Homework 6

ECE 631A Cyber-Security of Critical Infrastructure

Due Date: November 16, 2019 by 11:55 PM via canvas

This homework is for you to use a ***diffusion map-based dimensionality reduction*** followed by any unsupervised classification algorithm -- to create a model to distinguish between stressed grid condition and nominal grid condition. You have had used the same data for your Homework 4. So, the data should be familiar to you. How the data is generated is described in the paper "Methodology for a Security/Dependability Adaptive Protection Scheme Based on Data Mining" by Bernabeu, Thorp, Centeno. Check particularly in Section IV A of the paper to know how the data was generated.

Refer to the paper: An Introduction to Diffusion Maps, by J. de la Porte, B. M. Herbst, W. Hereman, S. J. van der Walt – covered in class – now available under Files -> Week12

The very first column in the data table is marked as '0' if it is a normal condition data and marked '1' if it is a stressed condition data. The data represents positive sequence voltages and angles at various substations, and line voltage/currents for different transmission lines.

However, for this homework, you need to only select those rows that are marked 0, as you are supposed to work with unsupervised learning – which means the data pertaining to the nominal condition should be used to train your model. After you have trained your unsupervised model – you need to use the data marked '1', as well as 25% of the data you set aside from data marked '0' to test the efficacy of your model.

We will be testing on data that is randomly picked from the given data to test your model and see the accuracy and false negative figures reported by you are in the ball park range.

The final submission should contain a pdf file with brief description of your methods and the model used followed by a table with accuracy, false positive, false negative etc. as customary.

You must submit the entire code, the pdf containing report as stated, and any other relevant files to run your model – along with a README file detailing how to run your model.

A group of up to 3 people can submit one homework if they work together. The TAs will interview the entire team to understand the contributions of all the members and ask the team members to run the model and show results.

The grading scheme will as follow:

Modeling including coding, and efficiency of the training and the actual classification on a few data: 60 points

Accuracy, Quality and details in the report: 20 points

Interview of team members during the demo run: 20 points (this will be different for each team member)

The data is available at canvas website in Files → Week4 and Homework4 files.