## Title: Learning From Time Series Sensor Data

**Project Description:** Many applications, ranging from traffic data to financial markets and wearable sensors in health analytics, generate large amounts of time series data. In most cases, this data is multivariate, where each timestamped observation consists of readings from multiple entities, or sensors. These long time series can often be broken down into a sequence of states, each defined by a simple pattern, where the states can reoccur many times. For example, raw sensor data from a fitness tracking device can be interpreted as a temporal sequence of actions (i.e., walking for 10 minutes, running for 30 minutes, sitting for 1 hour, then running again for 45 minutes). This representation can be used to discover repeated patterns, understand trends, detect anomalies and more generally, better interpret large and high-dimensional datasets. To achieve this representation, it is necessary to simultaneously segment and cluster the time series. This problem is more difficult than standard time series segmentation, since multiple segments can belong to the same cluster.

In this project, we propose to study a newly developed method for multivariate time series clustering, called Toeplitz inverse covariance-based clustering. In this method, we learn each cluster by estimating a sparse Gaussian inverse covariance matrix; and the Toeplitz constraint ensures that the cluster definitions are time-invariant, so the clustering assignment does not depend on the exact starting position of the subsequence. To solve this problem, an expectation maximization (EM)-like approach, based on alternating minimization, is used.

**Duties/Activities:** The intern will run and test different instances of the machine learning code on real data. The code in python will be provided, as well as the time series data (health and traffic).

**Required Skills:** Python

**Preferred Intern Academic Level:** B.Sc.

**Learning Opportunities:** The intern will enhance his/her programming skills in Python and acquire new knowledge in machine learning in particular in unsupervised learning for time series data on real problems (traffic analytics and health analytics).

**Expected Team Size:** 1 or 2 students

**Mentors**

Dr. Abdelkader Baggag email: abaggag@hbku.edu.qa

Dr. Michael Aupetit email: maupetit@hbku.edu.qa