
Optimal transport-based machine learning to match specific expression patterns in omics data

Thi Thanh Yen Nguyen^{*1}, Antoine Chambaz¹, Olivier Bouaziz¹, Warith Harchaoui¹,
and Christian Neri²

¹Mathématiques Appliquées Paris 5 – Université de Paris : UMR8145 – France

²CNRS UMR 8256, INSERM ERL U1164 – Sorbonne Université, Brain-C Lab, Paris – France

Résumé

We present two algorithms designed to learn a pattern of correspondence between two data sets in situations where it is desirable to match elements that exhibit an affine relationship. In the motivating case study, the challenge is to better understand micro-RNA (miRNA) regulation in the striatum of Huntington's disease (HD) model mice. The two data sets contain miRNA and messenger-RNA (mRNA) data, respectively, each data point consisting in a multi-dimensional profile. The biological hypothesis is that if a miRNA induces the degradation of a target mRNA or blocks its translation into proteins, or both, then the profile of the former should be similar to minus the profile of the latter (a particular form of affine relationship).

The algorithms unfold in two stages. During the first stage, an optimal transport plan P and an optimal affine transformation are learned, using the Sinkhorn-Knopp algorithm and a mini-batch gradient descent. During the second stage, P is exploited to derive either several co-clusters or several sets of matched elements.

A simulation study illustrates how the algorithms work and perform. A brief summary of the real data application in the motivating case-study further illustrates the applicability and interest of the algorithms.

^{*}Intervenant