Part II: Term Project Report

The term project report presents, illustrates and explains your experimental analysis and interprets the results for the electricity consumption datasets provided for this purpose. In addition to plain text, this requires diagrams, graphs and tables showing the experiments performed and comparing the outcomes. Specifically, this includes:

1) Comparing the importance of each response variable relative to the other ones as the result of performing PCA and illustrating the rational for your final of response variables;
2) Explaining the selection of response variables and the time window chosen for the analysis;
3) An overview of the log-likelihood and BIC values for different numbers of HMM states to show to justify your selection of most reasonable models;
4) Showing your choice for partitioning the data into train data and test data;
5) Comparing the normalized training log-likelihood and test log-likelihood in order to show how good your model fits the data;
6) Illustrating the anomalies across the three different data sets with injected anomalies.

The project report is to be completed and submitted by April 12, 2021. Oral presentations of the essential project outcomes as presented in the report follow on April 13, 14, and 16. Each group will have about 20 mins. (including Q&A) for presenting their work. A copy of the presentation slides needs to be submitted by April 12 as well.

OVERVIEW

Project Scope. Automation enhances cost efficiency, quality of service delivery and safe operation of critical assets. Electric power grids, public water utilities and smart transportation networks routinely rely on supervisory control systems, with steadily increasing integration of computation, networking and physical processes. Increasing reliance on automation also increases the attack surface for advanced persistent threats and amplifies the risk of cascading effects. Existing vulnerabilities expose critical infrastructure to a range of adversarial scenarios. The project explores anomaly detection based intrusion detection methods used for cyber situational awareness in the analysis of automated control processes.

Challenges. A number of inescapable 'external factors' make anomaly detection in time series data streamed from the operation of a real-world supervisory control system challenging. Typical examples include: imperfections in the data, such as missing or corrupted values; lack of ground truth in historic data, unavailability of labels to differentiate normal observations from outliers; types of anomalies depending on the particular application context; striking a good balance between *precision* and *recall*, specifically also reducing the false alarm rate to make anomaly detection practical in any real application context with resource constraints.

PROJECT REPORT

The report documents your team's work on the term project and the essential outcomes. Technical reports are routinely used in industry for communicating ideas, facts, problem descriptions and possible solutions for a technical subject matter. Common standards expected from a professionally written technical report are detailed below.

The term project report explains and illustrates at a technical level: (1) the **problem** being addressed; (2) the **methodology** used for solving the problem; (3) the **characteristics** of the solution and a **rational** for the underlying design choices; (4) major **problems** encountered in the course of the project; and (5) the **lessons learned**.

Technical writing is about a particular technical subject that requires direction, instruction, or explanation. This style of writing serves a different purpose and has different characteristics than other writing styles such as creative writing, academic writing or business writing. It is a clear and efficient way of explaining something and how it works.

Project Report Structure

Proper logical organization and clear structuring of the project report calls for:

- a **title page** containing a title, name of all authors, student ID numbers, the course and semester, an abstract (i.e., a one paragraph outline of your report);
- some concise but meaningful **conclusions** (e.g., what you have accomplished, future work);
- page numbers and **numbered headings** of sections, subsections, etc.;
- a **table of contents** and a table of figures;
- a **list of references** (i.e., bibliographic items).

Note that online references are perfectly acceptable; you may want to give references to web pages or online documents or a reference to a specific web sub-page if referencing a particular point from that particular link.

Example of a bibliographic item:

> Zahra Zohrevand, Uwe Glässer, Mohammad A. Tayebi, Hamed Yaghoubi Shahir, Mehdi Shirmaleki, and Amir Yaghoubi Shahir. Deep-Learning Based Forecasting of Critical Infrastructure Data. *In Proceedings of the 26th ACM International Conference on Information and Knowledge Management*, Singapore (2017), pages 1129-1138.

Using illustrations (diagrams, graphs) and tables to complement explanations in plain text enhances the clarity of the presentation. The **body** of your report (excluding the title page, table of contents, list of references, etc.) should be about 16 pages double spaced including figures and tables. It should start by introducing the **problem scope** and **technical background**, and provide a basic rational for the concepts on which your solution builds. List the main contributions to the project and the report of **each team member**.

We hope that you will find this project a rewarding experience.

Thank you for your cooperation!