Continuous 1-Wasserstein Distance

Petr Mokrov

Deep Generative Models course at MIPT

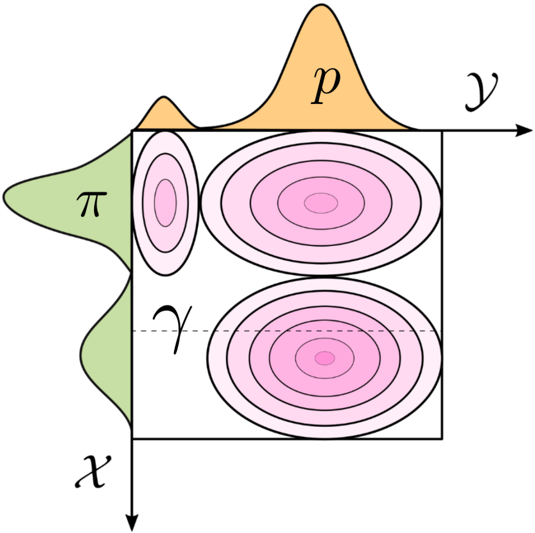
2023

1-Wasserstein distance

$$W_c(\pi, p) = \inf_{\gamma \in \prod(\pi, p)} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\gamma} c(x, y) = \inf_{\gamma \in \prod(\pi, p)} \int c(x, y)\gamma(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y}$$

- $\prod(\pi, p)$ – the set of all joint distributions $\gamma(\mathbf{x}, \mathbf{y})$ with marginals π and p ($\int \gamma(\mathbf{x}, \mathbf{y}) d\mathbf{x} = p(\mathbf{y})$, $\int \gamma(\mathbf{x}, \mathbf{y}) d\mathbf{y} = \pi(\mathbf{x})$)
- $\gamma(\mathbf{x}, \mathbf{y})$ – transportation plan (the amount of "dirt" that should be transported from point $\mathbf{x}$ to point $\mathbf{y}$).
- $\gamma(\mathbf{x}, \mathbf{y})$ – the amount, $c(x, y)$ – cost (not necessary to be metric).
- Of our interest is metric case: $\mathbf{x} \in X = \mathbb{R}^D$, $\mathbf{y} \in Y = \mathbb{R}^D$, $c(x, y) = \|x - y\|$.

1-Wasserstein distance



Wasserstein distance and Kantorovich Duality

Theorem 1 (Kantorovich Duality)

Let X and Y be Polish spaces, π an p are probability measures on X and Y, $c(x, y) : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ be a lower semi-continious cost function. Let

$$J(\phi, \psi) = \int_X \phi d\pi + \int_Y \psi dp$$

Then

$$\inf_{\gamma \in \prod(\pi, p)} \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \gamma} c(x, y) = \sup_{\phi(x) + \psi(y) \leq c(x, y)} J(\phi, \psi)$$

C. Villani: Topics in optimal transportation

Kantorovich Duality. Insight 1

Definition

$c(x, y) : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ be a lower semi-continious cost function. Let $\phi : X \to \mathbb{R} \cup \{-\infty\}$. Then $\psi(y) = \inf_{x \in X}(c(x, y) - \phi(x))$ is **c-transform** of ϕ (denoted as ϕ^c)

One can write the Kantorovich Duality in the following form:

Theorem 1* (Kantorovich Duality)

In conditions of **Theorem 1**:

$$\inf_{\gamma \in \prod(\pi, p)} \mathbb{E}_{(\mathbf{x},\mathbf{y}) \sim \gamma} c(x, y) = \sup_{\phi(x)} J(\phi, \phi^c) = \sup_{\phi(x)} \left(\mathbb{E}_{\mathbf{x} \sim \pi} \phi(x) + \mathbb{E}_{\mathbf{y} \sim p} \phi^c(y)\right)$$

C. Villani: Topics in optimal transportation

Kantorovich Duality. Insight 1

Theorem 1** (Kantorovich Duality)

In conditions of **Theorem 1**:

$$\inf_{\gamma \in \prod(\pi, p)} \mathbb{E}_{(\mathbf{x},\mathbf{y}) \sim \gamma} c(x, y) = \sup_{\phi(x)} J(\phi^{cc}, \phi^{c}) =$$

$$= \sup_{\phi(x)} \left(\mathbb{E}_{\mathbf{x} \sim \pi} \phi^{cc}(x) + \mathbb{E}_{\mathbf{y} \sim p} \phi^{c}(y) \right)$$

Where:

$$\phi^{c}(y) = \inf_{x \in X} (c(x, y) - \phi(x))$$

$$\phi^{cc}(x) = \inf_{y \in Y} (c(x, y) - \phi^{c}(y))$$

C. Villani: Topics in optimal transportation

Kantorovich duality: Insight 2

Let $W_c(\pi, p) = \inf_{\gamma \in \prod(\pi, p)} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\gamma} c(x, y) < +\infty$

Kantorovich duality. Characterization of the optimal potential:

Let:

$$\gamma^* \in \arg \inf_{\gamma \in \prod(\pi, p)} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\gamma} c(x, y)$$

Then there exists $\phi_{\text{opt}} : X \to \mathbb{R}$, such that $\phi^c_{\text{opt}}(y) + \phi^{cc}_{\text{opt}}(x) = c(x, y)$ γ^* - almost surely and :

$$J(\phi^{cc}_{\text{opt}}, \phi^c_{\text{opt}}) =$$
$$= \mathbb{E}_{\mathbf{x}\sim\pi} \phi^{cc}_{\text{opt}}(x) + \mathbb{E}_{\mathbf{y}\sim p} \phi^c_{\text{opt}}(y) = W_c(\pi, p)$$

C. Villani: Optimal Transport: Old and New

Kantorovich duality: metric case

Consider $X = Y = \mathbb{R}^D$ and $c(x, y) = \|x - y\|$

Proposition 1: Lipschitzness

Let $\phi : \mathbb{R}^D \to \mathbb{R}$. Then $\phi^{\|\cdot\|}$ is $1-$**Lipschitz**

Proposition 2: $\|\cdot\|$ - conjugate property:

Let $\phi : \mathbb{R}^D \to \mathbb{R}$. Then $\phi^{\|\cdot\|\|\cdot\|}(x) = -\phi^{\|\cdot\|}(x)$

Kantorovich-Rubinstein duality

$$W_{\|\cdot\|}(\pi, p) = \max_{\|f\|_L \leq 1} \left[\mathbb{E}_{\mathbf{x}\sim\pi} f(\mathbf{x}) - \mathbb{E}_{\mathbf{x}\sim p} f(\mathbf{x})\right],$$

where $\|f\|_L \leq 1$ are $1-$Lipschitz continuous functions ($f : X \to \mathbb{R}$):

$$|f(\mathbf{x}_1) - f(\mathbf{x}_2)| \leq \|\mathbf{x}_1 - \mathbf{x}_2\|, \quad \text{for all } \mathbf{x}_1, \mathbf{x}_2 \in X.$$

Kantorovich duality: metric case

Proposition 3: **optimal potentials characterization**

Let $\gamma^* \in \arg\inf_{\gamma \in \prod(\pi, p)} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\gamma} \|x - y\|$. Then there exists optimal $f^* : \mathbb{R}^D \to \mathbb{R}$, $\|f^*\|_L \leq 1$:

$$f^*(y) - f^*(x) = \|y - x\| \quad \gamma^* \text{ almost surely}$$