An Invitation to Statistics in Wasserstein Space

Wang Chao 2020212429

Shanghai University of Finance and Economics

January 16, 2022

Outline

Content

OT map

Definition (measure-preserving map)

A map $T: X \to Y$ is measure-preserving if for any measurable set $B \subset Y$,we have $\mu(T^{-1}(B)) = v(B)$.The measure-preserving condition is denoted as $T_{\#}\mu = v$, where $T_{\#}\mu$ is the push-forward measure induced by T.

Definition (OT map)

The solution to Monge's problem are called the OT maps,i.e. the OP maps are

$$\underset{T_{\#}\mu=v}{\textbf{argmin}} \int_X c(x, T(x)) d\mu(x)$$

The total transportation cost of an OT map is called the Wasserstein distance between μ and v, denoted as $W_c(\mu, v)$,

$$W_c(\mu, v) = \inf_{T_{\#}\mu=v} \int_X c(x, T(x)) d\mu(x)$$

Kantorovich Problem

Let the projection maps formally be $\pi_x(x,y) = x, \pi_y(x,y) = y$, then define the joint measure class as follows:

$$\Pi(\mu, v) := \{\rho(x,y) : \ X \times Y \to \mathbb{R} : (\pi_x)_{\#}\rho = \mu, (\pi_y)_{\#}\rho = v\}$$

Kontarovich's Problem (KP)

Given a transport cost function $c(x,y) : \ X \times Y \to \mathbb{R}^+$,find the joint probability measure $\rho(x,y) : \ X \times Y \to \mathbb{R}^+$ that minimizes the total transport cost.

$$\text{(KP)} \quad W_c(\mu, v) = \inf_{\rho \in \Pi(\mu, v)} \int_{X \times Y} c(x,y) d\rho(x,y)$$

Content

Discrete Case

(Prime Problem)

$$\begin{aligned} \textbf{min} \quad & b^T y \\ \text{s.t.} \quad & A^T y = c \\ & y \geq 0 \end{aligned}$$

$$L(y, \lambda) = b^T y - \lambda^T (A^T y - c)$$

$$\begin{aligned} g(\lambda) &= \inf_{y \geq 0} L(y, \lambda) \\ &= \inf_{y \geq 0} (b^T y - \lambda^T (A^T y - c)) \\ &= \inf_{y \geq 0} ((b^T - \lambda^T A^T) y + \lambda^T c) \end{aligned}$$

$$= \begin{cases} +\infty, & A^T\lambda > b \\ c^T\lambda, & A^T\lambda \le b \end{cases}$$

(Duality Problem)

$$\begin{aligned} \textbf{max} \quad & c^T x \\ \text{s.t.} \quad & A^T x \le b \end{aligned}$$

Discrete Case

(Discrete KP)

$$\begin{aligned} \textbf{min} \quad & \sum_{i=1}^{n}\sum_{j=1}^{m} \pi_{ij} c_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^{m} \pi_{ij} = \mu_i, i = 1, ..., n \\ & \sum_{i=1}^{n} \pi_{ij} = v_i, j = 1, .., n \\ & \pi_{ij} \geq 0 \end{aligned}$$

Discrete Case

(Discrete DP)

$$\begin{aligned} \textbf{max} \quad & \sum_{i=1}^{n} \varphi_i \mu_i + \sum_{j=1}^{m} \psi_j v_j \\ \text{s.t.} \quad & \varphi_i + \psi_j \leq c_{ij} \end{aligned}$$

Kantorovich Duality

Duality of KP

Given a transport cost function $c(x, y): \ X \times Y \to \mathbb{R}^+$,find the real integrable functions $\varphi: \ X \to \mathbb{R}$ and $\psi: \ Y \to \mathbb{R}$ such that

$$\text{(DP)} \quad \sup_{\varphi,\psi} [\int_X \varphi(x) d\mu + \int_Y \psi(y) dv : \ \varphi(x) + \psi(y) \leq c(x, y)]$$

We can directly derive the weak duality by

$$\int_X \varphi(x)d\mu(x) + \int_Y \psi(y)dv = \int_{X\times Y} c(x,y)d\rho(x,y) \leq W_c(\mu,v)$$

Theorem (Kontarovich Duality)

Let μ and v be probability measures on complete separable metric space X and Y,respectively ,and let $c(x,y):\ X\times Y\to\mathbb{R}^+$ be a measurable function. Then

$$\inf_{\rho\in\Pi(\mu,v)} \int_{X\times Y} c(x,y)d\rho(x,y) = \sup_{\varphi,\psi}\,[\int_X \varphi(x)d\mu + \int_Y \psi(y)dv : \varphi(x)+\psi(y)\leq c(x,y)]$$

Kantorovich Duality

Next,we consider the case that μ are absolutely continuous with respect to Lebesgue measure and the cost takes form as follows:

$$c(x,y) = \frac{\|x-y\|^2}{2}$$

Definition (Legendre transform)

Given a function $\varphi: \mathbb{R}^n \to \mathbb{R}$,its Legendre transform is defined as follows:

$$\varphi^*(y) = \sup_x \left[\langle x, y\rangle - \varphi(x)\right]$$

Legendre transform is of fundamental importance in convex analysis. Suppose that an oracle tells us that some φ is a good candidate.Then the largest possible ψ satisfying $(\varphi,\psi) \in \{(\varphi,\psi) : \varphi(x) + \psi(y) \leq c(x,y)\}$ is

$$\psi(y) = \inf_{x\in\mathbb{R}^d} \left[\frac{\|x-y\|^2}{2} - \varphi(x)\right] = \frac{\|y\|^2}{2} + \inf_{x\in\mathbb{R}^d} \left[\frac{\|x\|^2}{2} - \varphi(x) - \langle x, y\rangle\right]$$

Kantorovich Duality

In other words,

$$\tilde{\psi} := \frac{\|y\|^2}{2} - \psi(y) = \tilde{\varphi}^*, \quad \tilde{\varphi} = \frac{\|x\|^2}{2} - \varphi(x).$$

Now by symmetry,one can also replace $\tilde{\varphi}$ by $(\tilde{\psi})^* = (\tilde{\varphi})^{**}$,so it is reasonable to expect that an optimal dual pair should take the form $(\|\cdot\|^2/2 - \tilde{\varphi},\ \|\cdot\|^2/2 - (\tilde{\varphi})^*)$. The alternative representation of the dual objective value as

$$\int_{\mathbb{R}^d} \varphi d\mu + \int_{\mathbb{R}^d} \psi dv = \frac{1}{2}\int_{\mathbb{R}^d} \|x\|^2 d\mu + \frac{1}{2}\int_{\mathbb{R}^d} \|y\|^2 dv - \int_{\mathbb{R}^d} \tilde{\varphi} d\mu - \int_{\mathbb{R}^d} (\tilde{\varphi})^* dv.$$

Suppose that an optimal φ is found,according to the duality ,a necessary and sufficient condition is that

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\|x-y\|^2}{2} d\rho(x,y) = \int_{\mathbb{R}^d} \varphi(x) d\mu + \int_{\mathbb{R}^d} \psi(y) dv$$

Kantorovich Duality

Equivalently,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}[\tilde{\varphi}(x)+(\tilde{\varphi})^*(y)-\langle x,y\rangle]d\rho(x,y)=0.$$

Since we have $\tilde{\varphi}(x)+(\tilde{\varphi})^*(y)\geq\langle x,y\rangle$ everywhere,the integrand is nonnegative.Hence,the integral vanishes iff ρ is concentrated on the set of (x,y) such that $\tilde{\varphi}(x)+(\tilde{\varphi})^*(y)=\langle x,y\rangle$.By the definition of the Legendre transform as a supremum, this happans if and only if the supremum defining $(\tilde{\varphi})^*(y)$ is attained at x.Equibalently,

$$\tilde{\varphi}(z)-\tilde{\varphi}(x)\geq\ \langle z-x,y\rangle,\quad z\in\mathbb{R}^d.$$

This condition is precisely the definition of y being a subgradient of $\tilde{\varphi}$ at x. When $\tilde{\varphi}$ is differentiable at x,its unique subgradient is the gradient $y=\nabla\tilde{\varphi}(x)$. The problem, of course, is that $\tilde{\varphi}$ may fail to be differentiable μ-almost surely. This is remedied by assuming some regularity on the source measure μ in order to make sure that any convex function be differentiable μ-almost surely, and is done via the following regularity result, which, roughly speaking, states that convex functions are differentiable almost surely.

Theorem 1.2 (Differentiability of Convex Functions)

Let $f: \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be a convex function with domain $\text{dom} f = \{x \in \mathbb{R}^d : f(x) < \infty\}$ and let $\mathcal{N}$ be the set of points at which f is not differentiable.Then $\mathcal{N} \cap \text{dom} f$ has Lebesgue measure 0.

Kantorovich Duality

Based on the argument above,we can give a fundamental existence and uniqueness result for the Monge-Kantorovich problem.

Theorem 1.3 (Quadratic Cost in Euclidean Space)

Let μ and v be probability measures on $\mathbb{R}^d$ with finite second moments, and suppose that μ is absolutely continuous with respect to Lebesgue measure. Then the solution to the Kantorovich problem is unique, and is induced from a transport map T that equals μ-almost surely the gradient of a convex function φ. Furthermore, the pair $(\|x\|^2/2 - \phi,\ \|y\|^2/2 - \phi^*)$ is optimal for the dual problem.