Data Visualization
Supervised Dimensionality Reduction Methods for Visualization of High-Content Data

Supervisors
Olivier Morère – olivier.morere@gmail.com
Antoine Veillard – antoine.veillard.2004@polytechnique.org

Keywords
Supervised Machine Learning; Data Visualization; Big-Data; Spectral Methods; Topological Methods; Kernels

Scope of work
With the ongoing big-data trend, much of the recent research effort in the field of machine learning was spent on the adaptation of traditional learning algorithms for their use with high-dimensional and/or heterogeneous data. Such methods are often intended as black-boxes to be used with large training sets which have to be constructed with a specific task in mind.

Meanwhile, the current trend in the industry is increasingly to the accumulation of as much data as realistically possible without a clear prior purpose, whereas the question of the actual use of the data is relegated to when the need arises. In this context, the need for exploratory data analysis tools is a strong requirement which is not adequately satisfied by the current state-of-the-art.

During this internship, the candidate will work on the design and the implementation of data-visualization tools based on dimensionality reduction methods (either spectral or topological) in order to provide various specific insight into the same data with minimal inputs from the user. Datasets from a computer-aided diagnosis project (FLexMIm) will provide a validation framework for the proposed methods.

Required profile
- Engineer degree or master degree student (final year)
- Strong background in mathematics and machine learning
- Experience with technical computing languages (MatLab/Python)

Duration 6 months

Gratification Approx. 800 € / 1,400 SGD per month