Optimal transport for Topological Data Analysis

Final course thesis

Gonzalo Ortega

Wednesday 22$^{\text{nd}}$ 2025

Contents

Definition (Push-forward measure)

Let $T : X \to Y$ be a Borel map, and $\mu \in \mathcal{P}(X)$. Let $A \in$ Bar. The *push-forward measure* $T_{\#}\mu \in \mathcal{P}(Y)$ is defined as

$$T_{\#}\mu(A) := \mu(T^{-1}(A)).$$

Definition (Transport map)

Given $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, a *transport map from μ to ν* is a Borel map $T : X \to Y$ that satisfies $T_{\#}\mu = \nu$.

Definition (Transport plan)

Let $\pi_X : (X \times Y) \to X$ and $\pi_Y : (X \times Y) \to Y$ such that for every $(x, y) \in (X, Y)$, $\pi_X(x, y) = x$ and $\pi_Y(x, y) = y$. A *transport plan between* μ *and* ν is a probability measure $\gamma \in \mathcal{P}(X \times Y)$ where

$$(\pi_X)_\#\gamma = \mu \text{ and } (\pi_Y)_\#\gamma = \nu.$$

The set of all couplings between μ and ν is denoted $\Gamma(\mu, \nu)$.

Definition (Transport problems)

Fix $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and consider a lower semicontinous map $c : X \times Y \to [0, \infty]$. Then

$$C_M(\mu, \nu) := \inf \left\{ \int_X c(x, T(x)) d\mu(x) : T_{\#}\mu = \nu \right\},$$

$$C_K(\mu, \nu) := \inf \left\{ \int_{X \times Y} c(x, y) d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu) \right\}.$$

Wasserstein distance over probability spaces

Definition (Probability measures with finite p-moment)

Let (X, d) be a locally compact and separable, metric space. Let $1 \leq p < \infty$. The *set of probability measures with finite p-moment* is defined As

$$\mathcal{P}_p(X) := \left\{ \sigma \in \mathcal{P}(X) : \int_X d(x, x_0)^p d\mu(x) < \infty \text{ for some } x_0 \in X \right\}.$$

Definition (p-Wasserstein distance)

Given $u, v \in \mathcal{P}_p(X)$, the *p-Wasserstein distance* is defined as

$$W_p(u, v) := \left(\inf_{\gamma \in \Gamma(u,v)} \int_{X \times X} d(x, y)^p d\gamma(x, y) \right)^{\frac{1}{p}}.$$

Contents

Definition (Persistence diagram)

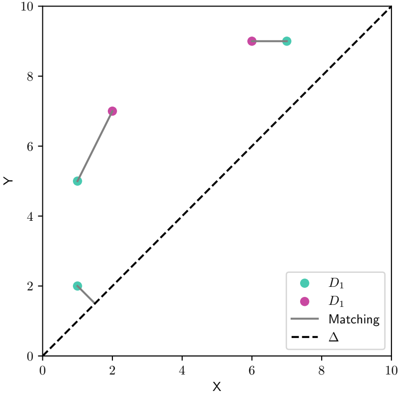
Let I be a countable set. A *persistence diagram* is a function $D : I \to \mathbb{R}^2_<$.

Definition (Partial matching)

Let $D_1 : I_1 \to \mathbb{R}^2_<$ and $D_2 : I_2 \to \mathbb{R}^2_<$ be persistence diagrams. A *partial matching* between D_1 and D_2 is the triple (I_1', I_2', f) such that $f : I_1' \to I_2'$ is a bijection with $I_1' \subseteq I_1$ and $I_2' \subseteq I_2$.

Wasserstein distance over persistence diagrams

Tiny diagrams matching example



Definition (Chebyshev distance)

Let $a, b \in \mathbb{R}^2$ with $a = (a_x, a_y)$ and $b = (b_x, b_y)$. The *Chebyshev distance* is defined as

$$d_\infty(a, b) := ||a - b||_\infty := \max\{|a_x - b_x|, |a_y - b_y|\}.$$

Definition (p-cost)

Let $D_1 : I_1 \to \mathbb{R}^2_<$ and $D_2 : I_2 \to \mathbb{R}^2_<$ be persistence diagrams. Let (I_1', I_2', f) be a partial matching between them. If $p < \infty$, the *p-cost of f* is defined as

$$\mathrm{cost}_p(f) := \left(\sum_{i \in I_1'} d_\infty(D_1(i), D_2(f(i)))^p + \sum_{i \in I_1 \setminus I_1'} d_\infty(D_1(i), \Delta)^p + \sum_{i \in I_2 \setminus I_2'} d_\infty(D_2(i), \Delta)^p \right)^{\frac{1}{p}}.$$

For $p = \infty$, the *∞-cost of f* is defined as

$$\mathrm{cost}_\infty(f) := \max \left\{ \sup_{i \in I_1'} d_\infty(D_1(i), D_2(f_i)), \sup_{i \in I_1 \setminus I_1'} d_\infty(D_1(i), \Delta), \sup_{i \in I_2 \setminus I_2'} d_\infty(D_2(i), \Delta) \right\}.$$

Definition (p-Wasserstein distance)

Let D_1, D_2 be persistence diagrams. Let $1 \leq p \leq \infty$. Define

$$\tilde{\omega}_p(D_1, D_2) = \inf\{\mathrm{cost}_p(f) : f \text{ is a partial matching between } D_1 \text{ and } D_2\}.$$

Let $\emptyset$ denote the unique persistence diagram with empty indexing set. Let $(\mathrm{Dgm}_p, \omega_p)$ be the space of persistence diagrams D that satisfy $\tilde{\omega}_p(D, \emptyset) < \infty$ modulo the equivalence relation $D_1 \sim D_2$ if $\tilde{\omega}_p(D_1, D_2) = 0$. The metric ω_p is called the *p-Wasserstein distance*.

Definition (Bottleneck distance)

If $p = \infty$, the metric ω_∞ is called the *bottleneck distance*.

Definition (Isometric embedding)

Let $(X, d_X), (Y, d_Y)$ be metric spaces. An *isometric embedding* $\eta : (X, d_X) \to (Y, d_Y)$ is a mapping that satisfies

$$d_X(x_1, x_2) = d_Y(\eta(x_1), \eta(x_2))$$

for all $x_1, x_2 \in X$.

Definition (Ball in persistence diagrams)

Let $1 \leq p \leq \infty$. Let $D_0 \in \mathrm{Dgm}_p$. The *ball* at the space of persistence diagrams is defined as $B_p(D_0, r) := \{D \in \mathrm{Dgm}_p : w_p(D, D_0) < r\}$.

Theorem (Isometric embeeding of metric spaces into persistance diagrams)

Let (X, d) be a separable, bounded metric space. Then there exists an isometric embedding to the space of persistence diagrams $\eta : (X, d) \to (\mathrm{Dgm}_\infty, \omega_\infty)$ such that $\eta(X) \subseteq B(\emptyset, \frac{3c}{c}) \backslash B(\emptyset, c)$.

As (X, d) is bounded, we can let $c > \sup\{d(x, y) : x, y \in X\}$. As (X, d) is separable, we can take $\{x_k\}_{k=1}^{\infty}$, a countable, dense subset of (X, d). Consider

$$\begin{aligned} \eta : (X, d) &\to (\mathrm{Dgm}_\infty, \omega_\infty) \\ x &\mapsto \{(2c(k-1), 2ck + d(x, x_k))\}_{k=1}^{\infty} \end{aligned}$$