

# Time Series Assignment

Assigned: Oct 4, 2021

Planning Due: Oct 15, 2021 at 11:59pm

Due: Nov 1, 2021 at 11:59pm

You will be exploring representation and classification. This will be done using a synthetic control data set from the University of California Irvine. The set has 60 columns of data (60-time points) with values between 0 and 100 and 6 different classes. It is grouped in order by its classification.

<https://archive.ics.uci.edu/ml/datasets/Synthetic+Control+Chart+Time+Series>

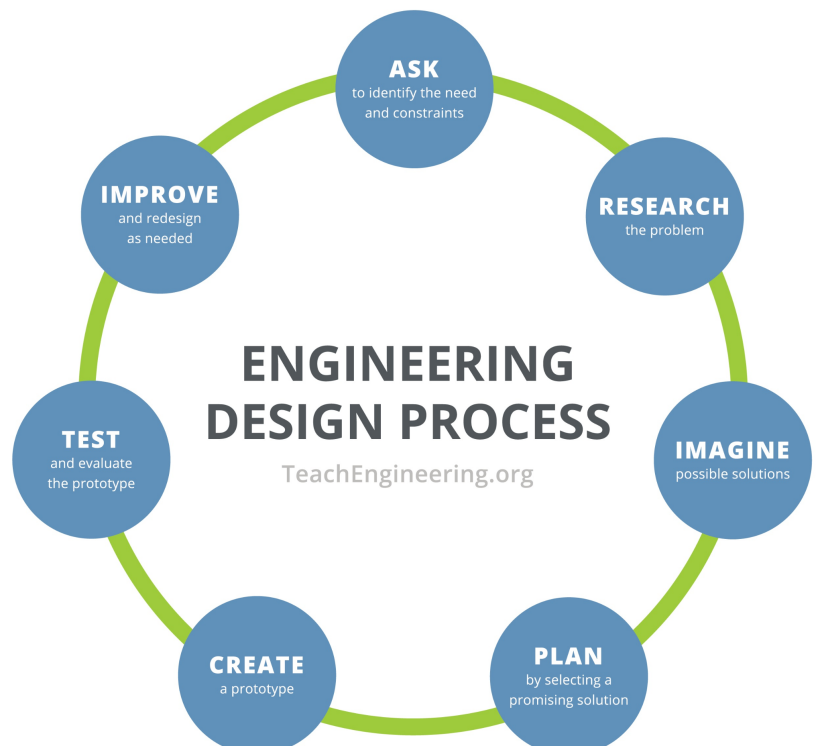
Create two new representation data sets that reduce the number of samples needed to represent the data. The first set will utilize PAA and the second will use SAX.

Explore classification behaviors with the original and PAA sets. Use an “expert” approach for training. Perform the classification using Euclidean Distance and Manhattan Distance. Compare which is better on each set. Are there any trends observed?

If you use any libraries for either representation/classification task, it must also be documented. To give you the opportunity to document and communicate what you did, you will submit your answers in a report form in addition to submitting your code.

## Planning (20 Points)

Create a document that includes the key planning elements that are needed to support the successful implementation of this project. This document must include the overview/objective summary. Then consider the following engineering design process diagram to help you determine what planning should be done prior to attempting to implement. Submit it as a PDF.



## Code submission/requirements (50 Points)

1. Processing to create PAA
2. Processing to create SAX
3. Classification using original data
4. Classification using PAA data set
5. Proper commenting

Submit as your single MatLab file. If you have more than one file or UB Learns doesn't like the MatLab file, zip them together and submit the zipped file.

## Report (80 Points)

Questions to be addressed in the report:

1. Describe the approach used to implement PAA on this data set
2. Plot your original time series and the PAA representation
3. Explain your approach to create the SAX implementation, including the alphabet used
4. Plot your original time series data and your SAX representation
5. Describe the processing process to establish your training and testing data generation
6. Explain your classification process and results

The report should be appropriate technical writing about the project, not about you. The overall writing quality of the report will factor in for 10 points. The overall base elements in a report must also be present and appropriate, for 10 points.

I gave up, I don't know, I ran out of time...All unacceptable to be in the report and read me. If it occurs in any of your documents that is where grading ends.

Submit as a PDF.

## Reflection Blog Post (10 Points)

- Describe your experience
- Challenges
- Easy things
- At least 1 spot identified for potential improvement

Submit as a PDF.

## Showcase (20 Points)

Create a GitHub Page that is well organized to show off this work. Include your planning stuff, a readme to describe the functionality, and report.... Other stuff is up to you. Submit the link to your repository with the rest of the submission.