Internship : deep clustering using variational autoencoder

Laboratoire ERIC, Université Lyon 2, in collaboration with EDF

1. Context

Clustering is the task of organizing similar objects into meaningful groups. With the big data phenomenon, modern data are now high dimensional and /or heterogeneous. This provides new challenges and there is a need to develop new clustering methods adapted to such data. One approach for clustering such data relies on data embedding, which consists in embedding the original data into a low-dimensional space in which the clustering can be easily performed. Recently, deep learning models have been successfully used in many domains, and in particular for the task of data embedding.

2. Subject

The goal of the internship is to develop a clustering approach based on deep learning models. The main missions are :

- to study the recent development in clustering methods using deep learning
- to develop a model on the basis of an original idea proposed by the internship supervisor,
- to test this model on simulated data and onto data provided by EDF (electricity consumption) and Thalès.

A publication presenting the model will be written during the internship. The intern candidate should have high skills in statistical learning, machine learning and R programming. A knowledge of Tensorflow framework would be appreciated.

3. References

DLC2018: IEEE ICDM Workshop on deep learning and clustering, November 17th 2018 , Singapore

N. Dilokthanakul, P. A. M. Mediano, M. Garnelo, M.C. H. Lee, H. Salimbeni, K. Arulkumaran & M. Shanahan, Deep unsupervised clustering with gaussian mixture variational autoencoders, working paper 2017.

4. Internship conditions

Location : the intern will join the Data Mining & Decision team of the ERIC lab., which is composed of 11 permanent researchers in statistics and computer science.
Duration: 6 months, starting in March 2020
Salary: approx. 550€ / month
Contact: julien.jacques@univ-lyon2.fr