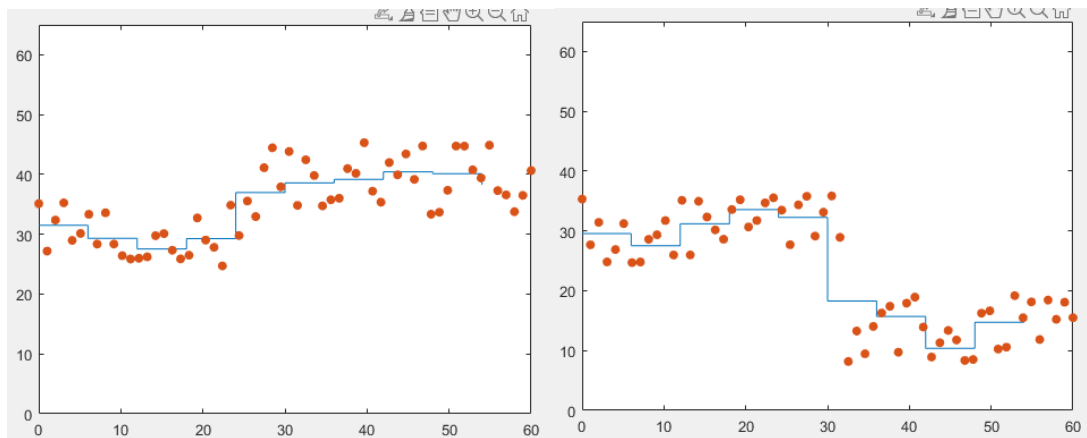
Index 91 – Cyclic



Index 181 – Increasing Trend



Index 271 – Decreasing Trend



Index 361 – Upward Shift

Index 451 – Downward Shift

SAX Implementation:

The mean and standard deviation of each row in the PAA matrix is calculated. this is used to adjust the cut points of the letter differentiation without the need to normalize the data to average zero beforehand. The PAA data points are then mapped to the cut points by row and then further mapped into the 5 letters: a, b, c, d, e used to represent the data.

Data Preprocessing:

The last 10 sets of each class of charts are separated to create the testing set. The remainder are used for the training set. Labels are created to be in line with each class in both sets.

Classification:

Using either Euclidean or Manhattan Distance formulas, the training set is compared to each entry in the testing set with each value in the temporal axis acting as a dimension/position in Euclidean space. The results of the calculations are mapped to a matrix and the index of the minimum distance for each row is retrieved. The index corresponds to the index of the testing set that best matched the entry in the training set and the corresponding testing set label is assigned. Accuracy is determined by the comparison of the assigned label to the true labels of the training set.