Jesse Both
Time Series Check In

Overview

The purpose of this assignment is to gain experience exploring representation and classification techniques. The techniques that will be used are PAA and SAX. The classification will be performed using Euclidean and Manhattan distance equations. It will then be determined which equation is better for the provided set.

Ask

Objective:

Create 2 representation data sets utilizing PAA and SAX techniques.

Inputs:

- Data Set
- c - number of dimensions

Outputs:

- Array with the calculated values.from PAA or SAX
    - Pass that array into the distance calculation to create a confusion matrix.

Constraints:

- Use the Synthetic Control data set that is provided.
- Classification techniques
    - Euclidean Distance
    - Manhattan Distance

Research

Definitions:

PAA - Piecewise Aggregate Approximation
SAX - Symbolic Aggregate approXimation

Purpose:

The purpose of these representations is to simplify a scatter plot into something that a computer can understand. Humans are able to look at a chart and see the trends, but computers need help with this. The use of PAA and SAX reducees the dimensionality with a strict set of rules that are defined.

Plan

- Start with creating a PAA plot to visualize the data.
- Modify the PAA function to return data that is usable.

- Utilize the required distance methods to determine the closeness of the data.
- Create the confusion matrix.
- Compare the confusion matrices for the Euclidean and Manhattan distances.
- Do the same steps for the SAX implementation.