L10: Distances for Distributions

Feb 10, 2025

Jeff M. Phillips

Abstract Representation

1. Data point → Set
2. Data point → Vector

Today 3. Data point → Distribution

Discrete Distribution

m possible outcomes

$m = 100{,}000$ words English

$m = 29$ countries in Utah

↳ vector in $\mathbb{R}^m$ too flexible

Discrete Prob. Dist   m states

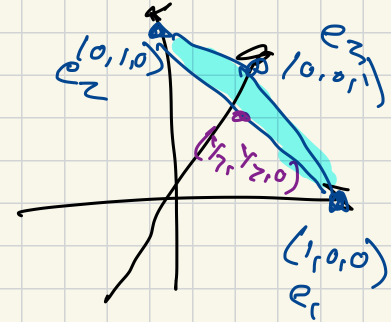
$v \in \mathbb{R}^m$, but restrict.

v_j = probability in state $j \in [m]$

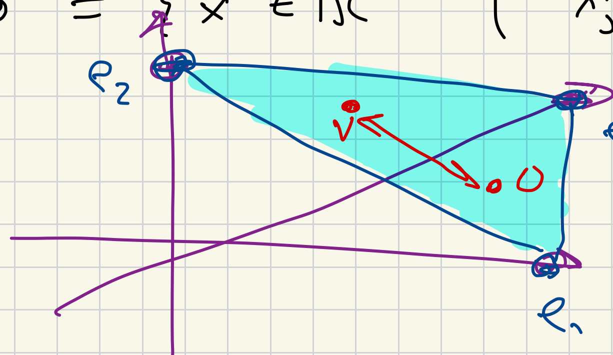
$v_j \in [0,1]$

$\sum_{j=1}^{m} v_j = 1$

$\mathbb{R}^3$



$$\Delta^{m-1} = \{ x \in \mathbb{R}^m \mid x_j \geq 0, \ \sum_{j=1}^{m} x_j = \|x\|_1 = 1 \}$$



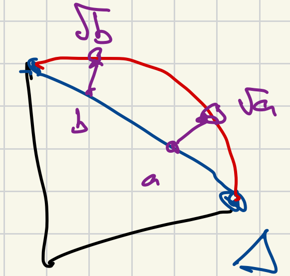
Δ^2

2-simplex

$v, u \in \Delta^m$

$D(v,u) = \|v - u\|$

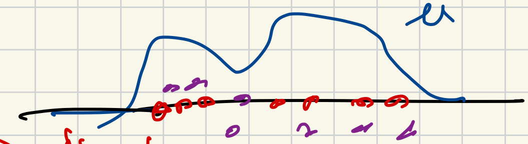
Data $a, b \in \Delta^{m-1}$



common distances

- $D_{KL}(a,b) = \sum_{j=1}^{m} a_j \ln(a_j/b_j)$
- $D_{H}(a,b) = \frac{1}{\sqrt{2}} \sqrt{\sum_{j=1}^{m} \left(\sqrt{a_j} - \sqrt{b_j}\right)^2}$
- $D_{TV}(a,b) = \max_{S \subset [m]} \sum_{j \in S} (a_j - b_j) = \frac{1}{2} \|a-b\|_1$

Distribution μ continuous



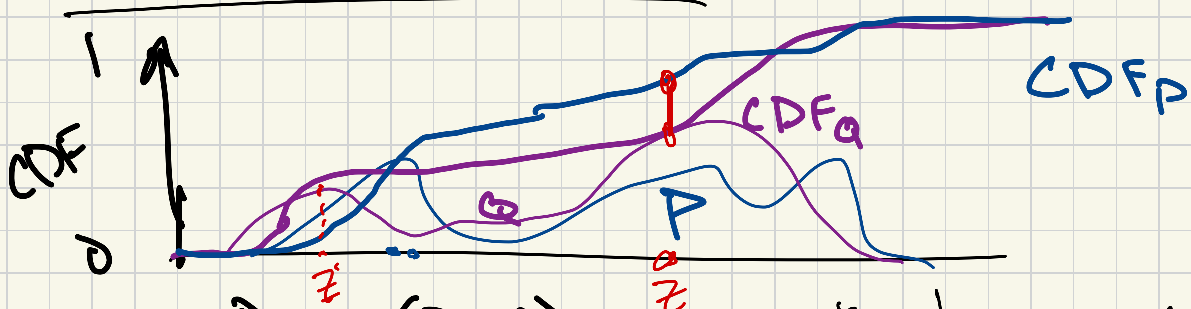
$\int \mu \, dx = 1$

sample $P \sim \mu$

$P = Alg_0 \, 1$

$Q = Alg_0 \, 2$

Kolmogorov - Smirnov (KS) Distance



$$D_{KS}(P, Q) = \max_{z \in \mathbb{R}} \left| CDF_P(z) - CDF_Q(z) \right|$$

0,321 inches

0,323 inches

Inputs continuous distributions on $\mathbb{R}^d$

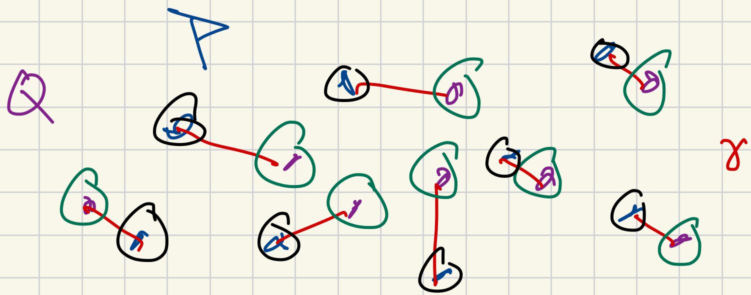
μ_P, ν_Q $\qquad$ $P \sim \mu_P$ $\quad$ $Q \sim \nu_Q$

$P, Q \subset \mathbb{R}^d$ $\qquad$ assume $|P| = |Q| = n$

- Wasserstein
  Earth Movers, Optimal Transport

- Kernel Distance
  Maximum Mean Discrepancy (MMD)

Assume input $P, Q \subset \mathbb{R}^d$, $|P| = |Q| = n$



Transportation Plan γ

$\gamma \in \Gamma(P, Q)$

Earth Movers Distance

$$\min_{\gamma \in \Gamma(P,Q)} \frac{1}{|P|} \sum_{(p,q) \in \gamma} \|p - q\| = W_1$$

Wasserstein W_s

$$W_s(P,Q) = \min_{\gamma \in \Gamma(P,Q)} \left(\frac{1}{|P|} \sum_{(p,q) \in \gamma} \|p-q\|^s \right)^{1/s}$$

metric for $s \in [1, \infty), \infty$

$$W_s(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \left(\mathop{E}_{(p,q) \sim \gamma} D(p,q)^s \right)^{1/s}$$

Computation on W_1, W_2

Need to find $\arg\min_{\gamma \in \Gamma(P, Q)}$

Algorithms $O(n^3 \log n)$ time

input might be n^2 size

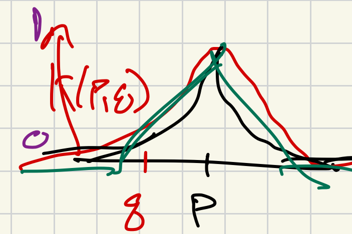
Approx W_2 $\Rightarrow$ Sinkhorn adds entropy regularization

↳ convex optimization $O(\frac{1}{\varepsilon} n^2 \log(n))$

ε approx

Kernel Distance (MMD)

Similarity $p, q \in \mathbb{R}^d$ Kernel



Gaussian $K(p,q) = \exp(-\|p-q\|^2)$

Laplace $K(p,q) = \exp(-\|p-q\|)$

Triangle $= \max\{0, 1-\|p-q\|\}$

general $K(p,q) = \exp(-D(p,q)^2)$

$\boxed{P1}$ $K(p,q) \in [0,1]$

$\boxed{P2}$ K is <u>positive definite</u>

For any X,
Gram matrix

$G \in \mathbb{R}^{n \times n}$

$G_{ij} = K(x_i, x_j)$

G has all real, pos. eigenvalues.

Analogy $\|p-q\|^2 = \|p\|^2 + \|q\|^2 - 2\langle p, q\rangle$

$= \langle p, p\rangle + \langle q, q\rangle - 2\langle p, q\rangle$

Use $K(p, q)$ as $\langle p, q\rangle$

$$D_K(p, q) = \sqrt{K(p, p) + K(q, q) - 2K(p, q)}$$

$$K(P, Q) = \frac{1}{|P|} \cdot \frac{1}{|Q|} \sum_{p \in P} \sum_{q \in Q} K(p, q)$$

$$D_K(P, Q) = \sqrt{K(P, P) + K(Q, Q) - 2K(P, Q)}$$

$X = P \cup Q \Rightarrow$ Gram matrix $G(X) \in \mathbb{R}^{2n \times 2n}$

$G_{ji} = K(x_i, x_j)$

Eigen decomposition $G \rightarrow [U_s, \lambda]$ $U = (u_1, u_2 \ldots u_{2n})$

basis

Project $X = P \cup Q \rightarrow$ basis U.

in basis for dist $= D_K(p, q) = \|p - q\|_U$

Reproducing Kernel Hilbert Space (RKHS)

$\mathcal{H}_K$ = linear combo $K(p, \cdot)$

$$K(p, \cdot) \in \mathcal{H}_K$$

Kernel Density Estimate $\quad KDE_P(x) = \frac{1}{|P|} \sum_{p \in P} K(p, x)$

$$KDE_P(\cdot) \in \mathcal{H}_K$$

$$D_K(P, Q) = \| KDE_P(\cdot) - KDE_Q(\cdot) \|_{\mathcal{H}_K}$$

$$= \| \mathop{E}_{p \sim P} K(p, \cdot) - \mathop{E}_{q \in Q} K(q, \cdot) \|_{\mathcal{H}_K}$$