Optimal transport & Wasserstein Distance

How to calculate distances between pairs of PROBABILITY DISTRIBUTIONS ?

DISTANCES :

- Simmetry
- Triangle inequality (Schwartz' inequality)

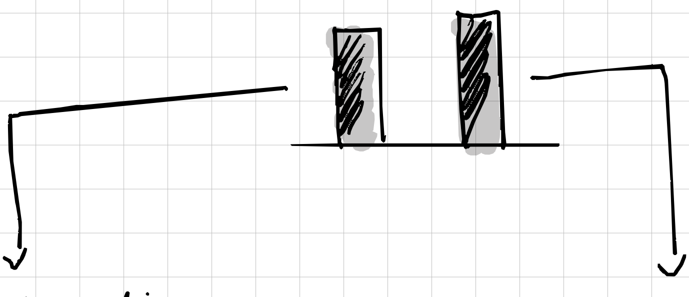
The weaker notion of distances are often called "divergens".

$$D_{KL}(P\|Q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx$$

① IT'S NOT SYMMETRIC :(

$$D_{KL}(P\|Q) \neq D_{KL}(Q\|P)$$

② $D_{KL}(P\|Q) \to +\infty$ if P & Q have different support.



SOLUTION 1:

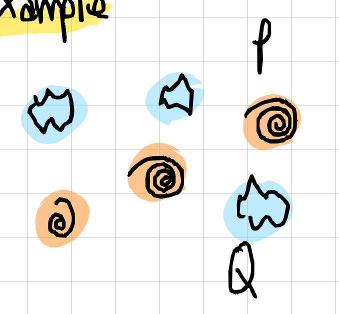
Smooth the distributions to match support

Use a different divergence

WASSERSTEIN DISTANCE

(• SYMMETRIC ✓
• TRIANGLE INEQUALITY ✓)

2D example



→ move the ☆ to fill the @

>> Want to find the best Transportation plan

"MOST EFFICIENT" → "MINIMIZE TRANSPORTATION COST" C

$$C(x_0, y_0, x_1, y_1) = (x_0 - x_1)^2 + (y_0 - y_1)^2$$

Then define a "TRANSPORTATION PLAN", how many units of dirt $(x_0, y_0) \to (x_1, y_1)$

$$T(x_0, y_0, x_1, y_1) = w$$

→ units of dirt from (x_0, y_0) to (x_1, y_1)

$T(x_0, y_0, x_1, y_1)$ must:

- $T \geq \emptyset$
- $\iint T(x_0, x, y_0, y)\,dx\,dy = p(x_0, y_0) \quad \forall (x_0, y_0)$
- $\iint T(x, x_1, y, y_1)\,dx\,dy = q(x_1, y_1) \quad \forall (x_1, y_1)$

Then:

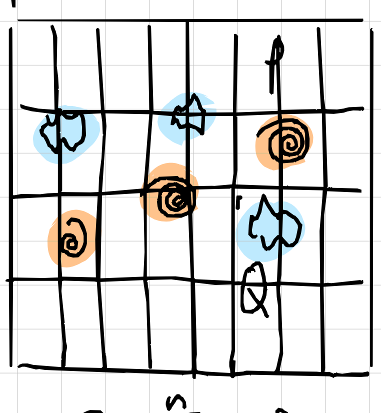
$$\text{TOTAL COST} =: \iiiint C((x_0, y_0, x, y)) \cdot T \, dx_0\, dy_0\, dx\, dx_2$$

→ Finding the best transport plan is a LINEAR PROGRAM after discretizing

"scipy.optimize.linprog(...)"

→ OPTIMAL TRANSPORT BETWEEN DISCRETE DISTRIBUTIONS

Difficult between continuos, not always analytically tractable



DISCRETIZATION ERROR but FINITE NUMBER OF BINS

$$O(d^3/\log d)$$

$$P = \sum_{i=1}^{n} p_i \delta_{x_i} \qquad Q = \sum_{i=1}^{n} q_i \delta_{x_i}$$

- $C_{ij} = \|x_i - x_j\|^2$ from bin $i \to j$
- $T \in \mathbb{R}^{n \times n}$
- total cost $= \langle T, C \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} T_{ij} C_{ij}$

→ has similar properties as before

So our goal is to $\min_T \langle T, C \rangle = T^*$

The W-distance $W(P,Q) = (\langle T^*, C \rangle)^{1/2}$

→ $\emptyset$ if $P = Q$

[• T^* turns out to be very sparse
• The peaks of T^* are those of the densities]

News : Entropy Regularization

important innovation that has galvanized recent work :

Penalize transport plans with small SHANNON ENTROPY

$$\text{minimize}_T \; \langle T, C \rangle - \epsilon H(T) \qquad \epsilon > \emptyset$$

$$(T\mathbf{1} = p,\; T^T\mathbf{1} = q,\; T \geq \emptyset)$$

$$H = -\sum_{ij} T_{ij} \log T_{ij} \quad \text{shannon entropy}$$

As $\epsilon \to \infty$, the problem becomes increasingly easy to solve and with solution $T_{ij}^* = p_i q_j$

→ After regularization, we can expect nearly linear time.