Generalized Projection Sinkhorn Divergences Conference Submissions

Anonymous authors
Paper under double-blind review

Abstract

A principled way of comparing probability distributions, that has seen recent success in machine learning are optimal transport distances (OT). However, they suffer from the curse of dimensionality as they scale cubically with the number of dimensions. Recent work that compares discrete measures in high-dimensions proposes maximizing the linear subspace projections and shows desirable robustness properties. In this work we extend their approach, showing that we can approximate robust OT distances by maximizing the pullback of the L2 distance through a general family of transformations f_ϕ, such as a neural network, that need only be 1-Lipschitz. This can be seen as learning a parametrized "ground" cost function and computing OT distances in a much lower dimensional space, $k << d$. Our method scales with the measures sample size n and the OT problem depends only on f_ϕ and k, making it constant in d and extremely efficient. Unbiased Sinkhorn divergences (SD) have been proven to interpolate between MMD and true OT distances and have the added benefit of differentiability and quadratic computational time. We use SD to demonstrate empirically the robustness and efficiency of our technique. Furthermore, we show how our approach can be used to train an OT-GAN on the MNIST dataset where the discriminator naturally clusters the data.