Optimal Transport Notes

Ivan Zhytkevych

October 27, 2023

Contents

1 Probability Measures

Probability vectors gives a probability point mass in a vector form. For each of the outcomes of the random variable corresponds one row/column in the vector.

$$x_0 = \begin{pmatrix} 0.25 & 0.5 & 0.1 & 0.15 \end{pmatrix}$$

(TODO: what is a measure)

1.1 Discrete measure

Definition 1.1 (Discrete measure)**.** A discrete measure with weights α and locations $x_1, \dots, x_n \in \mathcal{X}$ reads

$$a = \sum_{i=1}^{n} \alpha_i \delta_{x_i}$$

where δ_x is the Dirac delta function.

1.2 Simplex

Let $x = (x_0, x_1, x_2, \dots, x_n)$ be a probability vector

$$\sum_{i=1}^{n} x_i = 1$$

So simplex should be a set of probability vectors

$$\Sigma_n := \left\{ a \in \mathbb{R}_+^n : \sum_{i=1}^{n} a_i = 1 \right\}$$

2 Monge Problem

Discrete measures:

$$\alpha = \sum_{i=1}^{n} = \mathbf{a}_i \delta_{x_i} \quad \text{and} \quad \beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{y_j}$$

Seek for a map that associates to each point x_i a single point y_i and which must push the mass of α toward the mass of β:

$$T : \{x_1, \dots, x_n\} \to \{y_1, \dots, y_m\}$$

$$\forall j \in [[m]], \;\; \mathbf{b}_j = \sum_{i:T(x_i)=y_j} a_i$$

compactly

$$T_{\#}\alpha = \beta$$

This map should minimize the transportation cost which is the sum of each single point transportation:

$$\min_T \left\{ \sum_i c(x_i, T(x_i)) \ : \ T_\# \alpha = \beta \right\}$$

3 Kantorovich Relaxation

$$\mathbf{U}(\mathbf{a}, \mathbf{b}) := \left\{ \mathbf{P} \in \mathbb{R}_+^{n \times m} \ : \ \mathbf{P} \nVdash_m = \mathbf{a} \ \text{ and } \ \mathbf{P}^T \nVdash_n = \mathbf{b} \right\}$$
$$\nVdash_n = \left(a_{i,j} = 1 \ : \ i = n \right)$$

Kantorovich optimal transport reads:

$$L_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) := \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b})} \langle \mathbf{C}, \mathbf{P} \rangle := \sum_{i,j} \mathbf{C}_{i,j} \mathbf{P}_{i,j}$$

4 Wasserstein distance

Proposition 4.0.1. Suppose that $n = m$ and that for some $p \geq 1$

$$\mathbf{C} = \mathbf{D}^p = (\mathbf{D}_{i,j}^p)_{i,j} \in \mathbb{R}^{n \times n}$$

where $\mathbf{D} \in \mathbb{R}_+^{n \times n}$ is a distance on $[[n]]$, i.e.

1. $\mathbf{D} \in \mathbb{R}_+^{n \times n}$ is symmetric
2. $D_{i,j} = 0 \Leftrightarrow i = j$
3. $\forall (i, j, k) \in [[n]]^3, \ \mathbf{D}_{i,k} \leq \mathbf{D}_{i,j} + D_{j,k}$

Then

$$W_p(\mathbf{a}, \mathbf{b}) := L_{\mathbf{D}^p}(\mathbf{a}, \mathbf{b})^{1/p}$$

defines the *p-Wasserstein distance* on Σ_n, i.e. W_p is symmetric, positive, $W_p(\mathbf{a}, \mathbf{b}) = 0$ if and only of $\mathbf{a} = \mathbf{b}$, and it satisfies the triangle inequality.