

A Brief Report on Minibatch OT

Logistics

The minibatch method calculates the average of transport maps between smaller samples selected from the source and target distributions, and gets the OT map between the population source and target. The main advantage of considering this method is that we can compute maps for large datasets using smaller samples, which makes the process more computationally efficient.

The tradeoff with this algorithm is between the accuracy and computing costs. If the size of each batch is small, the computation will take less time, but the resulting OT map will be 'blurred out'.

The minibatch method has 'emd' and 'entropic' modes, meaning that a single transport map between the mini batch can be calculated by either the `ot.emd` function or the `ot.sinkhorn` function. The 'emd' option gives a much clearer OT map, but the computation takes longer and the sample size requirements are severe for high-dimensional data.

Issues

- Sample size requirements for the EMD option:

The reason we considered the minibatch method was because `ot.emd` does not converge when sample size(n) is small compared to sample dimension(d). However, if we use the minibatch method, batches that are used to compute maps are smaller than the population, which means that the chances of the emd function converging is not increased by the use of the minibatch method. The use of the minibatch method is only justified when we have a dataset in which the mini batches are also large enough for the emd function to converge.

- Sinkhorn method:

The sinkhorn method converges if the data is large enough for its dimensions. However, sinkhorn in turn is so efficient that it does not need the help of the minibatch method; for instance, computing the sinkhorn solution for $n = 25000$, $\text{dim} = 40$ takes 30 seconds on the Great Lakes cluster. Therefore, we also do not need to consider the minibatch method unless we are considering a ridiculously large dataset.

To Do

- Decide whether the use of minibatch method is valid