OT

Kouamé Kra

December 2024

1 Introduction

Entropic Regularization of Optimal Transport

We consider two input histograms $p, q \in_N$, where we denote the simplex in R^N:

$$_N p \in (R^+)^N \sum_i p_i = 1.$$

We consider the following discrete regularized transport:

$$W_\gamma(p,q)\pi \in \Pi(p,q)C\pi - \gamma E(\pi).$$

where the polytope of couplings is defined as:

$$\Pi(p,q)\pi \in (R^+)^{N\times N}\pi = p,\ \pi^T = q,$$

and $(1,\ldots,1)^T \in R^N$. For $\pi \in (R^+)^{N\times N}$, we define its entropy as:

$$E(\pi) - \sum_{i,j} \pi_{i,j}\left(\log(\pi_{i,j}) - 1\right).$$

When $\gamma = 0$, one recovers the classical (discrete) optimal transport. We refer to the monograph [Villani] for more details about OT. The idea of regularizing transport to allow for faster computation was introduced in [Cuturi].

Here, the matrix $C \in (R^+)^{N\times N}$ defines the ground cost, i.e., $C_{i,j}$ is the cost of moving mass from a bin indexed by i to a bin indexed by j.

The regularized transportation problem can be rewritten as a projection:

$$W_\gamma(p,q) = \gamma\pi \in \Pi(p,q)\pi\xi,$$

where:

$$\xi_{i,j} = e^{-\frac{C_{i,j}}{\gamma}}.$$

The Kullback-Leibler divergence between $\pi, \xi \in (R^+)^P$ is:

$$\pi\xi = \sum_{i,j} \pi_{i,j}\left(\log\left(\frac{\pi_{i,j}}{\xi_{i,j}}\right) - 1\right).$$

This interpretation of regularized transport as a KL projection and its numerical applications are detailed in [BenamouEtAl].

Given a convex set $\mathcal{C} \subset R^N$, the projection according to the Kullback-Leibler divergence is defined as:

$$_{\mathcal{C}}(\xi) = \pi \in \mathcal{C}\pi\xi.$$

—

Iterative Bregman Projection Algorithm

Given affine constraint sets $(\mathcal{C}_1, \ldots, \mathcal{C}_K)$, we aim to compute:

$$_{\mathcal{C}}(\xi) \quad where \quad \mathcal{C} = \mathcal{C}_1 \cap \ldots \cap \mathcal{C}_K.$$

This can be achieved, starting with $\pi_0 = \xi$, by iterating:

$$\forall \ell \geq 0, \quad \pi_{\ell+1} =_{\mathcal{C}_\ell} (\pi_\ell),$$

where the index of the constraints should be understood modulo K, i.e., $\mathcal{C}_{\ell+K} = \mathcal{C}_\ell$.

One can show that $\pi_\ell \to_{\mathcal{C}} (\bar{\pi})$. For more details on this algorithm and its extension to compute the projection on the intersection of convex sets (Dijkstra algorithm), see [BauschkeLewis].

—

Iterative Projection for Regularized Transport (Sinkhorn's Algorithm)

We re-cast the regularized optimal transport problem within this framework by introducing:

$$\mathcal{C}_1 \pi \in (R^+)^{N \times N} \pi = p, \quad \mathcal{C}_2 \pi \in (R^+)^{N \times N} \pi^T = q.$$

The KL projection on $\mathcal{C}_1$ is computed by divisive normalization of rows:

$$\forall (i,j), \quad \pi_{i,j} = \frac{p_i \bar{\pi}_{i,j}}{\sum_s \bar{\pi}_{i,s}},$$

and similarly for $_{\mathcal{C}_2}(\bar{\pi})$ by replacing rows with columns.

A key observation is that if $\bar{\pi} = (a)\xi(b)$ (a diagonal scaling of the kernel ξ), then:

$$_{\mathcal{C}_1}(\bar{\pi}) = (\tilde{a})\xi(b), \quad _{\mathcal{C}_2}(\bar{\pi}) = (a)\xi(\tilde{b}),$$

where the new scaling reads:

$$\tilde{a} = \frac{p}{\xi b}, \quad \tilde{b} = \frac{q}{\xi^T a}.$$

This means the iterates of Bregman projections, starting with a_0, always have the form:

$$\pi_\ell = (a_\ell)\xi(b_\ell),$$

and the diagonal scaling weights are updated as:

$$a_{\ell+1} \frac{p}{\xi b_\ell}, \quad b_{\ell+1} \frac{q}{\xi^T a_{\ell+1}}.$$

This algorithm is the well-known Sinkhorn algorithm [Sinkhorn].

—

Transport Between Histograms

We consider two measures p, q on a uniform grid $x_i = y_i = i/N$. These are often referred to as "histograms".

Histogram Definitions

- Size N of the histograms. - Define p, q as translated Gaussians. - Add minimal mass and normalize.

Gibbs Kernel

The Gibbs kernel is a Gaussian convolution:

$$\xi_{i,j} = e^{-\frac{(i/N - j/N)^2}{\gamma}}.$$

Sinkhorn Algorithm

- Initialize $b_0 =_N$. - Perform iterations of:

$$a_{\ell+1} \frac{p}{\xi b_\ell}, \quad b_{\ell+1} \frac{q}{\xi^T a_{\ell+1}}.$$

Barycentric Projection

Compute the transport plan approximation via the barycentric projection map:

$$t_i \mapsto s_j \frac{\sum_j \pi_{i,j} t_j}{\sum_j \pi_{i,j}} = \frac{[a \odot \xi(b \odot t)]_j}{p_i},$$

where $\odot$ is element-wise multiplication.

Visualization

- Display the coupling matrix in the log domain. - Overlay the transport map (barycentric projection).